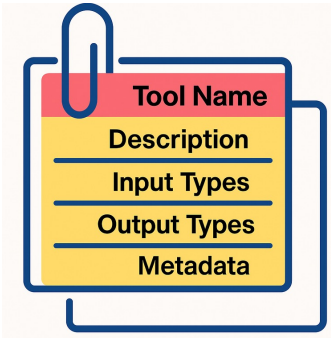
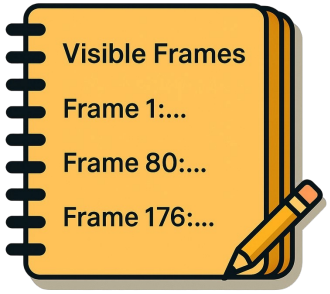
<0> <60> <120> <180> <240> <300> <360> <420> <480> <540> <600> <660> sec

Video Toolkit



Tool Card



Visible Frame Dictionary

In this video, by using a tractor, how many times faster is the cultivation of the same land compared to manual labor?

Calling **temporal-grounding-tool** to ground the tractor.

Temporal-grounding-tool: Tractor appears from frame <170> to <190>.

Calling **image-caption-tool** for the 20 frames from <170> to <190>.

Image-caption-tool: … <176>: The image is composed of two pictures. In the first picture, two people are working together to plow the field, while in the second picture, a single person is operating the tractor. The texts in the image indicate the time it takes to complete the task…

Calling **frame-selector** to select key frames.

Frame-selector: Look into all current visible frames, frame <176> selected.

Frame <176>:

It seems there are some texts in the image. Calling **OCR-tool** for frame <176>.

OCR-tool: “45 min” and “45 sec” in frame <176>.

Calling **summarizer** to summarize all the information.

Summarizer: Using a tractor is 10 times faster than manual labor.