

Feature Extraction

The diagram illustrates a feature extraction pipeline. It begins with an input image of St. Peter's Basilica. This image is processed by a series of layers: three gray layers, one orange layer, and one blue layer. The output of the blue layer is a feature map. This feature map is then processed by a dashed box containing two orange layers, which produces another feature map. Finally, this feature map is processed by a dashed box containing two blue layers, which produces the final output feature map. A red dashed line connects the input image to the final output feature map.

Sub-pixel Refinement

The diagram illustrates the Sub-pixel Refinement module. It starts with **Sample Patch Features** (green blocks) and **Proposal Matches** (dashed red line). These are processed by **Self&Cross Attention** (blue block, repeated $\times 1$), followed by **Scale Align** (green block) and **Center Features** (orange blocks). The output is then processed by **Spatial Attention** (green blocks) and **Conv & Softmax** (green blocks) to produce the final refined features. The diagram also shows a visual representation of the refined features as a grid of colored squares.

Feature Interaction With Co-visible Area Segmentation (FICAS)

The diagram illustrates the FICAS module architecture. It starts with input feature maps $F_{1/8}^{A_0}$ and $F_{1/8}^{B_0}$. These are processed by a **Self&Cross Attention** block (labeled $\times 1$) to produce intermediate features $F_{1/8}^{A_1}$ and $F_{1/8}^{B_1}$. $F_{1/8}^{A_1}$ is then processed by a **Decoder** to generate a query Q_A , and $F_{1/8}^{B_1}$ is processed by another **Decoder** to generate a query Q_B . These queries are used in two **Cross Attention** blocks (labeled $\times 1$) to produce features $F_{1/8}^{A_2}$ and $F_{1/8}^{B_2}$. Finally, these features are processed by another **Self&Cross Attention** block (labeled $\times 1$) to produce the final output features. The diagram also shows **Co-visible Area Segmentation Modules** that receive inputs from the feature maps and the queries Q_A and Q_B .

Adaptive Assignment and Filtering

The diagram illustrates the Adaptive Assignment and Filtering process. It starts with three inputs: P_A (a grayscale image), $F_{1/8}^{A_3}$ (a green feature map), and $F_{1/8}^{B_3}$ (an orange feature map). P_A and $F_{1/8}^{A_3}$ are processed by 'Patch-level Matching'. The output of 'Patch-level Matching' and $F_{1/8}^{B_3}$ are then processed by 'Matches Filtering'. The final output is a filtered feature map where patches are highlighted in yellow and orange.