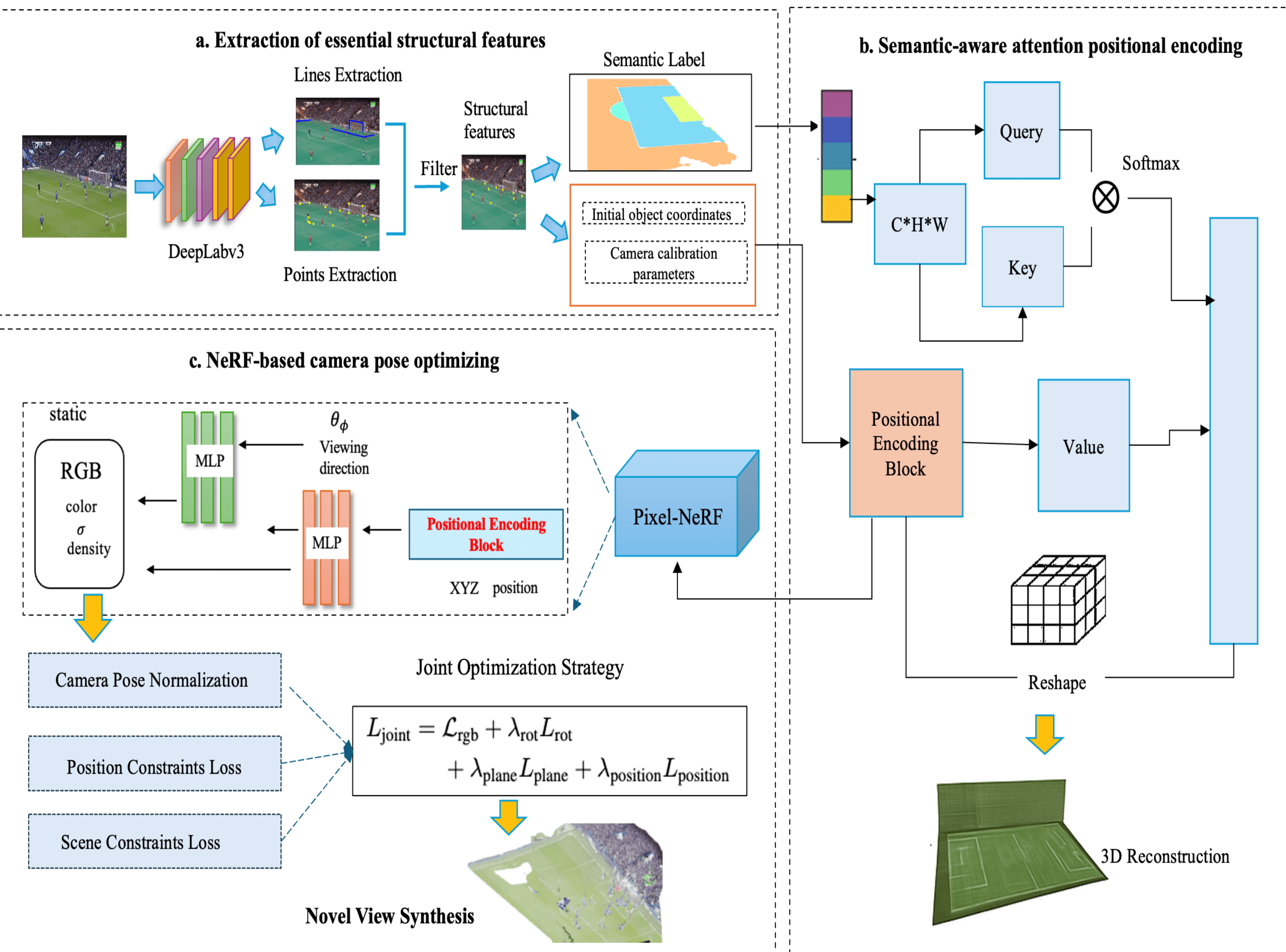


Motivation

Monocular cameras are widely used in sports video analysis due to their low cost, ease of deployment, and fast processing. However, in dynamic and complex sports environments like football fields, accurate pose estimation is challenging. We are motivated to develop a robust, data-efficient camera calibration framework that:

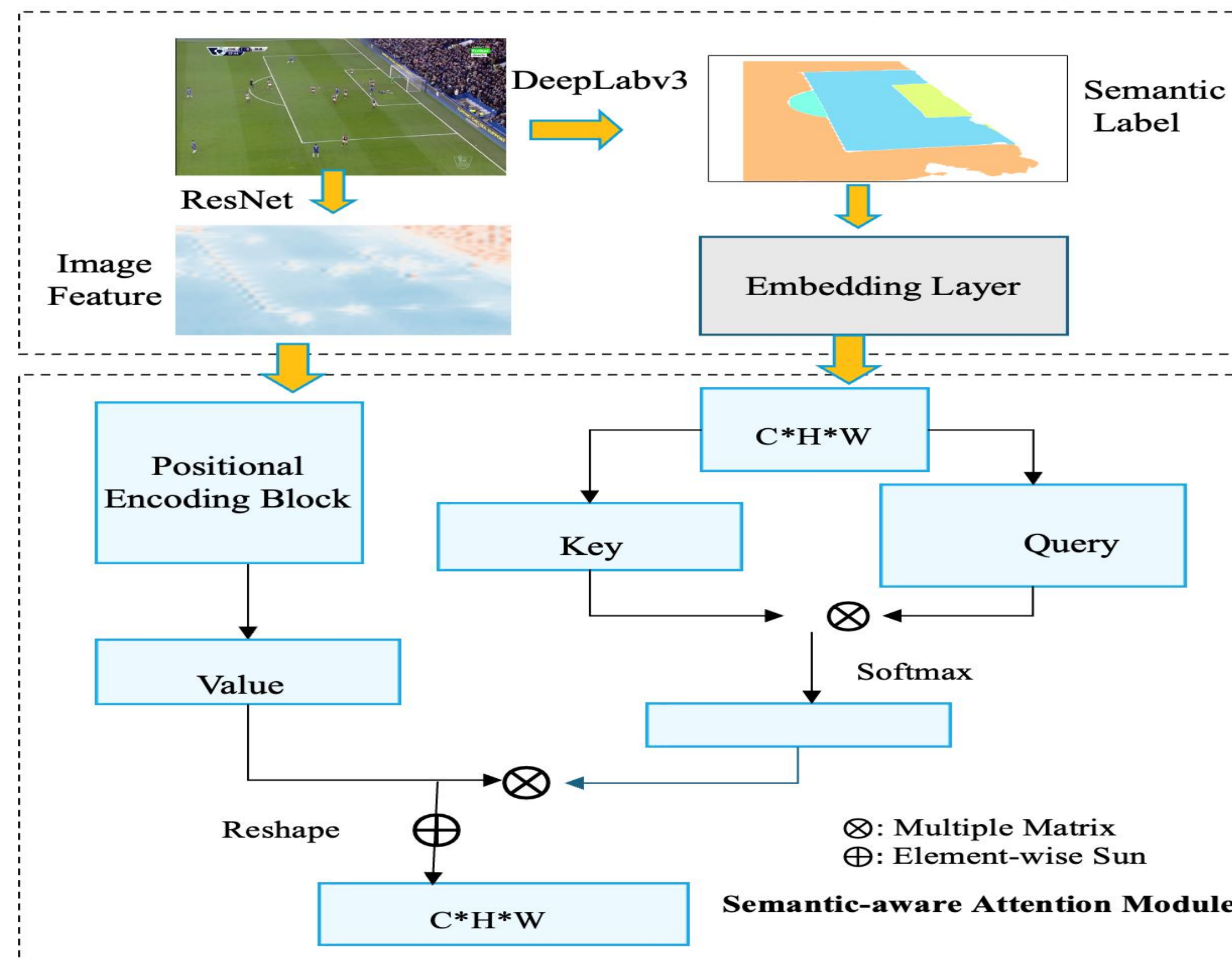
- 1, Leverages monocular depth and semantic scene understanding.
- 2, Enhances pose estimation via NeRF-based constraints and attention mechanisms
- 3, Reduces dependence on large datasets while improving accuracy.

➤ The framework of NeRF-guided Camera Optimization

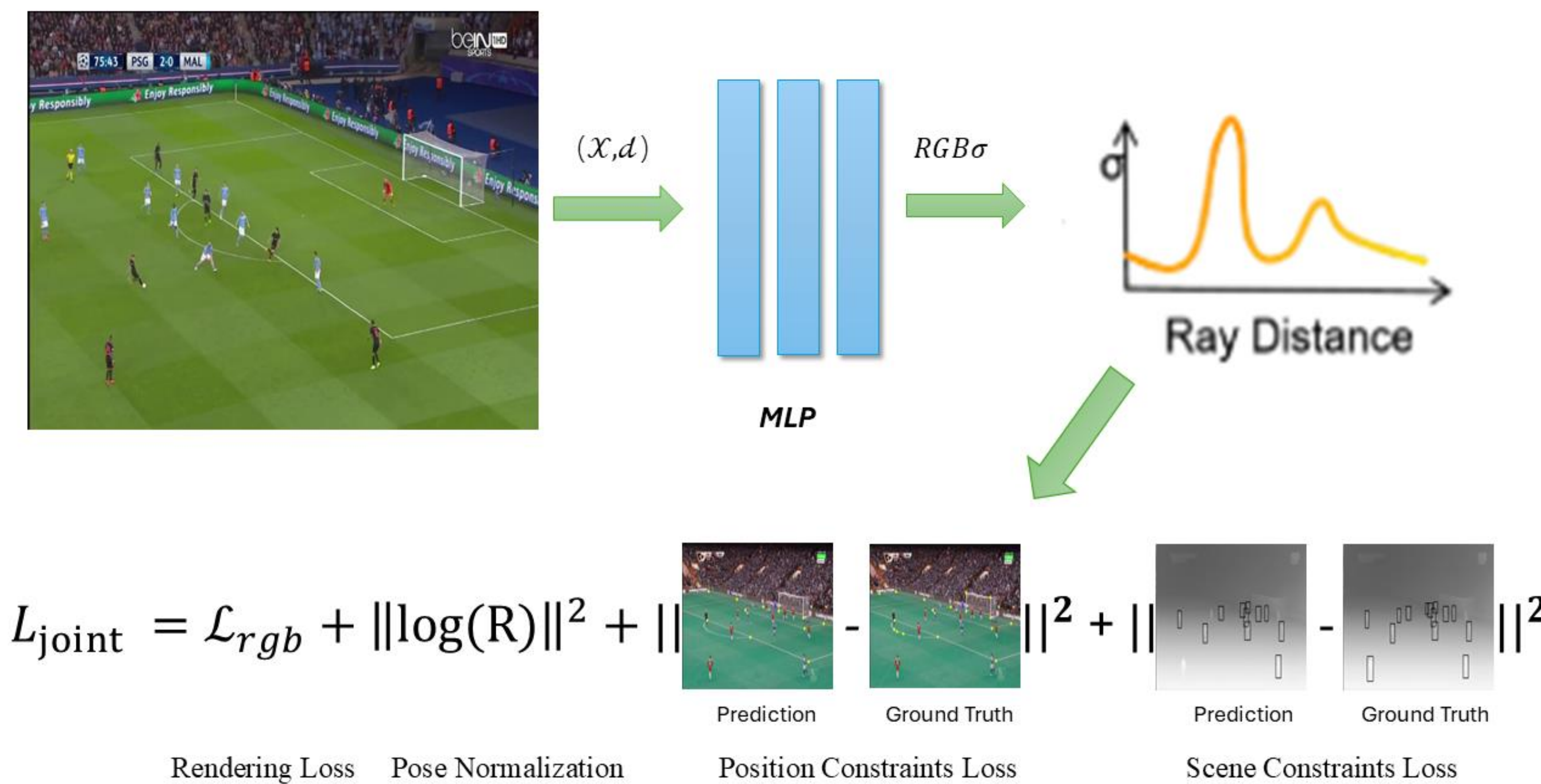


Method

➤ Semantic-aware attention positional encoding



➤ camera pose constraints



Results

➤ Evaluation for calibration accuracy and projection optimization

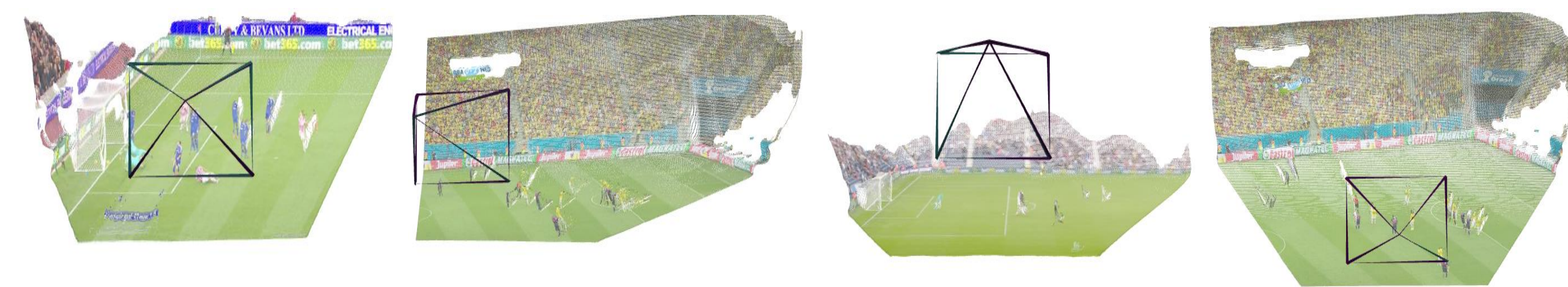
Approaches	IoU (Part)Med	IoU (Part)Med	IoU (Part)Med	IoU (Part)Med
CCBV-SN	79.8	81.7	88.5	92.3
SCCVSD	89.4	93.8	94.5	96.1
TVcalib	—	95.3	96.6	—
NBJW	89.4	93.0	94.4	96.9
[1]	89.8	92.9	95.1	96.7
Ours	92.1	94.6	96.3	97.8

➤ Evaluation the reprojection error induced by camera parameters

Approaches	JaC5	JaC10	CR
BroadTrack	75.3	86.8	69.8
TVCalib	52.9	73.4	66.5
NBJW	73.7	86.7	77.5
Ours	75.8	86.9	78.3



➤ Camera pose optimization



References

[1] Olivier D Faugeras, Q T Luong, and Stephen J Maybank. Camera self-calibration: theory and experiments. In Euro-pean Conference on Computer Vision (ECCV), pages 321–334, 1992. 7