An Adaptable and Extensible Library for Annotation of Video
Open Software Science Initiative Proposal

Kristin Branson, Mayank Kabra, Iljung Kwak

November 24, 2021

Video is captured in a wide variety of biology applications, including cell and animal behavior. Modern analysis of these data requires machine learning, used in problems such as cell or animal tracking and pose recognition, cell division detection, and behavior classification. As good training data is essential to the success of practical applications of machine learning, we need efficient methods for data annotation and visualization. We've been developing such applications for over about 15 years [1, 2, 3], and, for each new application we have worked on, we have had to write new, from-scratch interfaces. While each application has had differences, there are many commonalities that a well-designed library could exploit. Even more problematic is the fact that, throughout computer vision, we often artificially constrain ourselves to inference and learning over single time points because access to time series annotations is unavailable.

We propose developing a general-purpose, extensible library for visualizing and annotating video data. This would enable new machine learning applications to biology time series data, as it will allow developers to efficiently collect annotations. Our hope is that they/we could quickly and easily incorporate this library into a new application, and tailor it to the details of the application, allowing for efficient labeling.

There are a number of pieces of related software, but none that provide an adaptable library for video annotation, management, and visualization that can be used for multiple applications. However, all of these pieces of software have code, algorithms, or principles that we hope to draw from. Efforts from the computer vision community [4, 5, 6] are not adaptable, and are not well-designed for video, particularly video from biology applications [7, 8]. However, they provide great resources on principles related to how to annotate individual images, and how to be used for crowdsourcing annotations. Efforts from the computational biology community [9, 10, 11, 12, 13, 14, 15, 16] either are not well-suited for annotation of machine learning training data, or are not adaptable to time series data, particularly of behaving animals. However, they provide great resources for efficient visualization of large biology data sets that we hope to draw from.

We thus see that there is a need for the development of an adaptable library for image time series annotation for machine learning tasks. Our goal is for this library to be something that can be easily incorporated into new applications, and extended for efficient annotation in specific applications. Our proposal is to develop such a library, with APT and JAABA as initial, concrete applications, with extensibility of common components in mind. Common components we have identified include:

- Video reading and management:
  - Consistently and efficiently reading videos from common codecs and formats.
  - Caching or pre-fetching video frames for fast browsing.
  - Managing projects with linked videos that can be moved from between file systems, including checks for changes/missing videos, macros for translating between file systems/OSes.
- Annotation:
  - Tools for annotating and efficiently storing annotations (scribbling and polygons for segmentation, bounding boxes/shapes for detection, for points for pose recognition, curves for boundary detection, tags for behavior classification).
  - Associating annotations with target identities across frames.
  - Iterative prediction and annotation, and manual correction of predictions.
- Visualization of predictions and intermediate processing steps, or other information to better understand the how the algorithm is working.
- Visualization for comparison of annotations from different annotators or predictions from different

algorithms.
- Multi-stage prediction, e.g. object detection followed by pose recognition/semantic segmentation.
- Efficient navigation, including identification of frames of interest, and clickable timelines showing annotations, predictions, and other time series data.
- Visualization and annotation of multi-camera, multi-view video.

While there are many commonalities, there are also many differences, for instance what is annotated, or how to navigate between frames. We anticipate differences, and adaptability, to be handled through definition of child classes.

One possibility would be to develop this library on top of napari, given its easy integration into Python applications, and the amount of resources devoted to its development. We also would like to explore the alternative of a Javascript (or React) based library that runs in a browser, given the amount of resources to draw from for web applications. We would like to work with Scientific Computing and the software developer recruited for this project to specify the scope, language, and platform for this library.

To ensure that the library will be adaptable, we hope to also explore its use as a component of our ongoing development of an ambitious system allowing 3D-estimation of animal joint locations via mesh models of appearance [17, 18, 19]. We hope to work with other researchers at and outside of Janelia to ensure that the library would be useful beyond our own applications. In particular, we hope to work with Carsen Stringer, Jan Funke, and David Van Valen's groups to ensure that it is useful for microscopy applications. A possible major extension related to this would be annotation of 4D time series data.

We are guessing this effort would require 6 months of developer time for a basic prototype implementing some of the most important features described above. The developer would need to have the wisdom necessary for developing good user interfaces, as well as the ability to develop adaptable code. They should be fluent in Python (and possibly Javascript if we choose a Javascript interface). They must be good at communicating and brainstorming ideas. We are very happy to work closely with the engineer recruited for this project on a day-to-day basis. The three of us would also like to contribute to the coding and design effort ourselves, in addition to testing and integration with applications in our lab.

Documentation necessary would be how to include this library in new applications, with demo examples, and how to extend the library through object inheritance. Tools for installation of requirements would also be necessary, e.g. Conda, Docker, Singularity. Documentation and tutorials for users of specific applications would also be necessary, but would be handled by our lab. It would be great to have a hackathon for adapting this library to new applications, both to facilitate its use and better understand commonalities and differences between applications.

For successful applications of this library, e.g. APT, we could imagine having virtual bootcamps for biologists on how to use the application. Summer students at Janelia have successfully used previous versions of APT for training trackers, and we envision that APT with this user-friendly interface could be a great way to introduce students to computer science and machine learning. Data sets constructed using this interface could be great for teaching machine learning, e.g. such as a notebook teaching the basics of training convolutional networks in PyTorch [20].

References

[1] Kristin Branson et al. "High-throughput ethomics in large groups of Drosophila". In: *Nature methods* 6.6 (2009), pp. 451–457.

[2] Mayank Kabra et al. "JAABA: interactive machine learning for automatic annotation of animal behavior". In: *Nature methods* 10.1 (2013), pp. 64–67.

[3] Mayank Kabra et al. *APT: The Animal Part Tracker*. Version commit 3639d235. Nov. 23, 2021. URL: `http://kristinbranson.github.io/APT/`.

[4] Intel. *CVAT: Computer Vision Annotation Tool*. `https://openvinotoolkit.github.io/cvat/docs/`. Accessed: 2021-11-24.

[5] Abhishek Dutta, Ankush Gupta, and Andrew Zisserman. *VIA: VGG Image Annotator*. `https://www.robots.ox.ac.uk/~vgg/software/via/`. Accessed: 2021-11-24.

[6] Microsoft. *VoTT: Visual Object Tagging Tool*. https://github.com/microsoft/VoTT. Accessed: 2021-11-24.

[7] Christoph Sager, Christian Janiesch, and Patrick Zschech. “A survey of image labelling for computer vision applications”. In: *Journal of Business Analytics* (2021), pp. 1–20.

[8] Eshan Gaur, Vikas Saxena, and Sandeep K Singh. “Video annotation tools: A Review”. In: *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*. IEEE. 2018, pp. 911–914.

[9] Johannes Schindelin et al. “Fiji: an open-source platform for biological-image analysis”. In: *Nature methods* 9.7 (2012), pp. 676–682.

[10] Christoph Sommer et al. “Ilastik: Interactive learning and segmentation toolkit”. In: *2011 IEEE international symposium on biomedical imaging: From nano to macro*. IEEE. 2011, pp. 230–233.

[11] Claire McQuin et al. “CellProfiler 3.0: Next-generation image processing for biology”. In: *PLoS biology* 16.7 (2018), e2005970.

[12] Nicholas Sofroniew et al. *napari/napari: 0.4.12rc2*. Version v0.4.12rc2. Oct. 2021. DOI: 10.5281/zenodo.5587893. URL: https://doi.org/10.5281/zenodo.5587893.

[13] Tobias Pietzsch et al. “BigDataViewer: visualization and processing for large image data sets”. In: *Nature methods* 12.6 (2015), pp. 481–483.

[14] Eric Perlman. “Visualizing and interacting with large imaging data”. In: *Microscopy and Microanalysis* 25.S2 (2019), pp. 1374–1375.

[15] Stephan Saalfeld et al. “CATMAID: collaborative annotation toolkit for massive amounts of image data”. In: *Bioinformatics* 25.15 (2009), pp. 1984–1986.

[16] Saalfeld Lab. *Paintera*. https://github.com/saalfeldlab/paintera.

[17] Jathushan Rajasegaran et al. “Tracking People with 3D Representations”. In: *arXiv preprint arXiv:2111.07868* (2021).

[18] Marc Badger et al. “3D Bird Reconstruction: a Dataset, Model, and Shape Recovery from a Single View”. In: *ECCV*. 2020.

[19] Silvia Zuffi et al. “3D Menagerie: Modeling the 3D Shape and Pose of Animals”. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. July 2017.

[20] Kristin Branson. https://colab.research.google.com/drive/1SLgOHcgo1bfMDx5wlnLqm05AZe6aVG0l?usp=sharing. 2021.