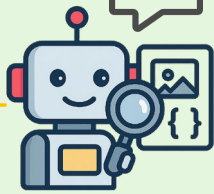


Stage 1

Scene Understanding Prompt

Key Frames

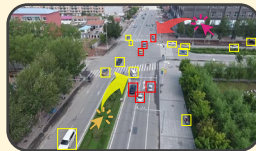
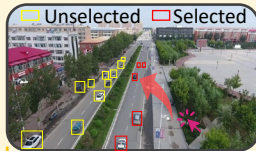
SUP
Agent



Scene Digest

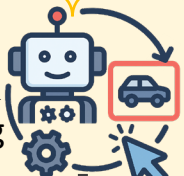
Stage 2

Semi-automated Object Labeling



Human
Click

Auto
labeling



SOL Agent

Cars travelling in the
same direction as us

Prompt Text

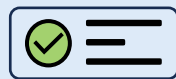
Bbox & Id
Text



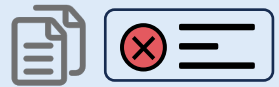
Expressions

Stage 3

Consistency Check (Optional)



Pass



Failed



CC Agent



Key Frames

Stage 4

Expressions Expansion

EE Agent



Human
Check



AerialMind Dataset

Timeline