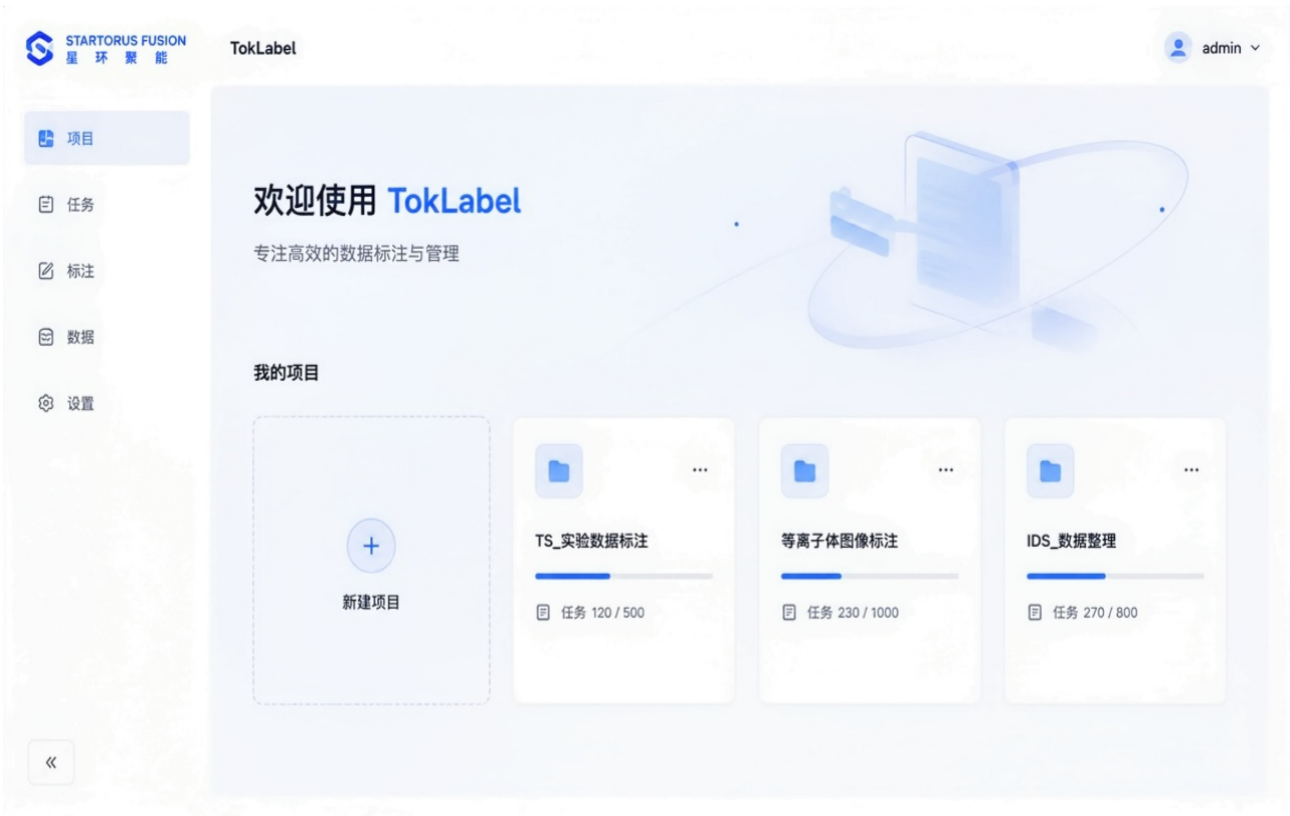
TokLabel

Introduction

TokLabel (Tokamak Label Suite) is an open-source fusion data labeling platform co-developed by Startorus Fusion and Tsinghua University, built on the open-source Label Studio and released under the Apache 2.0 license. Fusion AI researchers spend roughly 70% of their time on data preparation, and TokLabel targets exactly this stage.

It turns raw discharge data into labeled datasets ready for fusion AI. Researchers simply input a shot number to automatically retrieve data and generate labeling tasks — labeling feature moments such as breakdown, disruption, and end on one-dimensional time series, and configuration information such as the last closed flux surface on two-dimensional visible-light images. Labeling results are stored by shot number and directly consumed by downstream model training such as disruption prediction, discharge feature recognition, and configuration recognition, saving roughly 70% of data-preparation time.



Features

1. Fusion-specific: data is organized by shot. Dozens of discharge time-series signals and visible-light image frames all belong to one shot, and both data types are labeled together. General-purpose labeling tools treat data as independent samples and do not recognize this multimodal structure; TokLabel aligns data by shot number from the ground up.
2. One YAML file configures one project: project definition, channel computation, and data filtering rules all live in a single config file, and data retrieval, project creation, and storage setup are completed in one pass. To switch signals or channels, just edit the config — no coding or manual task creation required.
3. AI pre-labeling loop: the model first predicts feature moments such as breakdown, disruption, and end, leaving humans only to verify and correct. Corrections flow back into training, so the model gets more accurate with use and saves more labor as the labeling scale grows.

4. Labeling results stored by shot number: exported annotations go into PostgreSQL and are retrieved and reused by shot number, so a shot never needs to be prepared twice and feeds directly into downstream model training.
5. Open-source, extensible, with access control: built on Label Studio and open-sourced under the Apache 2.0 license, it fills the access-control gap missing from the open-source edition and supports multi-user collaborative labeling.

Parameters

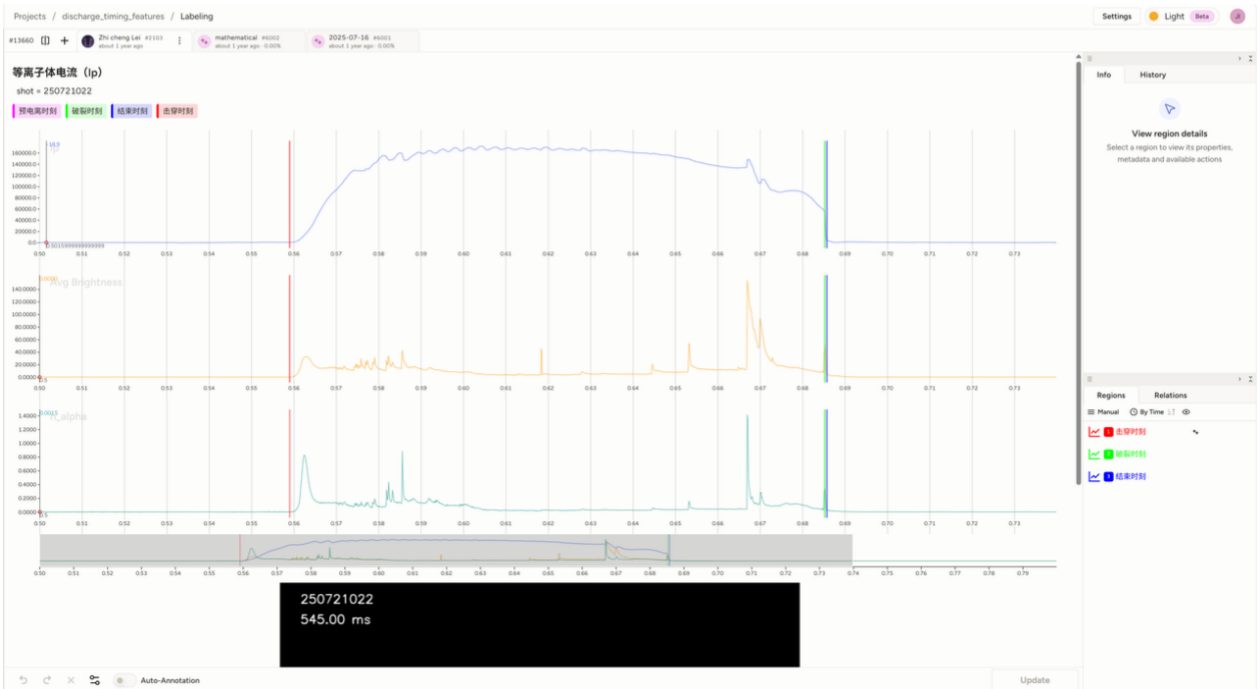
Parameter	Description
Labeling types	1D data (time series), 2D data (images)
Automatic data import	Connects to MDSplus / PostgreSQL; input a shot number to automatically retrieve data and generate labeling tasks
AI pre-labeling	The model predicts breakdown, disruption, and end feature moments; humans verify and correct, and corrections flow back into training
Access control	Supports multi-user collaboration, filling the gap in the open-source edition of Label Studio
Storage & retrieval	PostgreSQL + Redis; exported annotations are stored and retrieved/reused by shot number

Applications

Disruption prediction

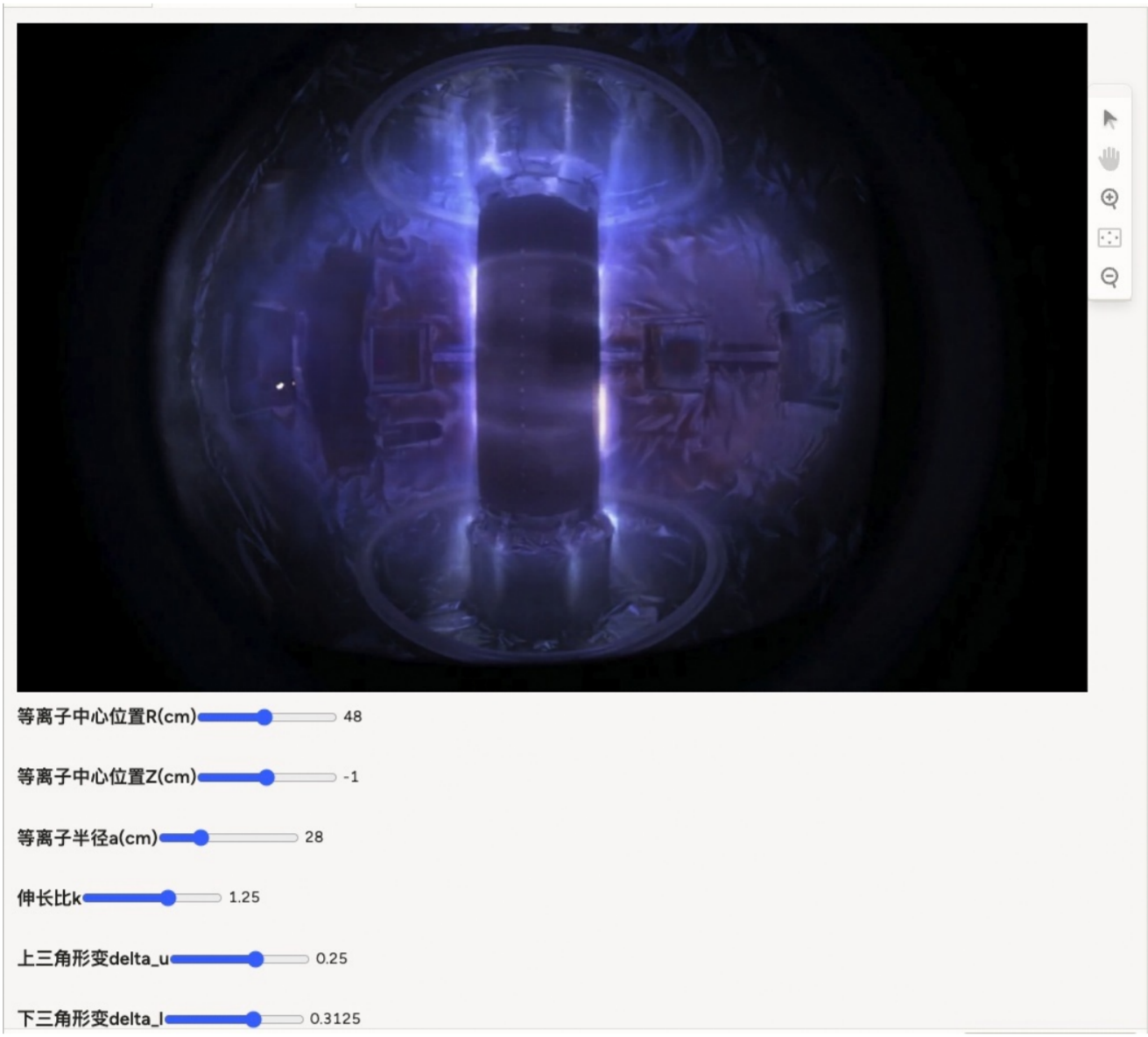
Train a model on disruption-labeled data to determine whether a shot is disruptive

Discharge feature recognition



Use breakdown / disruption / end time labels to train a model to automatically recognize key discharge events

Configuration recognition



Use last-closed-flux-surface labels on visible-light images to train a configuration recognition model