Q1. What algorithm or architecture do you think of which can satisfy the requirement (if any)? (Analogy will suffice), bonus point (optional) if you can think of some innovative method outside of present architectures like stacked LSTM, transformers, etc

Ans-1 Pre-trained transformers are the best architecture present according to me to get this done . The best part about deploying transformers is that it combines the benefits of convolutional [neural networks](https://www.analyticsinsight.net/backpropagation-in-neural-networks-how-it-helps/) (CNNs) and recurrent [neural networks](https://www.analyticsinsight.net/guide-building-deep-learning-neural-networks/) (RNNs). Both of these have had wide usage in pattern recognition, recognizing objects in pictures.

Q2. Any automatic (python script) approach can you think of to label the data for training?  
(Don’t overthink, think of simple approaches that you use as human)

Ans-2 The pipeline of our tool is as shown below that mainly consists of the detector & tracker, auto-label module, and I/O module for outputting & saving the machine annotated labels to the disk.

Diagram, schematic

Description automatically generated