

Vision Transformers (ViTs) are employed to assess images and videos captured from construction locations—like aerial shots or video feeds—to spot any discrepancies from the plans outlined in Building Information Modeling (BIM).