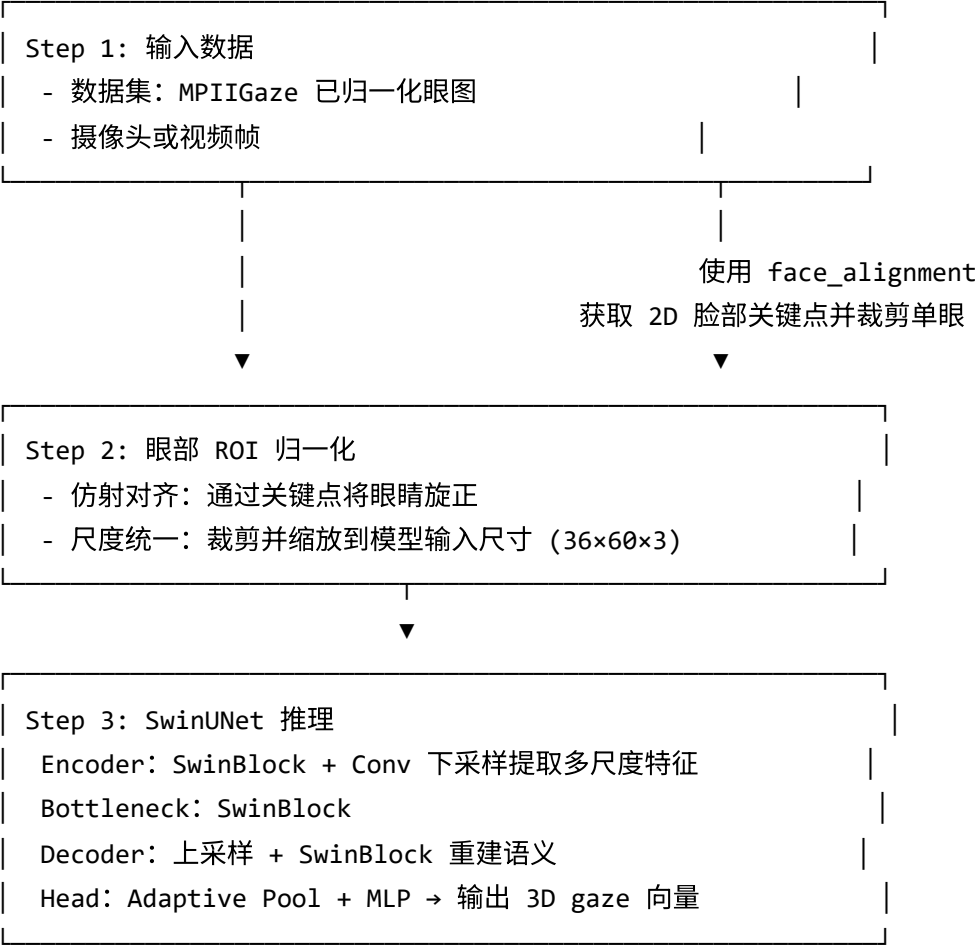


SwinUNet-VOG 算法概况

1. 算法 Pipeline

核心流程围绕“视频/图像 → 眼部 ROI → SwinUNet → 3D gaze”展开，分为三步：

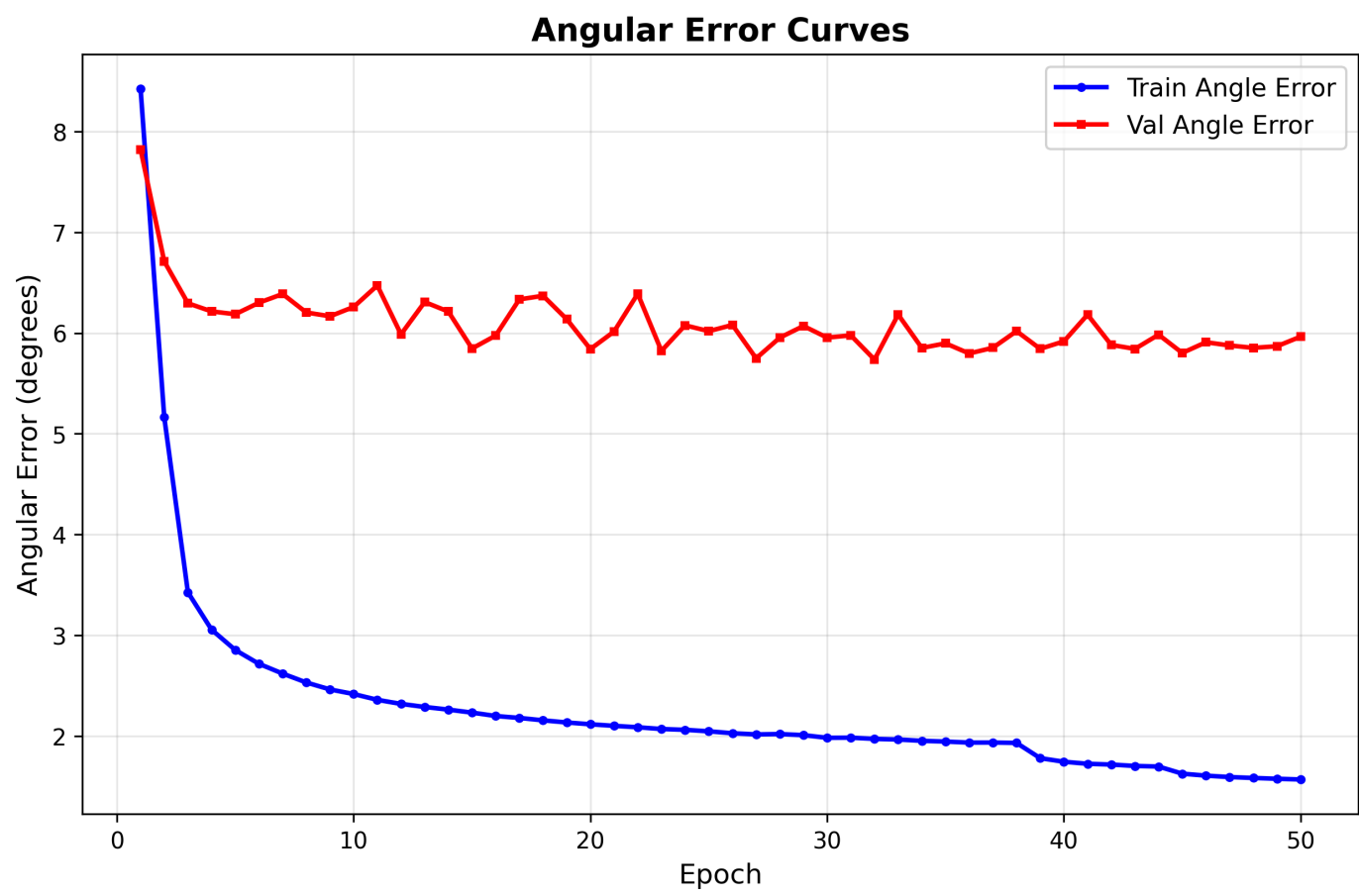


说明：

- **模型输入：**MPIIGaze 的归一化眼图直接满足输入尺寸，实际情况的摄像头采集的帧则通过 face_alignment + 仿射变换将原始帧裁剪为同样规格的 ROI。注：face_alignment 是基于 google 的 mediapipe 现有的 2D 关键点检测模型。
- **SwinUNet 结构：**编码端堆叠 SwinBlock + Conv 完成下采样，解码端使用上采样再接 SwinBlock 恢复分辨率，最后经全连接头输出 3×1 的 gaze 向量（注视点 x,y,z 坐标）。

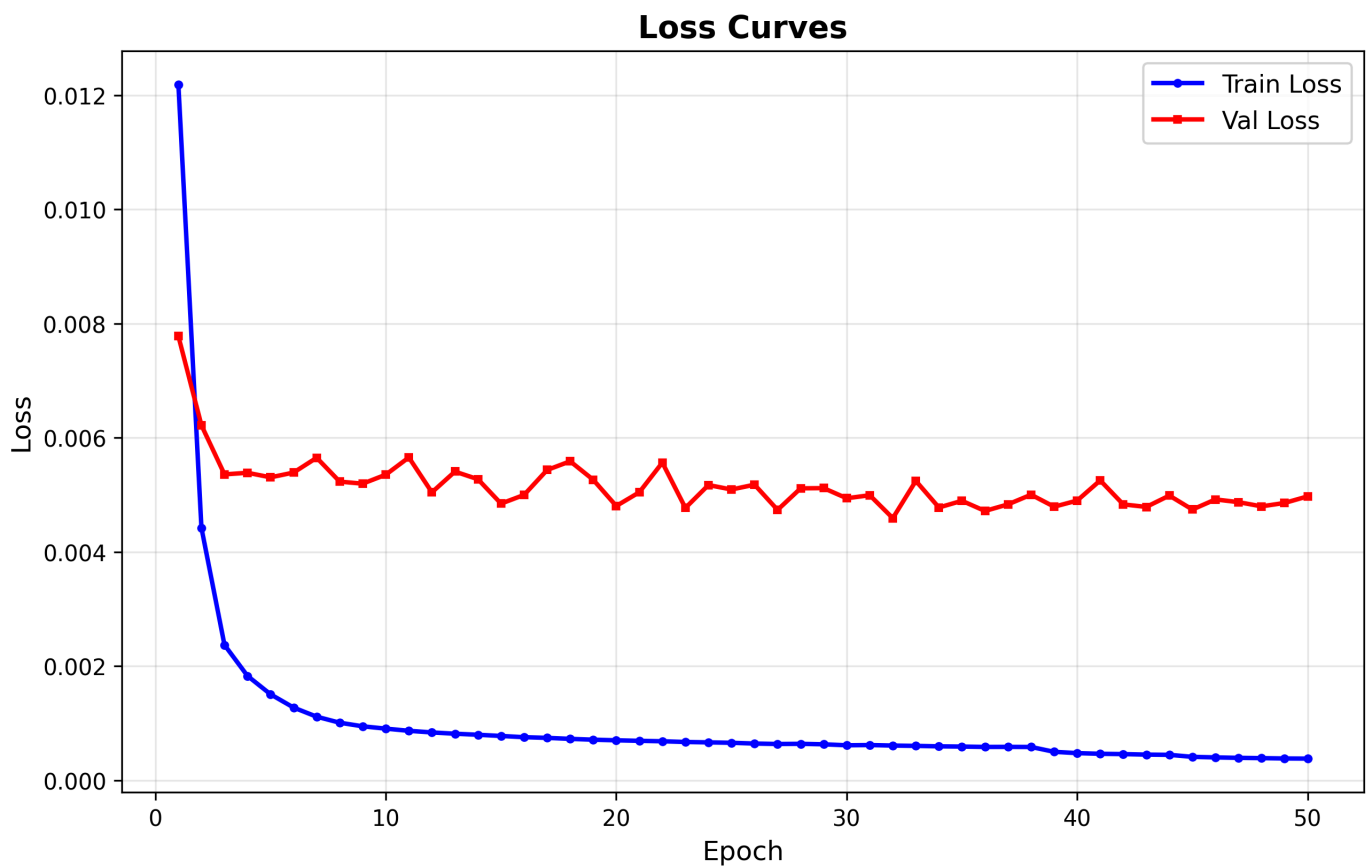
2. 训练记录

- 角度误差曲线



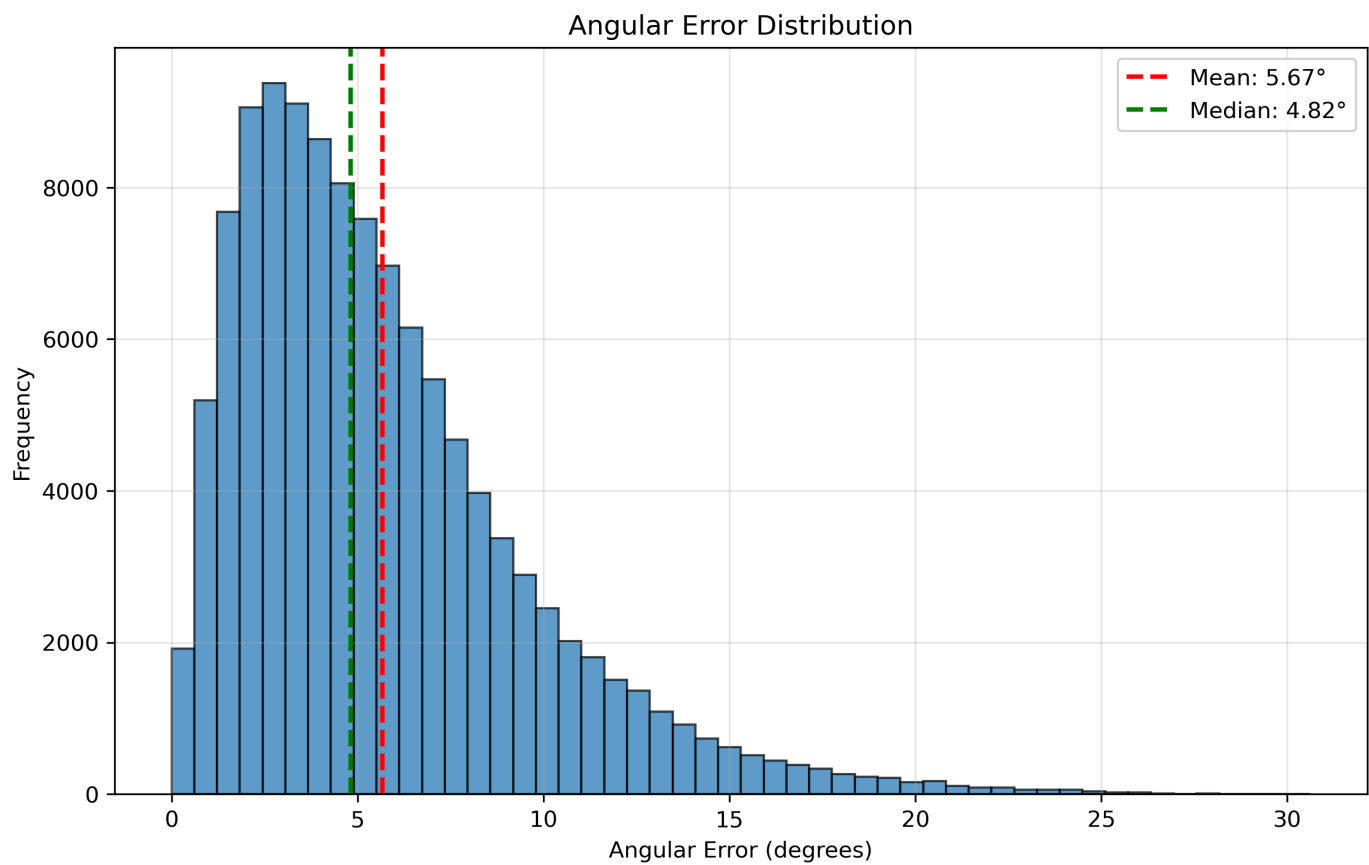
训练误差持续下降，验证误差在 6° 左右波动。

- 损失曲线



与角度误差趋势一致，训练集稳定收敛。

- 误差分布



平均角误差约 5.7°，中位数约 4.8°。

3. 效果展示

