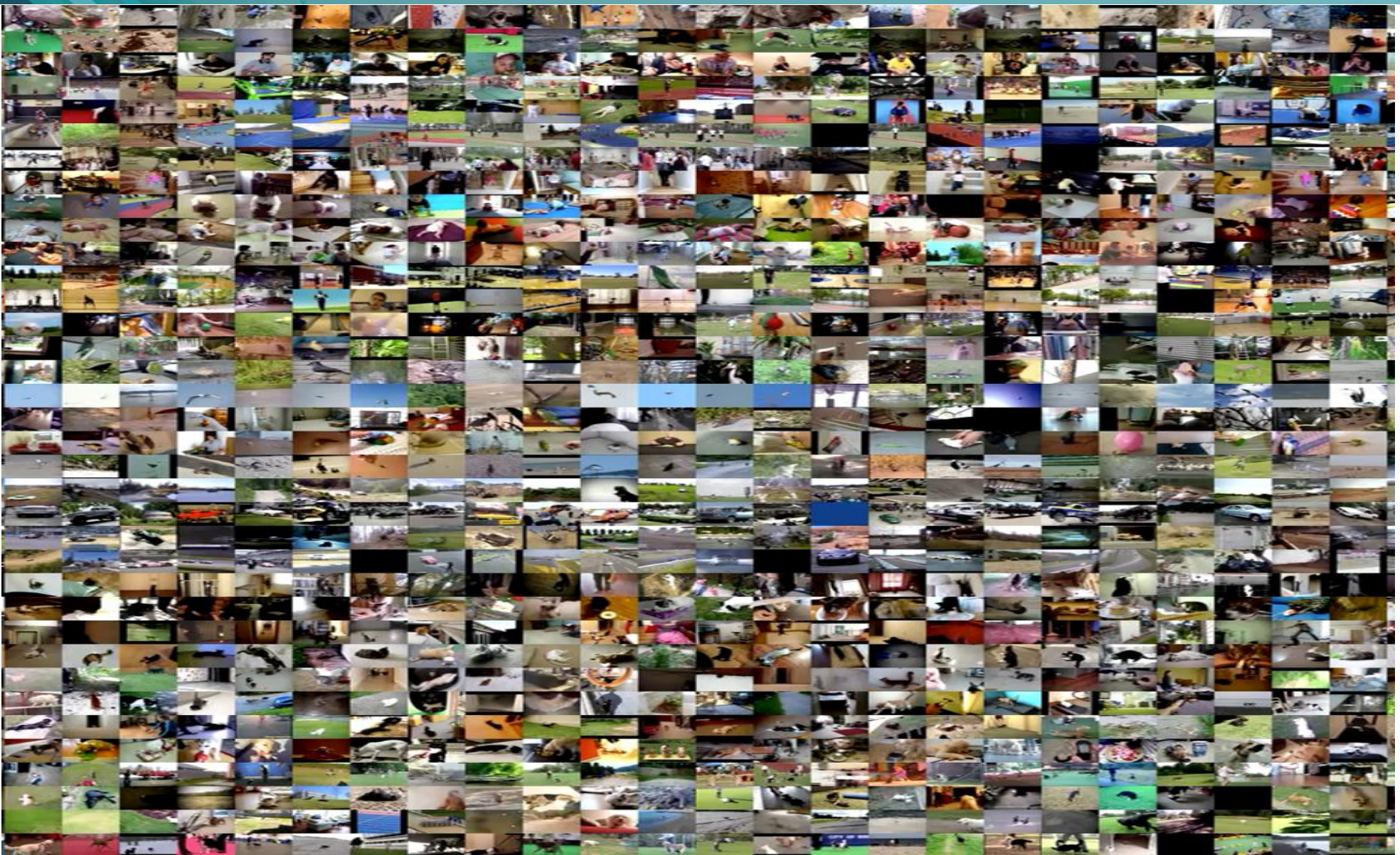
Creating Better Datasets For Training More Robust Models in FiftyOne

Jason Corso, PhD
CEO
Voxel51
https://fiftyone.ai

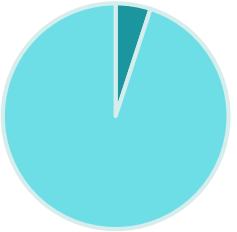
Data Eats Models for Lunch



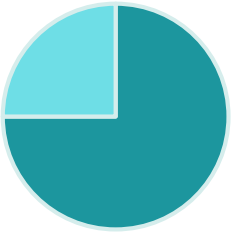
Model work

Data work

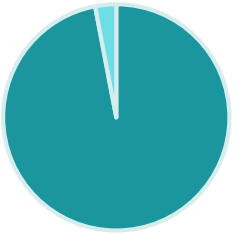
2015



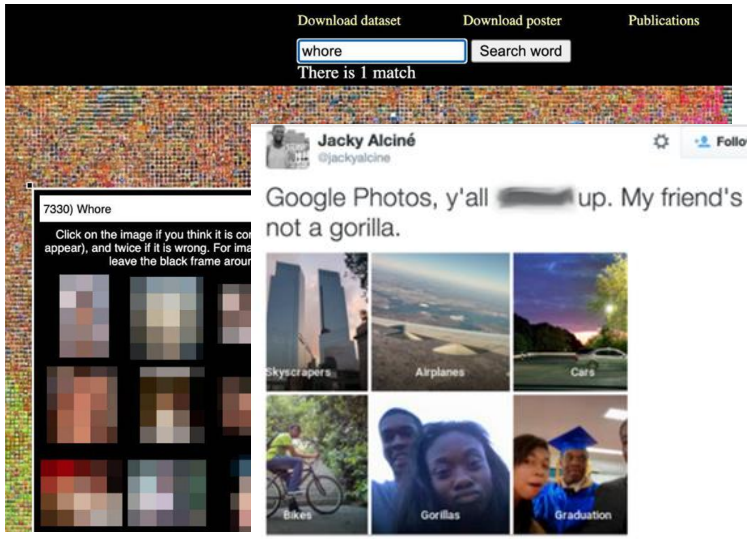
2021



2030



Poor Data? Big Problems



MIT's 80 Million Tiny Images Dataset has "Significant Ethical Issues"
Google Photos Racial Bias (WNY Studios)

Model Bias Issues

Two killed in Tesla Autopilot Crash with Parked Vehicle (NYT 2021/08/17)

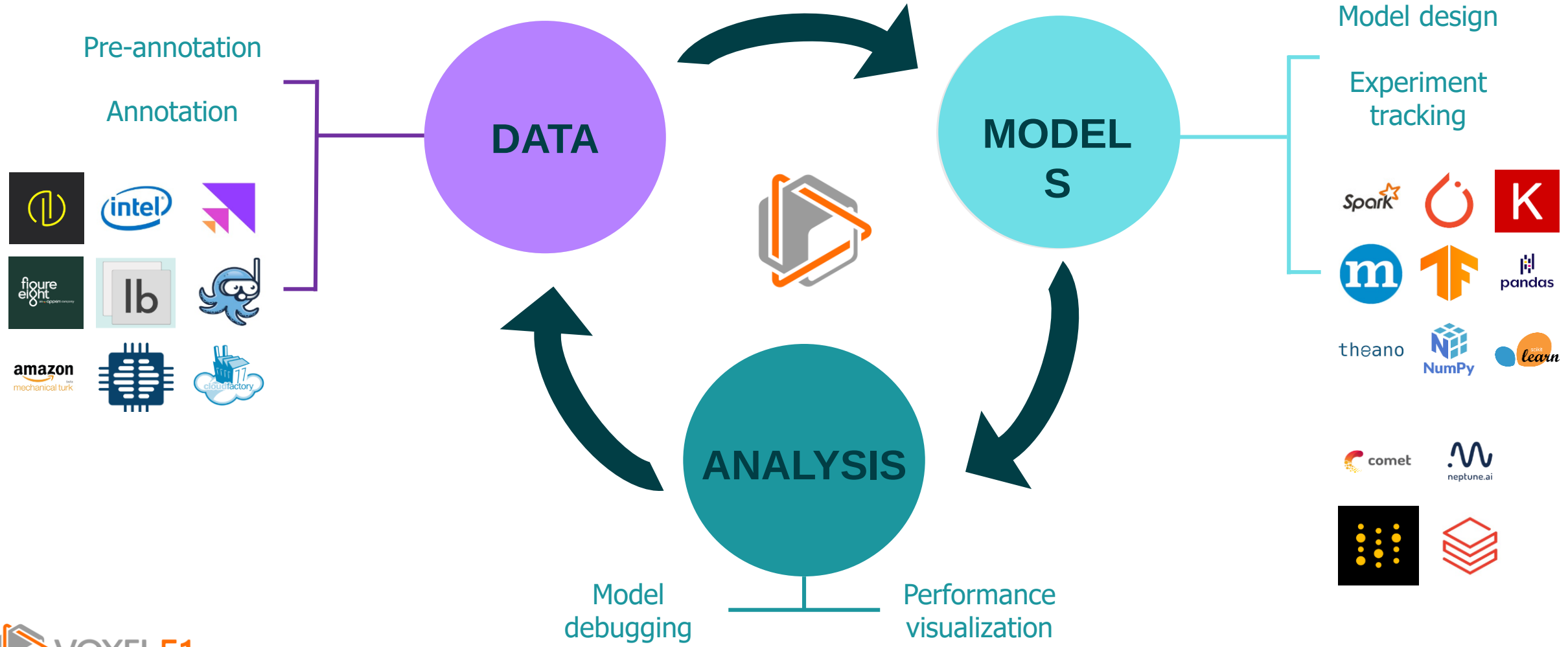
Lethal Physical Danger

Even Google's Open Images dataset has quality problems (blog post)

30% Reduction in Model Performance

ML Engineers Need Better Tools

To Bring Better Models to Production

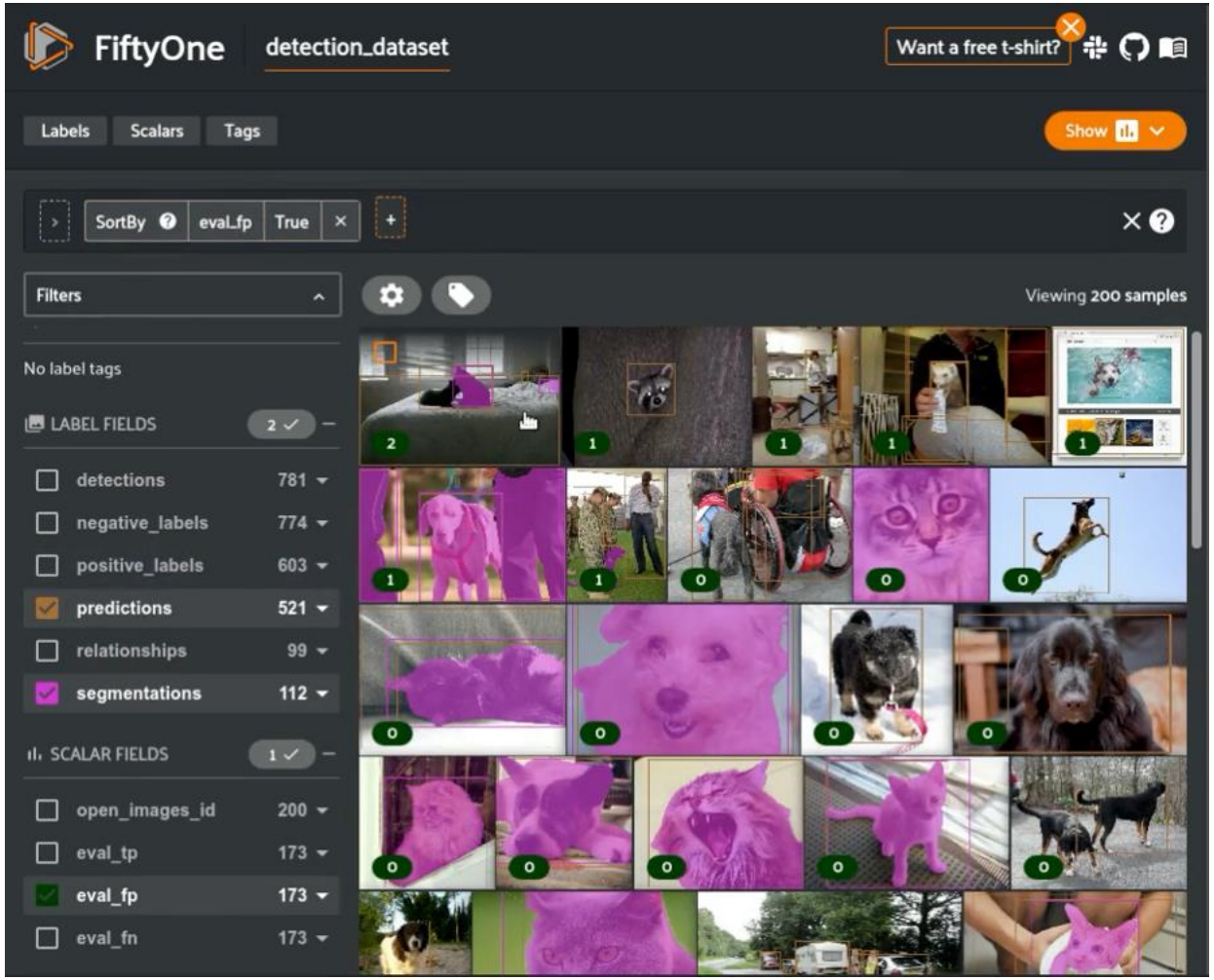


FiftyOne

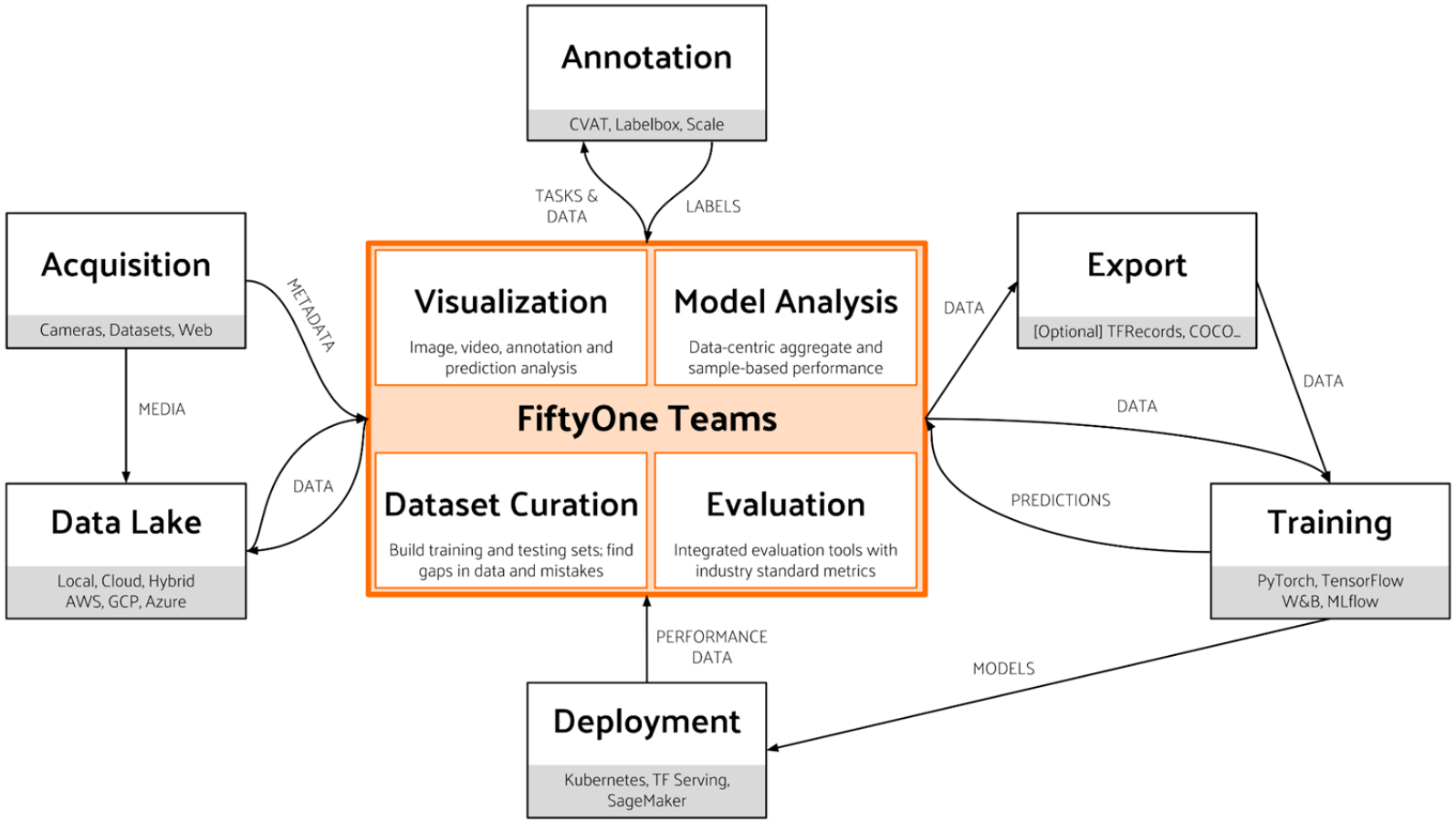
The only open-core data-centric ML solution

- Curate, visualize and analyze datasets
- Streamline annotation workflows
- Find and fix labeling mistakes
- Identify and correct model failures
- And dozens more workflows...

> FiftyOne has helped us rapidly test hypotheses, gain insights about our data, and understand both quantitatively and qualitatively where our models fall short.
> – Chris H, Staff Embedded Data Scientist at Vivint



How FiftyOne Fits

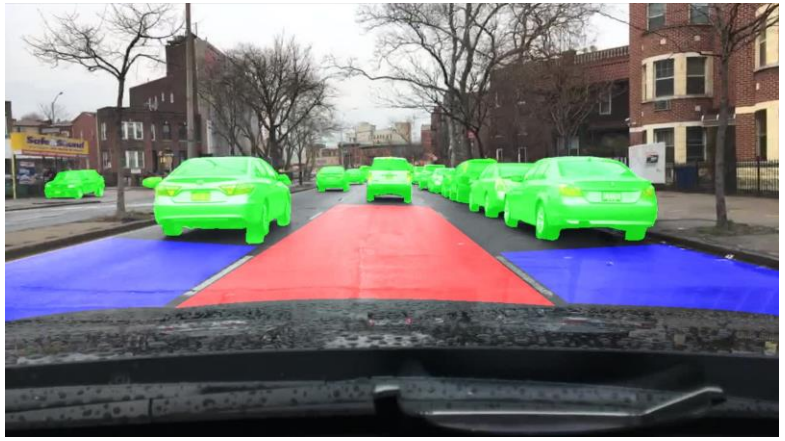


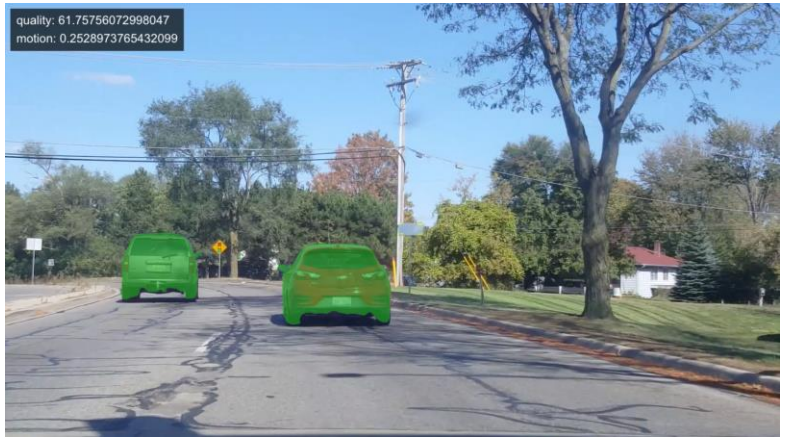
FiftyOne is the single source of truth for ML data on our team

– Mohammed A, Bosch

Supports All Popular Computer Vision Tasks

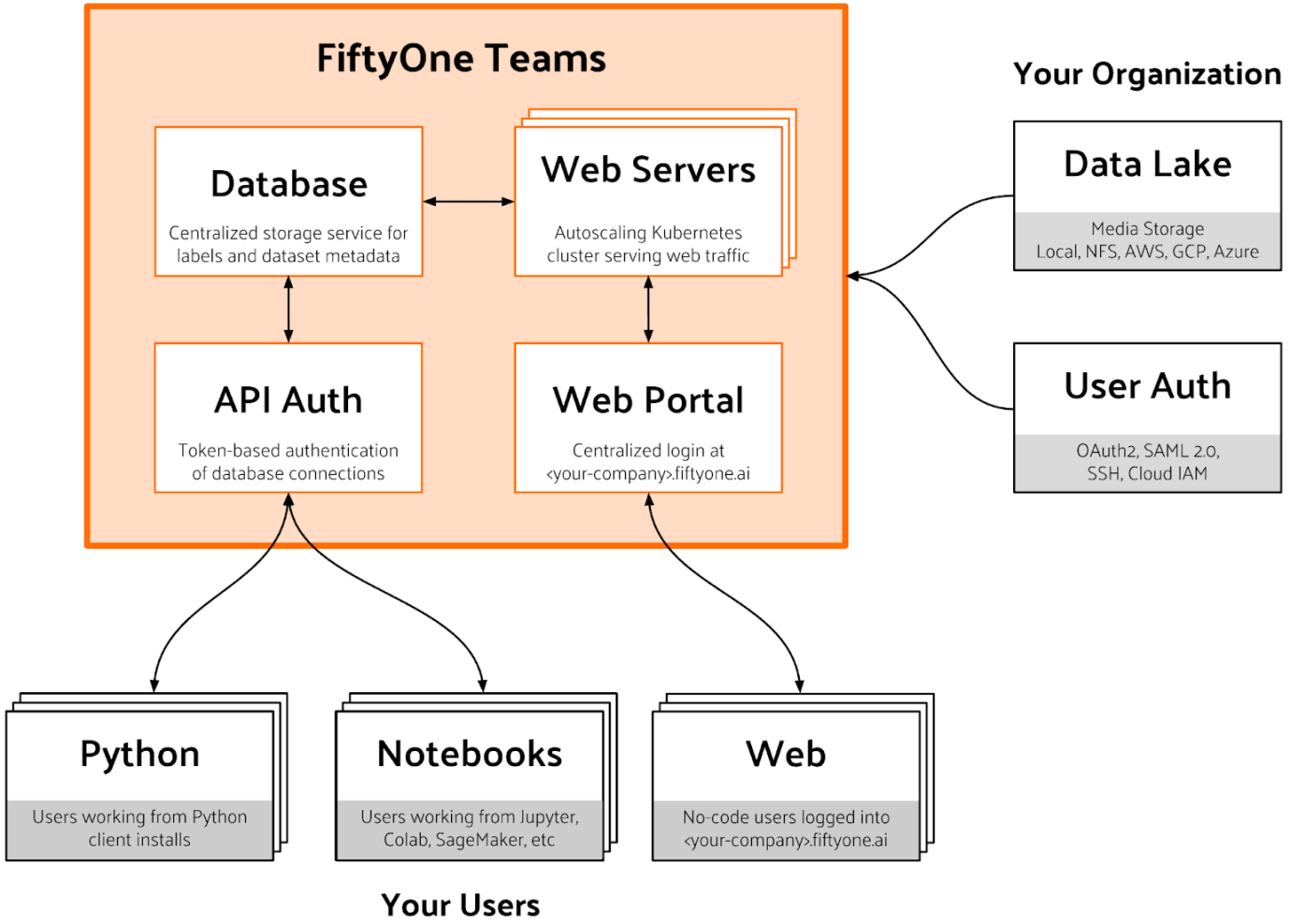
- Classification
- Detection
- Instance segmentation
- Semantic segmentation
- Polygons and polylines
- Keypoints
- **Image and video data**





Data Model and System Layout Brief

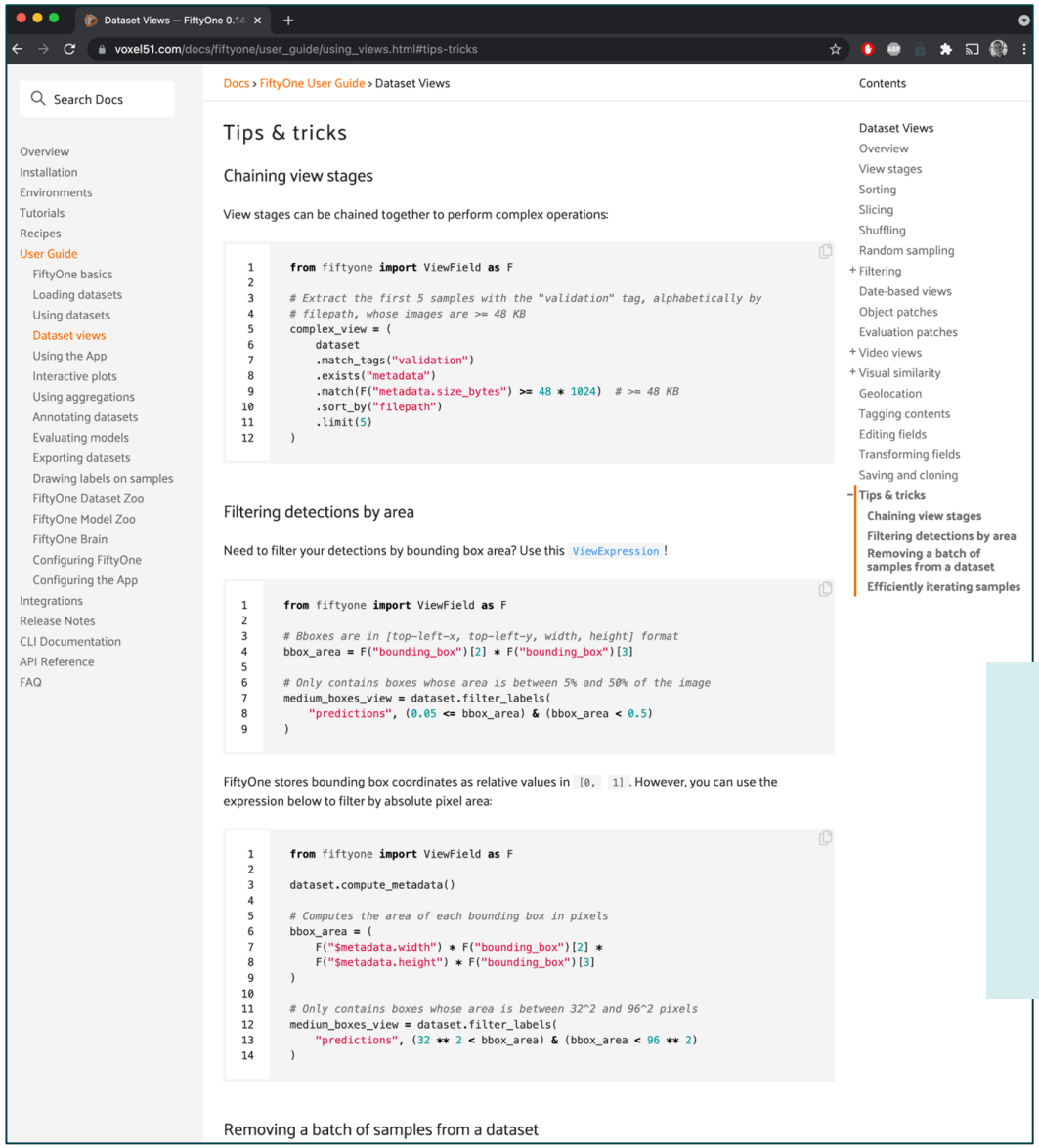
- FiftyOne has two key data *types*
  - Dataset
  - View
- Open-source → individuals
- **FiftyOne Teams → groups of users**
  - Same data model
  - Adds ”professional” features
    - Shared, cloud-backed data
    - Web portal
    - Versioning



Intuitive and Extensible API

FiftyOne integrates into existing workflows with a few lines of code

- Local, remote, and cloud data
- Handoff between code and App
- Core library is **open source**
- **Integrates with annotation** tools like CVAT, Scale, and Labelbox
- **Integrates with model training** loops in PyTorch, TensorFlow, etc.
- **Integrates with experiment tracking** tools like MLflow and W&B

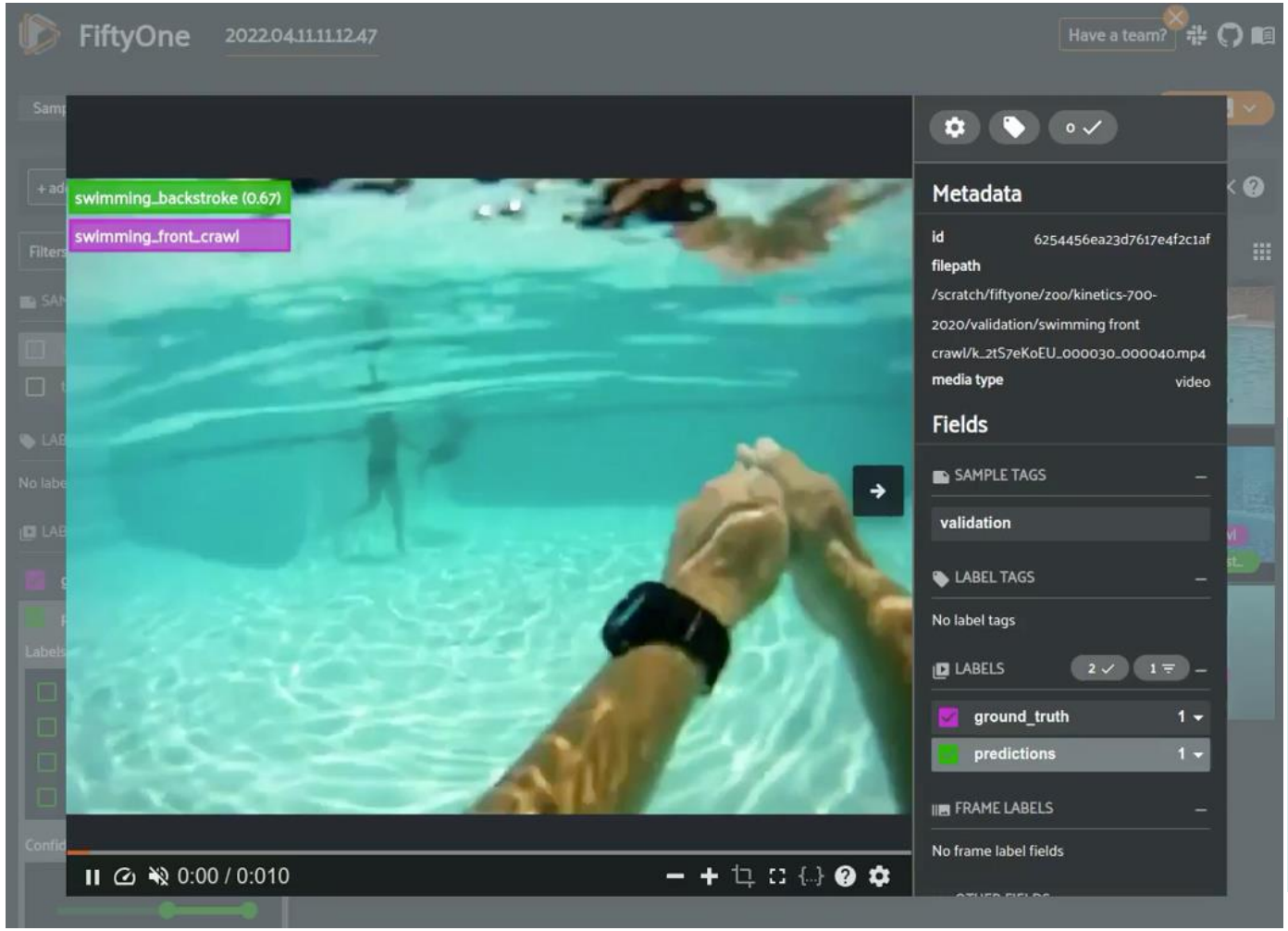


FiftyOne has rich documentation at https://fiftyone.ai

FiftyOne Integrations Enhance Workflows

Open-source nature lets FiftyOne extend to meet any ecosystem.

- For example, easily leverage open-source data
- Open Images, COCO, Activity-Net, Kinetics, ...
- Example here visualizes an annotation mistake in Kinetics
- Recent blog on kinetics

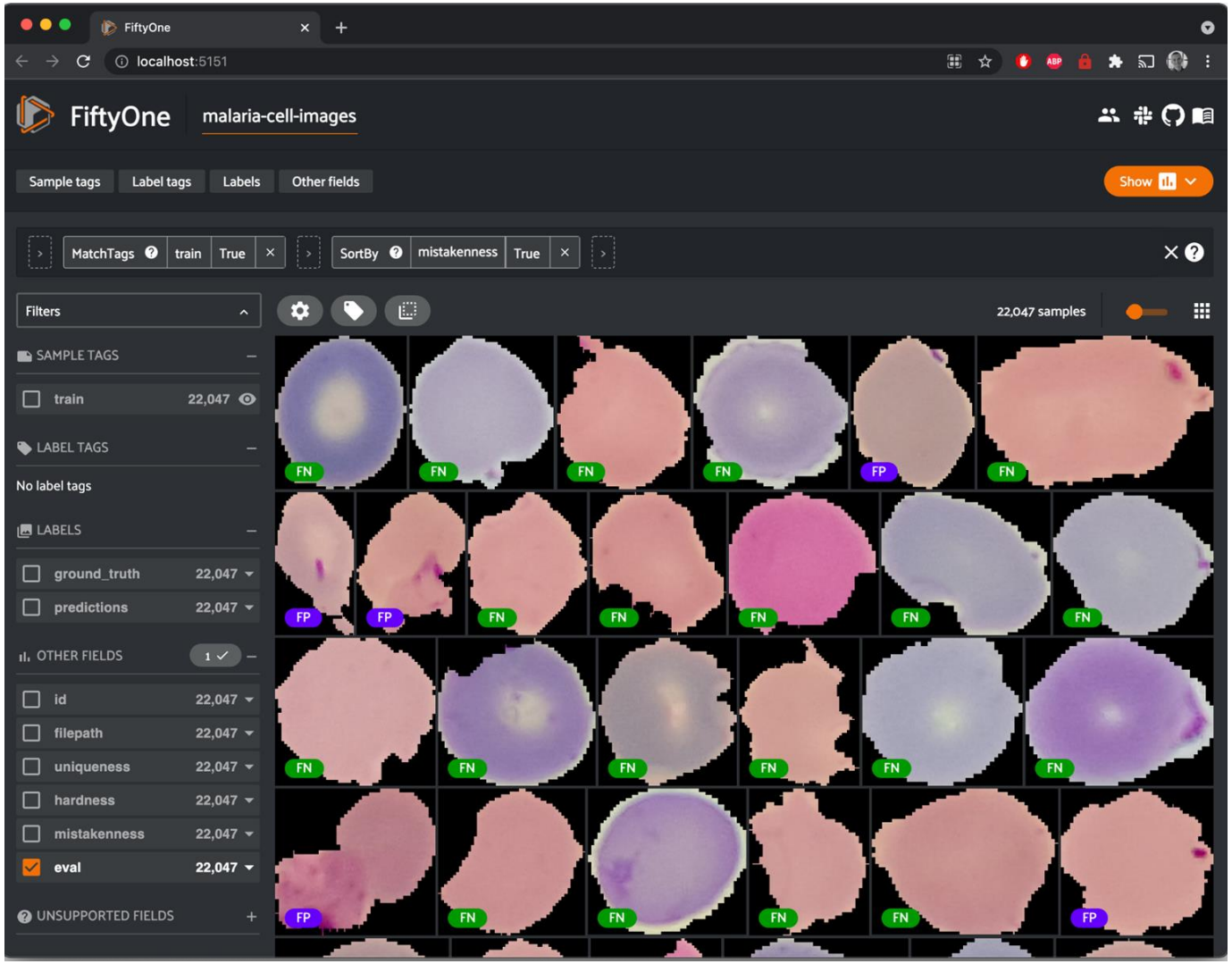


Getting closer to your data means **better datasets** and hence **better models**.

FiftyOne Brain

Automate the data-centric workflows you need to improve your datasets and models

- Identify model failure modes
- Recommend samples for annotation
- Find annotation mistakes
- Embedding-based workflows

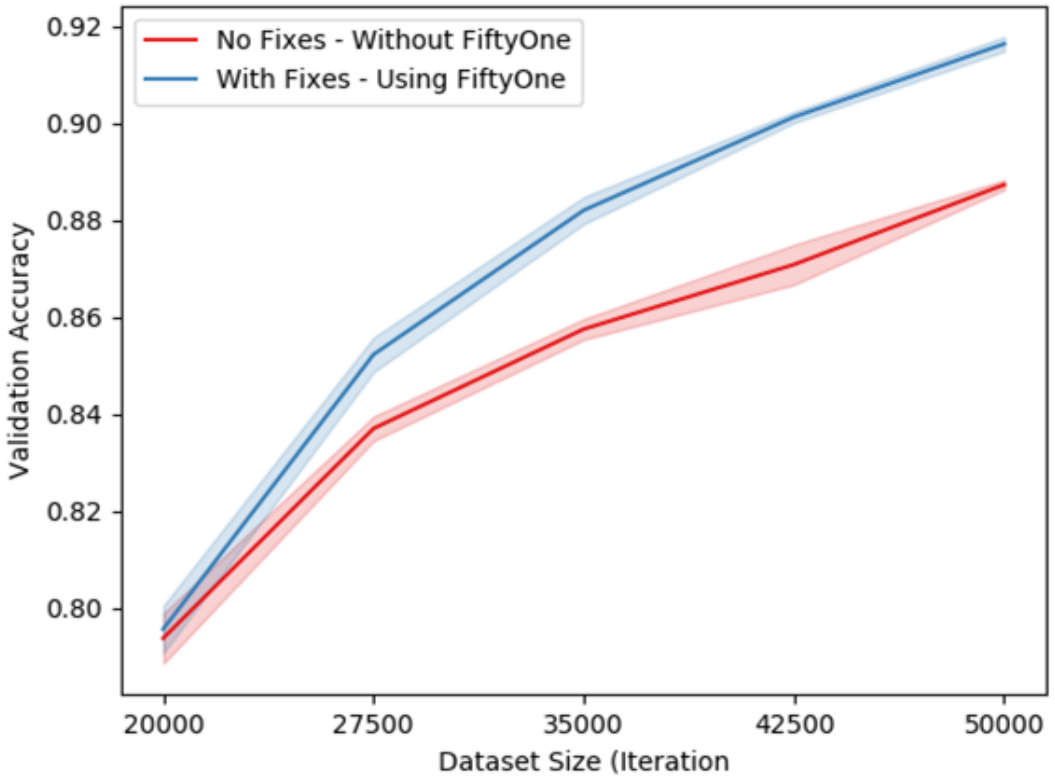


Exploring model predictions interactively in the FiftyOne App with best in class model analysis capabilities like false negative and false positive mining

Better Data, Better Models

FiftyOne **unlocks higher performance from your models** by identifying and removing the flaws in your data

- Example of using FiftyOne to find and fix annotation mistakes vs random fixes
- Real world scenario of iteratively growing your dataset as you improve your model



Using FiftyOne to find and correct labeling mistakes during model training ("with fixes") to accelerate training.

Take Home Message

- Data trumps models in production vision systems
- FiftyOne is the leading open-source software for data-centric machine learning workflows
- We offer both free OSS and commercial Teams-based software for augmenting nearly any CV workflow
- **Better Data; Better Models**

Documentation here!

As easy as
pip install fiftyone

https://fiftyone.ai

Documentation here!

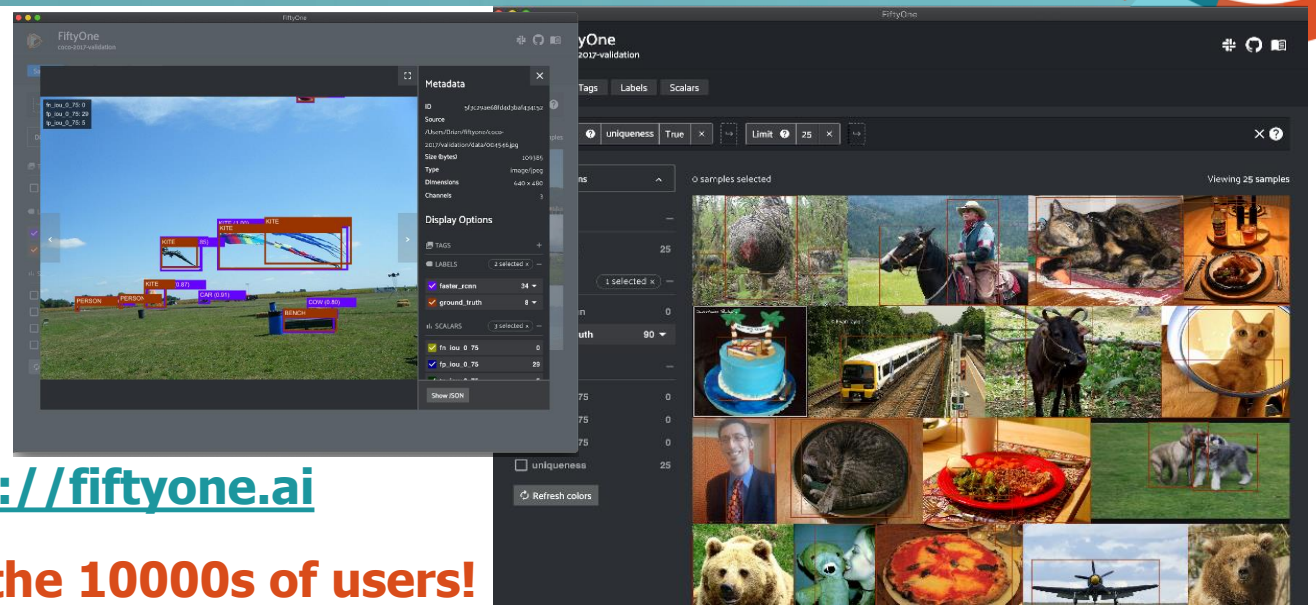
Resources

Next steps

- Like the project?
  Give us a star on GitHub
- Want to get involved?
  Join our Slack community

Light reading

- Overview blog post 1 and 2
- Installation guide
- Documentation
- Tutorials



https://fiftyone.ai

Join the 10000s of users!

Thank you!