Efficient Multi-purpose Video Annotation for Fast Labeling

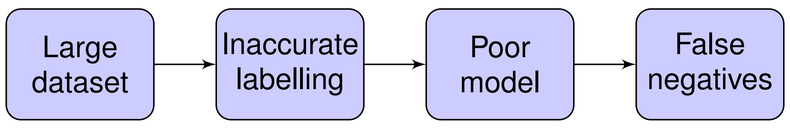
Mobina Mobaraki, Soodeh Ahani, Kwang Moo Yi, Mahyar Asadi, Klaske Van Heusden, Guy A. Dumont

Dept. of Electrical and Computer Engineering, The University of British Columbia, Canada
mobinamb@student.ubc.ca

Submitted to 2023 IEEE international conference on Visual Communications and Image Processing (VCIP)

Problem Definition

- Creating high-quality annotations can take up to 80% of the development time of a machine learning project.



Proposed Solution

- A **video annotation method** that **tracks the location of the cursor** while the video is playing and does not require the user to click on each individual frame, thereby **reducing the time needed for consistent labeling**.

Introducing the Interface

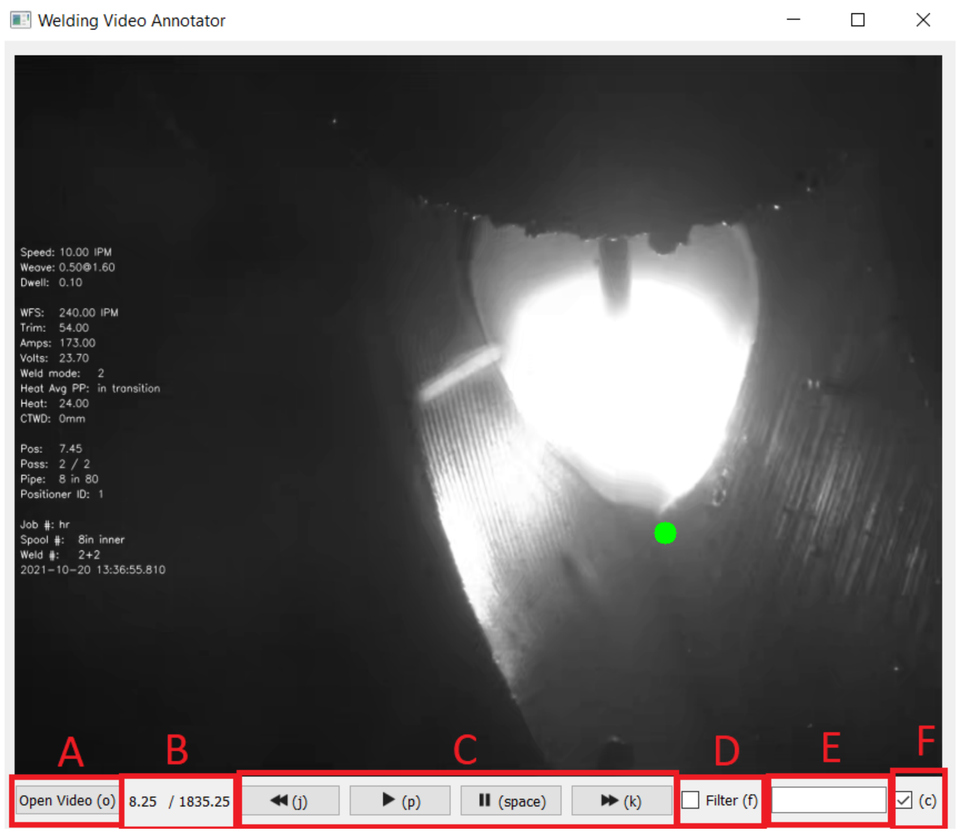


Fig. 1: The proposed video annotator interface.

- Button A (short key : “o”): data loading.
- Box B: current timing of the presented frame.
- Buttons C (short keys “j”, “p”, “space”, and “k”): video player features
- Checkbox D (short key : “f”): equalizer filter
- text box E: classification features
- Checkbox F (short key "c"): enables cursor tracking

conclusion

- We introduce a **fast, consistent, and open-access interface** for **video dataset labeling**, suitable for offline use and confidential data.
- Our interface accelerates the welding application labeling by **21 times**, offers dataset filtering, and supports multiple labeling tasks including **keypoint detection, object detection, and segmentation**.
- Future improvements include **machine learning predictions** for future frames and **collaborative annotation** features to boost annotation consistency.

Applications

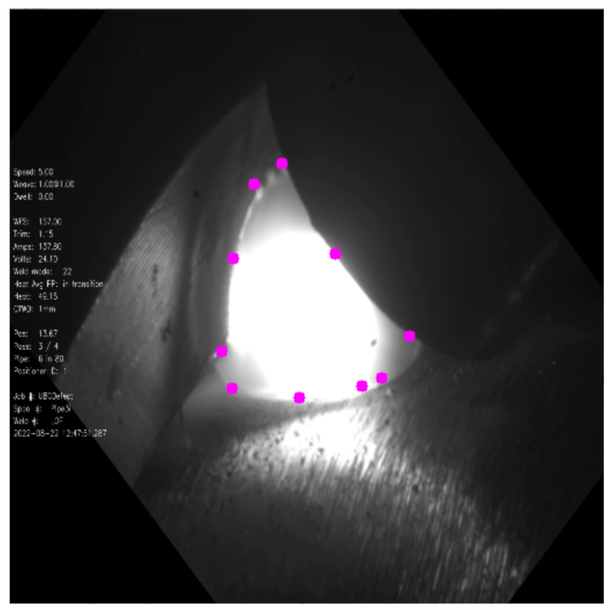


Fig. 2: The segmentation application. The users may use multiple points to create a boundary of an object for segmentation purposes when the shape of the object is homogeneous.

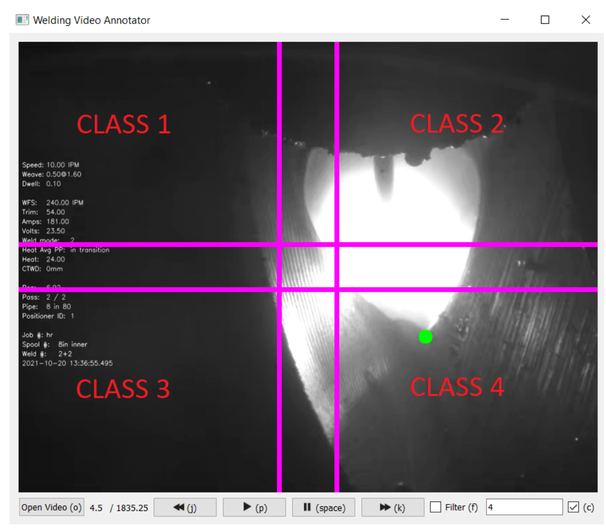


Fig. 3: The multi-class application. The interface divides the frame into 4 main areas and transit areas in between, which represent 4 classes and null one, respectively.

Results

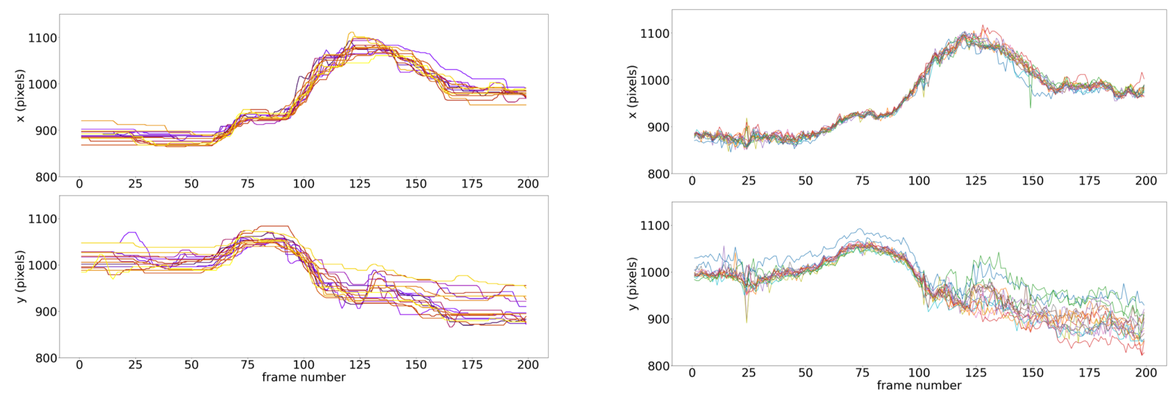


Fig. 4: Comparing the **time series** behavior of the (x,y) location of the cursor labeled by **sixteen annotators** to evaluate labeling consistency.

Metrics	Our interface	IVAT
ICC3k for x location of annotated point	0.99099	0.99929
ICC3k for y location of annotated point	0.98141	0.99370
average annotation time (s)	41.58	868.14

Table 1: Comparing the Intraclass Correlation Coefficient and the average annotation time for 16 annotators using our interface and IVAT interface

ICC value	Reliability level
0 to 0.5	poor
0.5 to 0.75	moderate
0.75 to 0.9	good
0.9 to 1	excellent

Table 2: Level of reliability based on Intraclass Correlation Coefficient