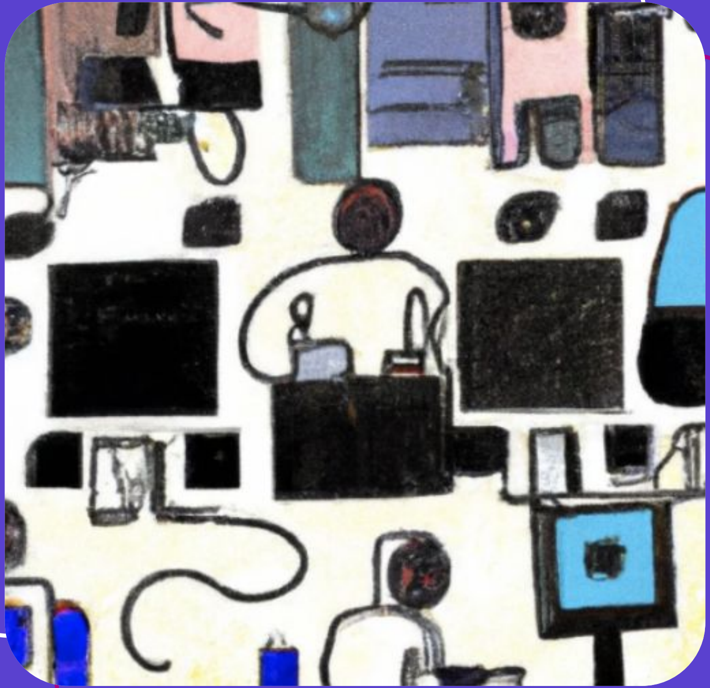
VOXEL WEBINAR

Re-Annotating MS-COCO, an Exploration of Pixel Tolerance

Jerome Pasquero
Eric Zimmermann

June 8, 2023



sama

Agenda

Quick intro to Sama

New trends in the business of generating computer vision training data

Why we re-labelled MS-COCO

Sama-COCO dataset exploration

Quality tolerance experiment

Quick Intro to Sama

sama

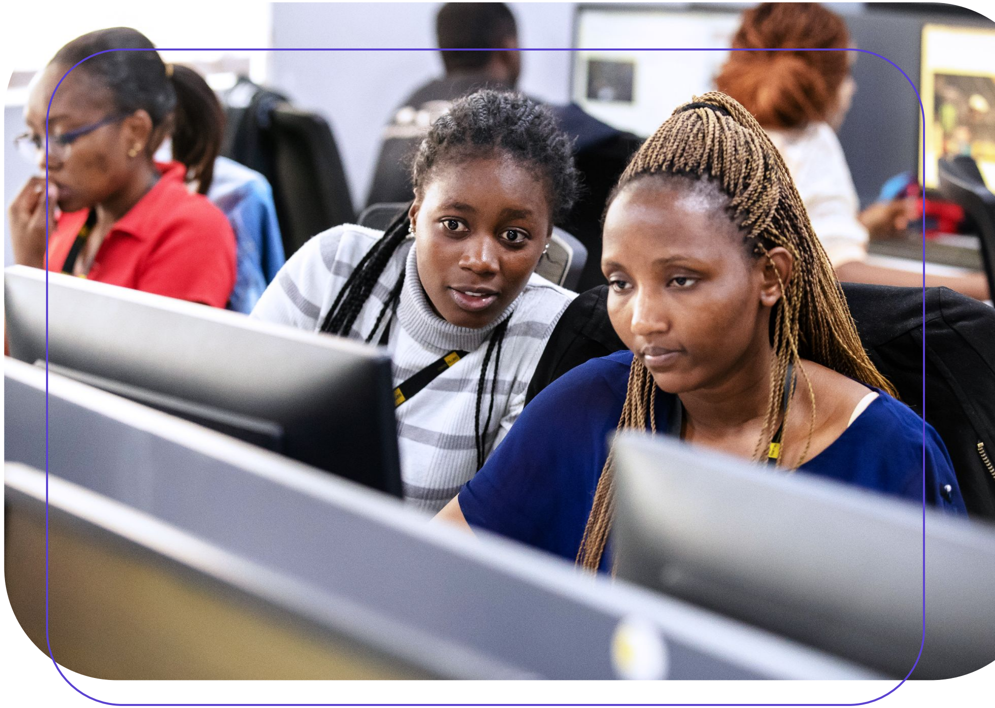
The Ethical AI Supply Chain

65,000+
Impacted since 2008

9,000+
Trained

15,000+
Hired

41,000+
Dependents

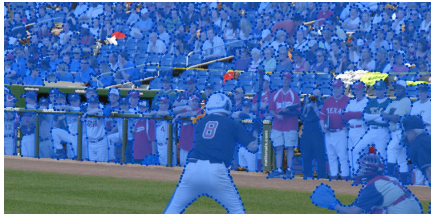


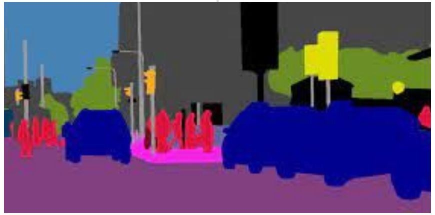
Advocating for an ethical AI supply chain

We advocate for more ethical AI alongside industry leaders and orgs like Partnership on AI and the Haas Center for Equity, Gender and Leadership

Task Complexities

Object detection | **Segmentation** | **Panoptic segmentation** | **3D point cloud** | **Sensor fusion**



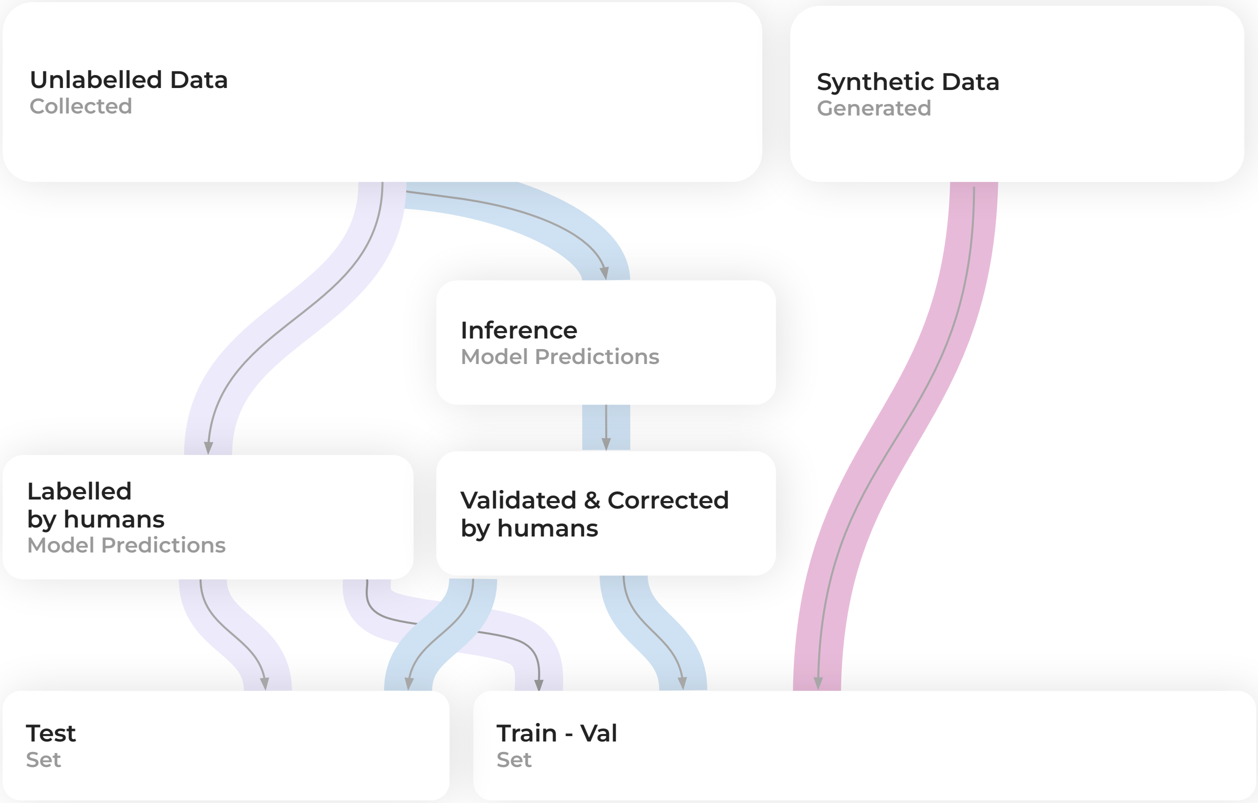


Pedestrian

Motor Cyclist

sama

Sources of Data



sama

New Trends in Computer Vision Training Data

Training Data is Changing

Foundation Models

Large scale models, such as Meta's SAM, are becoming available and can be leveraged for auto-labelling.

Model Maturity

Models in industry are near ready for production. They generate predictions of high quality.

Synthetic Data

High-fidelity synthetic data can be generate at scale and has been shown to help model performance.

Domain Knowledge

The data required for fine tuning models is becoming increasingly specialized.

sama

Training Data Providers Need to Adjust

Smaller Workflows

Workflows with less volume of data to process but with requirements for quicker turnaround times

Increased Training

Workforce needs to be continuously re-trained to meet domain knowledge expectations

Adding Value

More “surgical” approach that includes identifying and generating relevant data

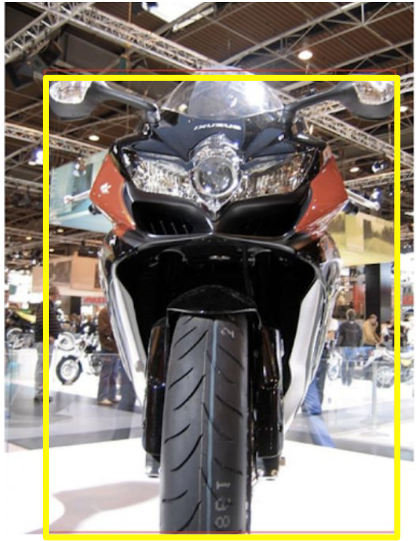
Monitoring

Workflows with pre-annotations in the form of model predictions that need to be validated and corrected.

sama

Why we re-labelled MS-COCO

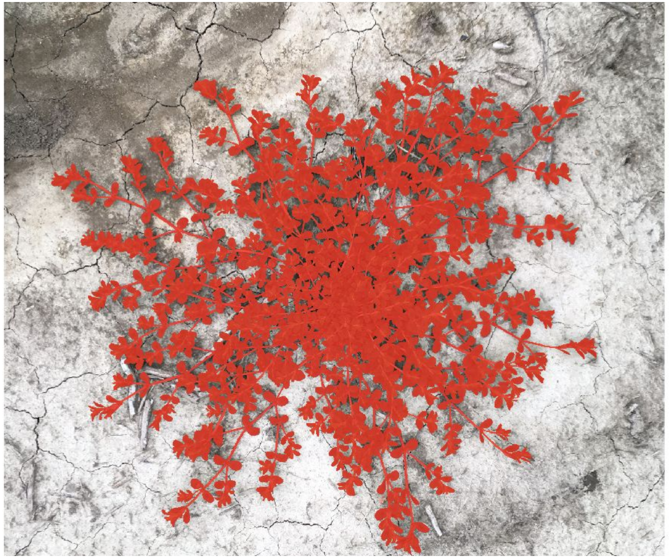
What should manual annotations look like?



PRICE	1x	10x	100x
VALUE	?	?	?

sama

What should manual annotations look like?



Validation Workflows

adding crop pixels

refining an annotation

correcting a model error

sama

Quality Rubric

Penalties are assigned for each type of annotation error.

Omission

Objects to be annotated was left out.

critical
penalty: 100%

Wrong Primary Label

Object main label was misclassified

medium
penalty: 20%

Object tracking

Object left the scene, but came back and was tracked by a different ID.

medium
penalty: 20%

Object Obfuscation

Parts of an object are missing.

medium
penalty: 20%

Inaccurate Annotation

Polygon or bounding box is too tight or too loose.

critical
penalty: 100%

"pixel tolerance. 2 or 3 pixels that appear for more than 1 instance is critical."

Inaccurate Attributes

At least one secondary object attribute is erroneous.

low
penalty: 5%

sama

Our Objectives

Raw dataset for running quality experiments

Illustrative Examples for Clients

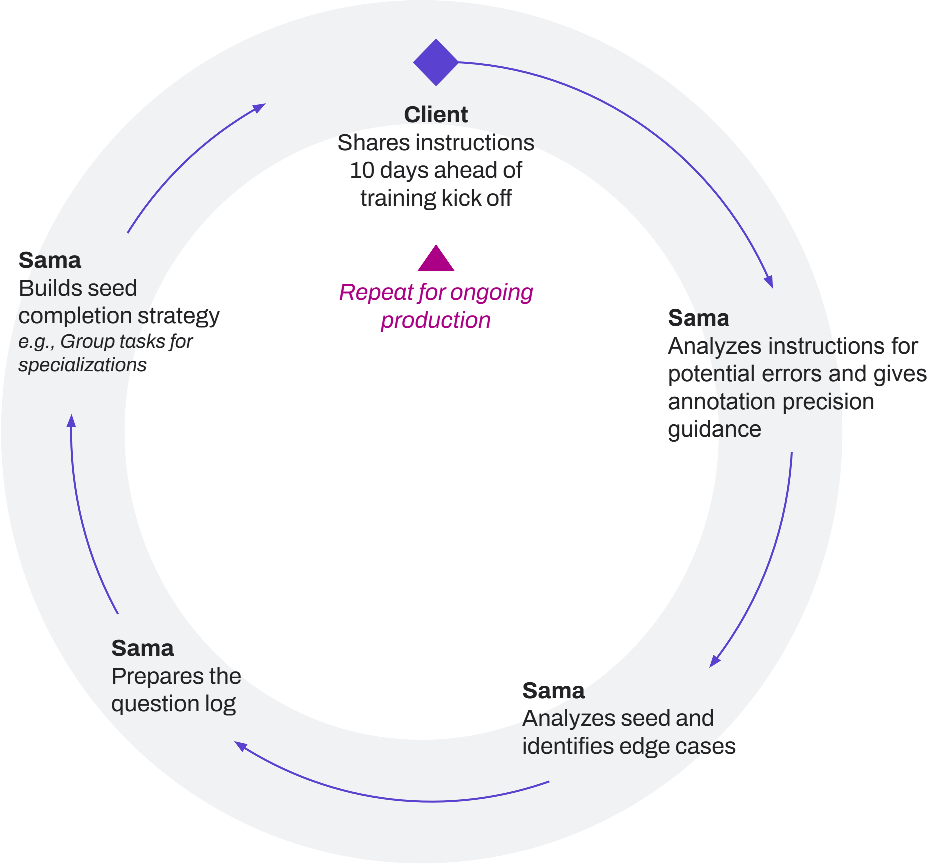
Open discussion on data quality

Contribute to the ML Community

Optimize our annotation processes

Foster collaborations & partnerships

R&D and Product Commissioned Sama-COCO



Simple Instructions...

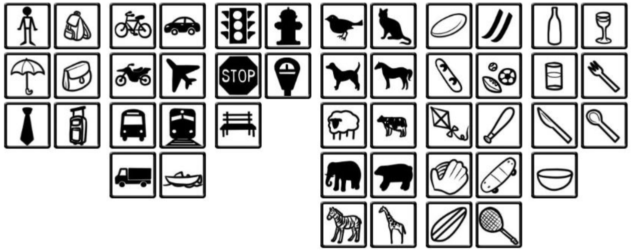
We started with simple instructions:

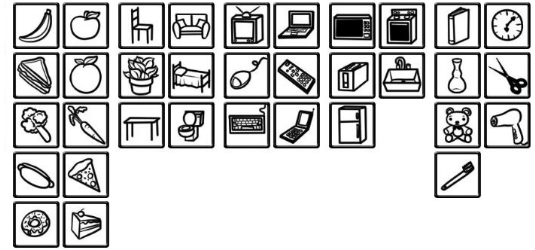
Be very precise and comprehensive when drawing polygons (pixel tolerance of 1 pixel)

Prioritize annotating instances of objects over crowds of objects.

If more than 75 instances of the same class are present, label remaining objects as a crowd.

Ignore small objects under 10 pixels in width or height.





Simple Instructions...

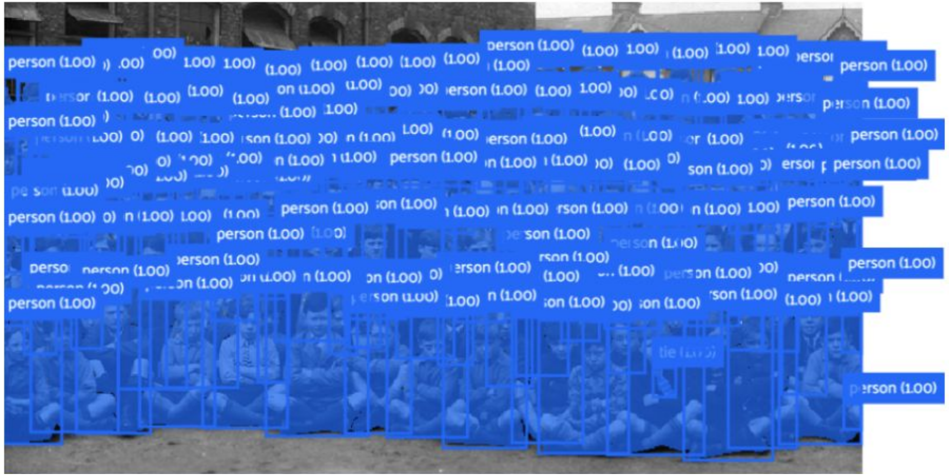
But...

we made adjustments over the course of a few weeks to balance quality and time allocated to labeling

The number of annotating agents fluctuated (from a dozen to a few dozens)

MS-COCO pre-annotations were used, but not always. Small objects (<10^2 pixels) were deleted, but not all.

In some case, agents continued to annotate instances of the same class even after the maximum number of 75 was reached.



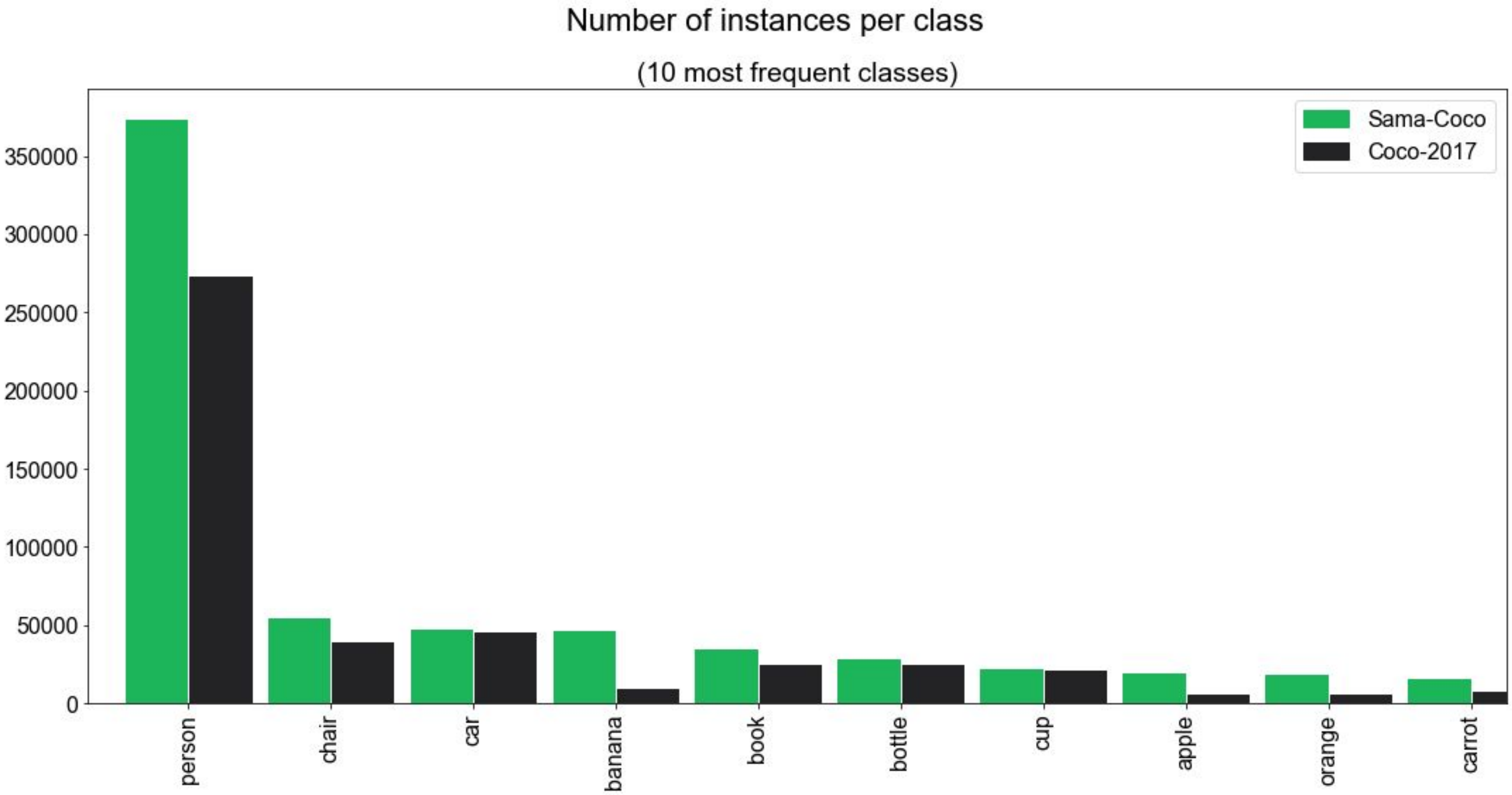
sama

Sama-COCO
Dataset Exploration

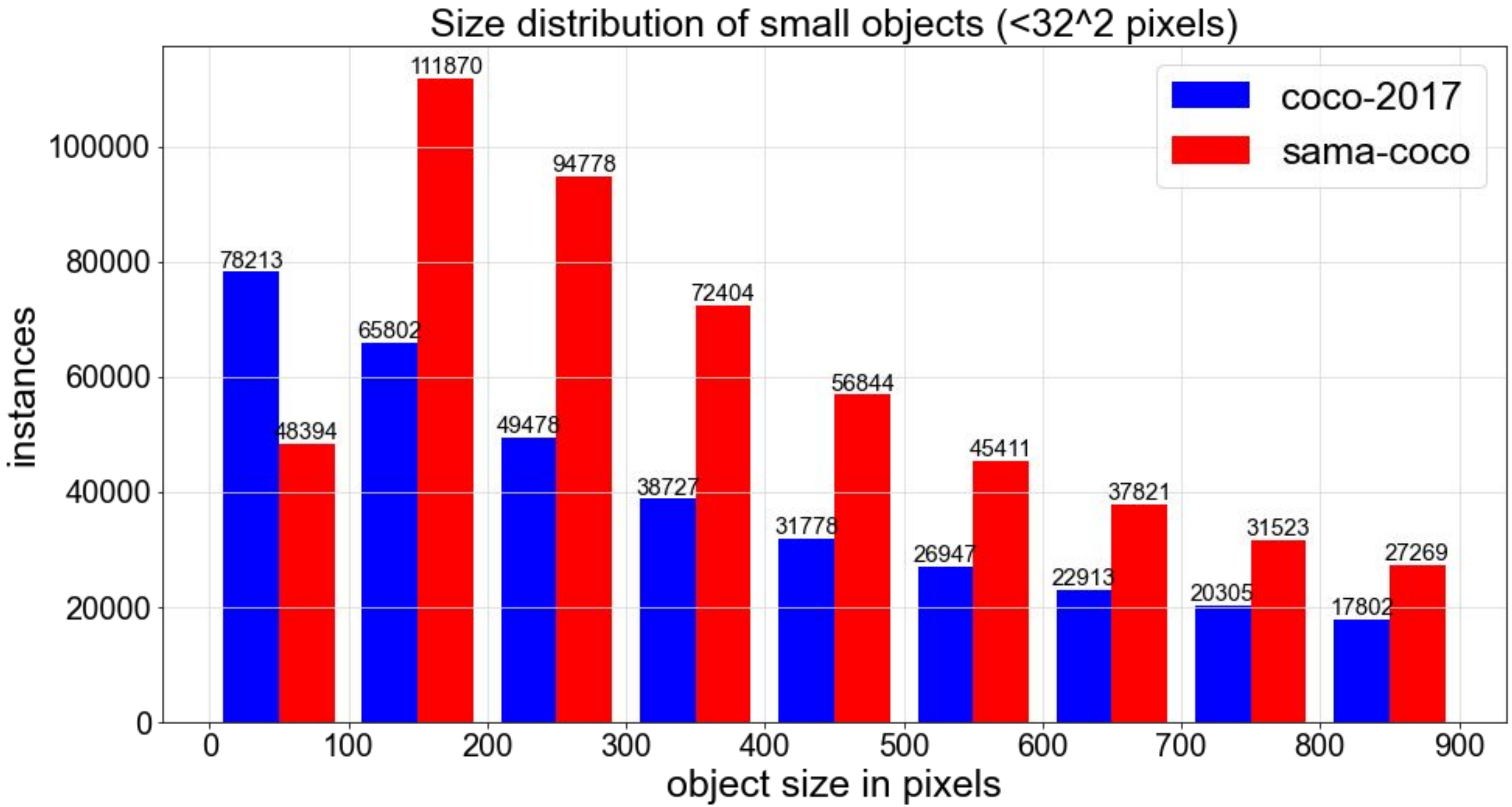
sama

Sama-COCO

	Coco - 2017	Sama-COCO	Difference
Number of images	123 287	123 287	
Number of instances	896 782	1 115 464	218 682 (**x1.24**)
Number of vertices	22 735 106	41 638 434	18 903 328 (**x1.8**)



Number of instances per class
(10 most frequent classes)
Sama-Coco
Coco-2017
350000
300000
250000
200000
150000
100000
50000
0
person
chair
car
banana
book
bottle
cup
apple
orange
carrot


Size distribution of small objects (<32^2 pixels)
coco-2017
sama-coco
instances
object size in pixels
0
20000
40000
60000
80000
100000
0
100
200
300
400
500
600
700
800
900
78213
48394
65802
111870
49478
94778
38727
72404
31778
56844
26947
45411
22913
37821
20305
31523
17802
27269

Ground Truth: **MS-COCO**
Predictions: **Sama-COCO**

	ALL OBJECTS	SMALL OBJECTS	MEDIUM OBJECTS	LARGE_OBJECTS
mAP	0.630	0.406	0.616	0.747
accuracy	0.633	0.299	0.502	0.755
precision	0.75	0.390	0.592	0.872
recall	0.797	0.563	0.767	0.850
fscore	0.775	0.461	0.668	0.861
support	1005999	267715	300290	294155

	ALL OBJECTS	SMALL OBJECTS	MEDIUM OBJECTS	LARGE_OBJECTS
True Positives (TP)	801448	150695	230196	250050
False Positives (FP)	260700	235870	158500	36837
False Negatives (FN)	204551	117020	70094	44105

Quality Tolerance Experiment

Sama-COCO: Quality Tolerance Experiment

Outline

- Defining objectives
- Properties of the intersection over union (IoU) as detection metric
- Impacts of annotation quality quantified by the IoU metric
- Analysis of stylistic annotation differences between MS-COCO and Sama-COCO
- Empirical study on effects of annotation noise and its impact on model performance

We show stylistic differences between datasets and demonstrate that maximum allowable pixel tolerances which preserve performance on a detection task are proportional to object size.

sama

Task Requirements

Detection and Segmentation

How does your application make a decision?

- Fine-grained understanding
  - Pixel level precision
  - VFX photo editing, medical intervention requiring pathology area or boundary
- Coarse understanding:
  - Object level precision
  - Localization, tracking, counting

Better understanding system requirements allows for better specifications of annotation quality

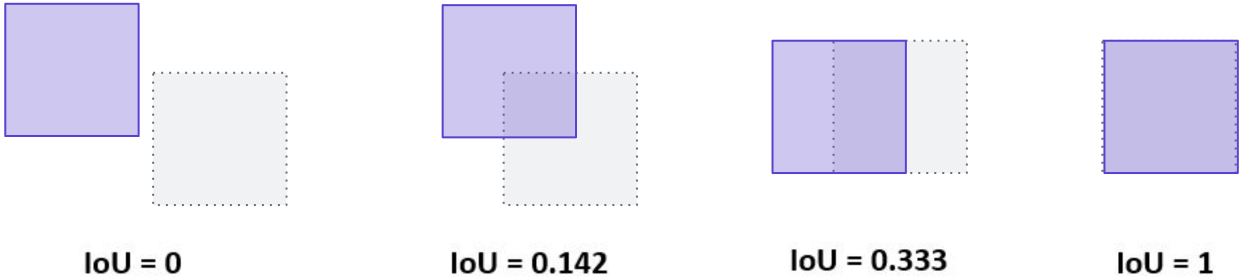
sama

Instance Detection and Segmentation

Metrics and Intersection Over Union

Metric to measure similarity of masks - Intersection over Union (IoU)

- IoU is the criteria used to determine what is considered to be a detection
- Measured by the intersection of two masks and normalized by the union



Intersection over Union Tolerances

Understanding IoU

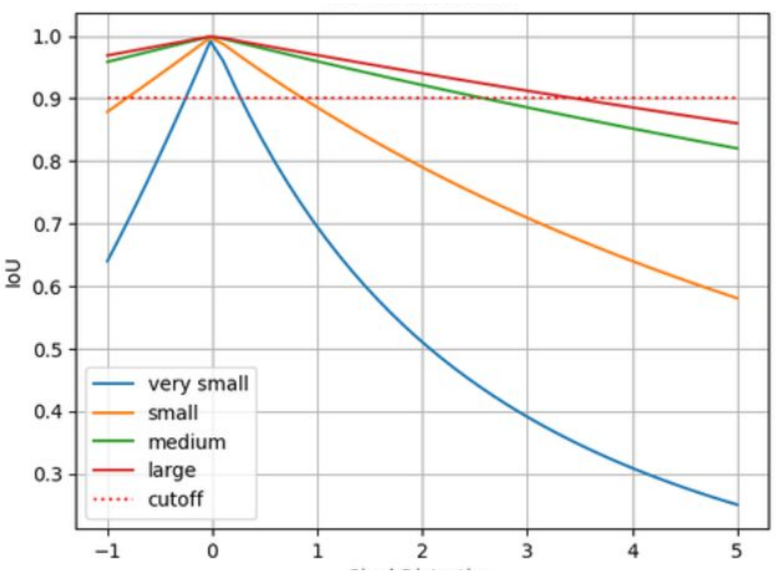
IoU is sensitive to:

- Absolute size of the masks being compared
- Relative differences between the masks
- Small masks are more sensitive to changes in boundaries
- Large masks are more tolerant to changes in boundaries

IoU does not:

- Characterize similarity by contour

Detection requires a calibration threshold to determine what is considered an adequate match between masks.



Theoretical IoU between an annotation and its modulated counterpart

sama

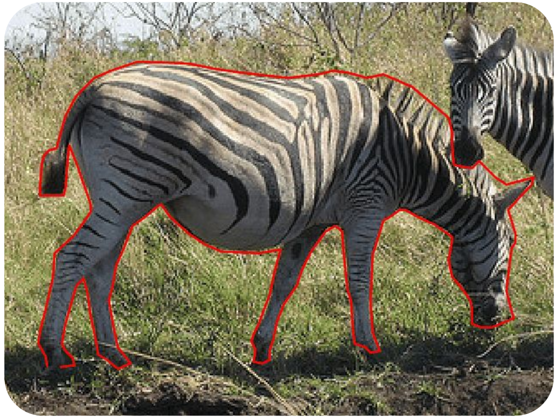
Annotation Uncertainties

What label noise is impactful?

Annotations come with uncertainties:

- Sensor uncertainties
    - Pixelated transitions between object boundaries
- Methodology uncertainties
    - Polygon rasterization errors

Pixel tolerances are task dependent and It is not always possible or to have small pixel tolerances.



Mask Dilations - 1,3,5,10 pixels

Assessing Sama-COCO

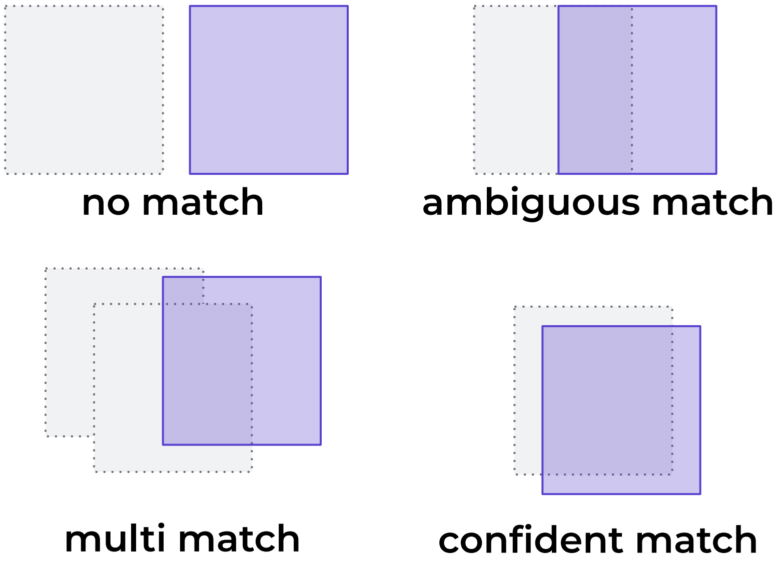
Accounting for differences

Reannotation procedure:

- Instances may have been added or removed
- No correspondence between dataset labels
- Change logs are not available

Analysis procedure:

- Must investigate trends for confident matches
- Quantify distribution shifts
- Assume baseline quality standards



We cannot gauge the absolute performance of a model across datasets due to stylistic differences and distribution shifts without assuming a “gold standard”

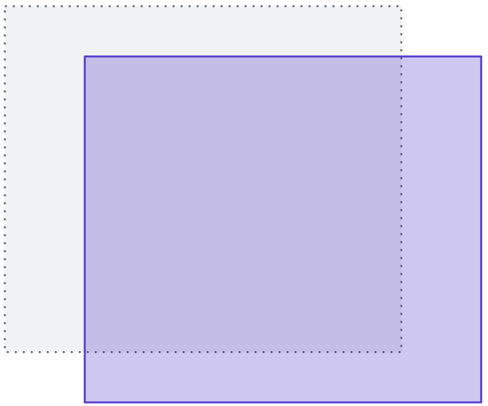
sama

Annotation Comparisons

Finding confident matches

Assumptions:

- Stylistic differences between corresponding matches are contour boundary dependent
- Deformations in a contour boundary have minimal changes on an assigned bounding box
- Matches can be mined using detection metrics



We analyze a subset of COCO's training and validation by finding matches based IoU

sama

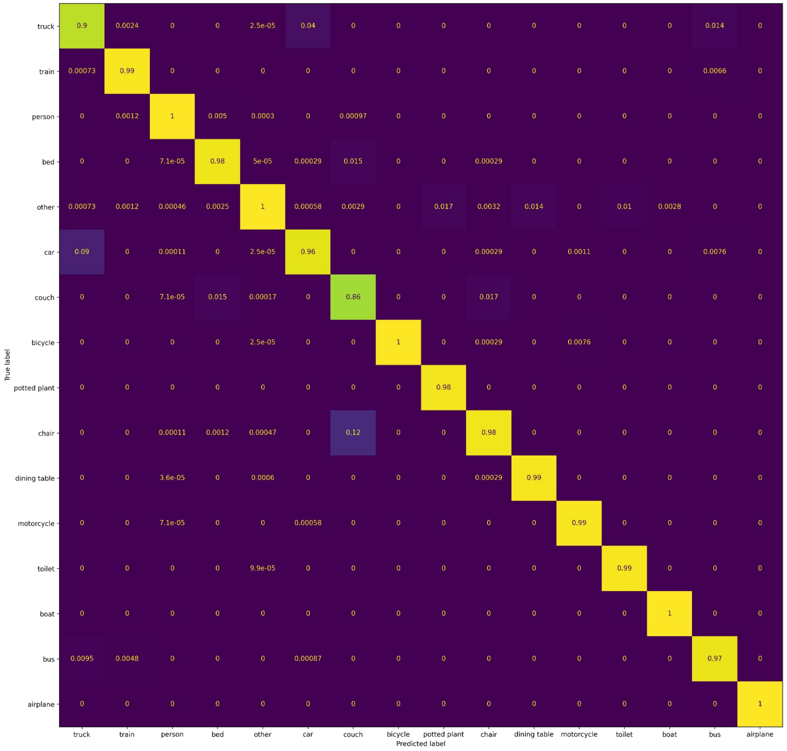
Class Consistencies

Changes in label distribution

Confusion between popular classes:

- People, vehicles, furniture
- Aggregation with “other” class
- Sama-COCO class labels are consistent with MS-COCO with minor differences`

Class confusion occurs between similar classes contained within the same superclass



sama

Estimating Contour Distances

Mining differences via surface distance

Compute sym-distance between two polygons:

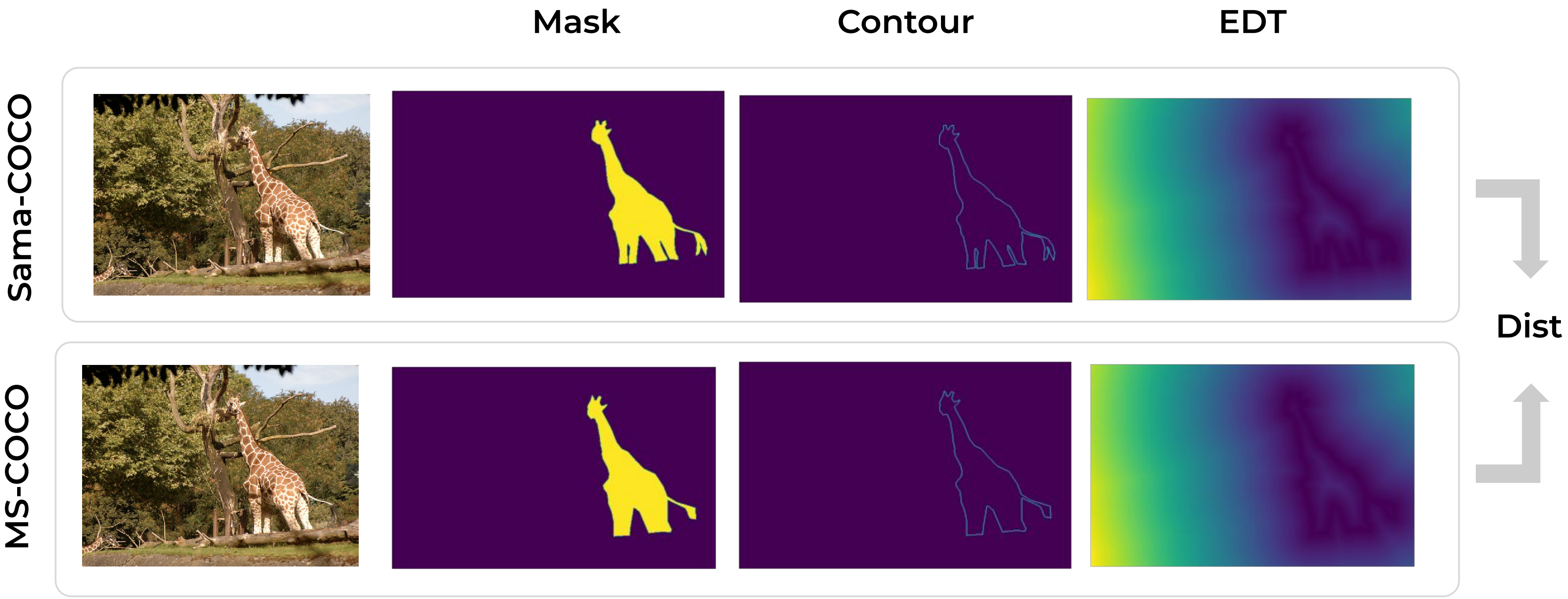
- Get binary mask (M) and extract binary boundary ∂M
    - $\partial M = M \oplus \mathrm{erode}(M)$
- Compute Exact Distance Transforms (EDT)
    - $D_M = \mathrm{EDT}(1 - \partial M)$
- Average distance over contour both contours
    - $d(A,B) = (\int D_B(p)\partial A + \int D_A(p)\partial B) / (\sqrt{2}\ (|\partial A| + |\partial B|))$
    - Consider sub-curves for locality
- Quantification of differences along a boundary and invariant to size of instance

Average distance between contours is the lower bound pixel differences between matched polygon annotations.

sama

Extraction Pipeline Visualization

Processing boundary distances



Large distance between polygons due to granularity on legs and tail

sama

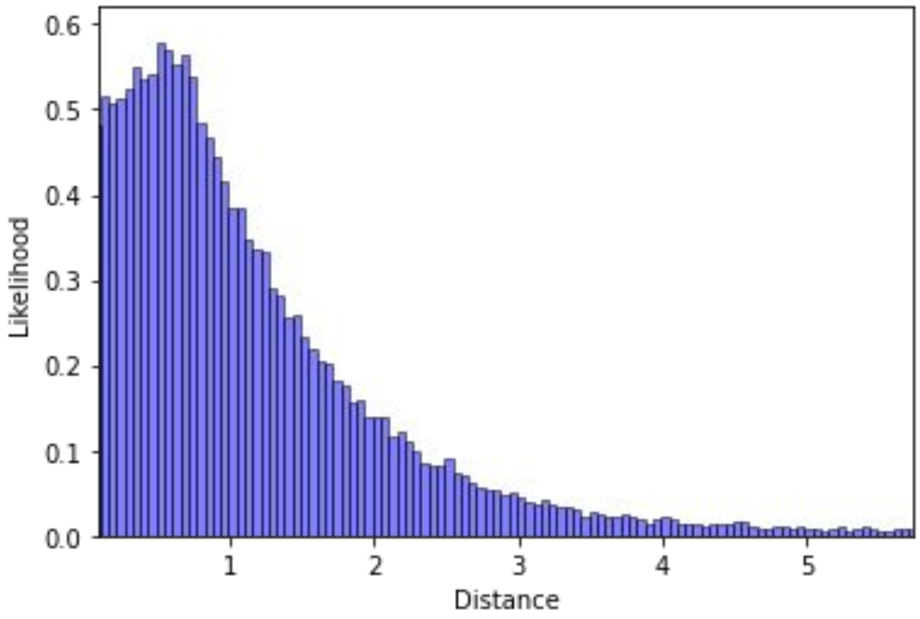
Empirical Distribution of Differences

What changes can we observe?

Estimated distance between confident samples

- Follows an Exponential distribution
- Can sample distributions to observe quantitative differences

Can mine samples of interest based on the distances observed

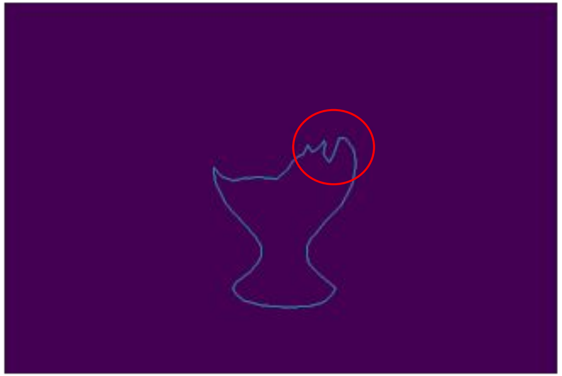


Truncated distribution of distances

sama

Examples

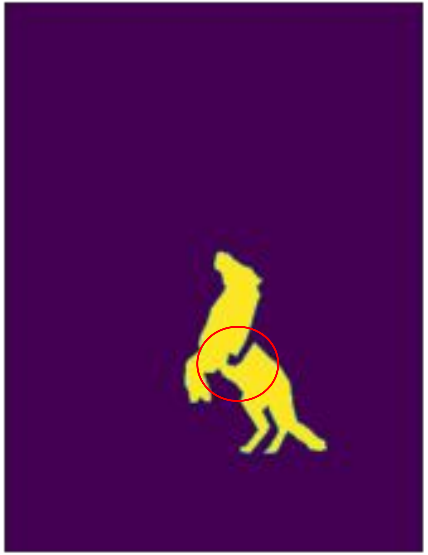
Observed trends

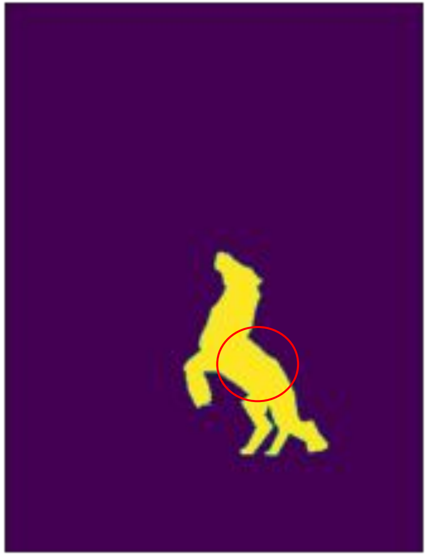




Minor difference in contour around occluded segment

Examples

Observed trends

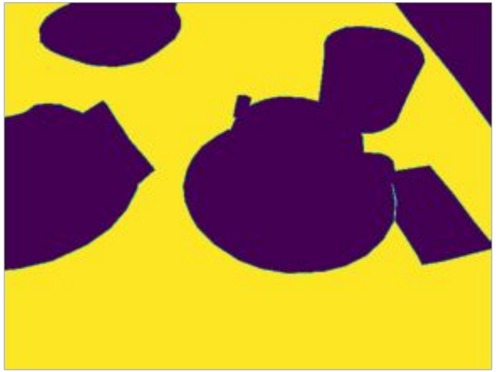




Major difference in style around occluded segment.

Examples

Observed trends



Major difference in content based. Wholes are present where the table is occluded

Small distances correlate to boundary noise while large distances correlate to boundary style

Label Noise and Model Performance

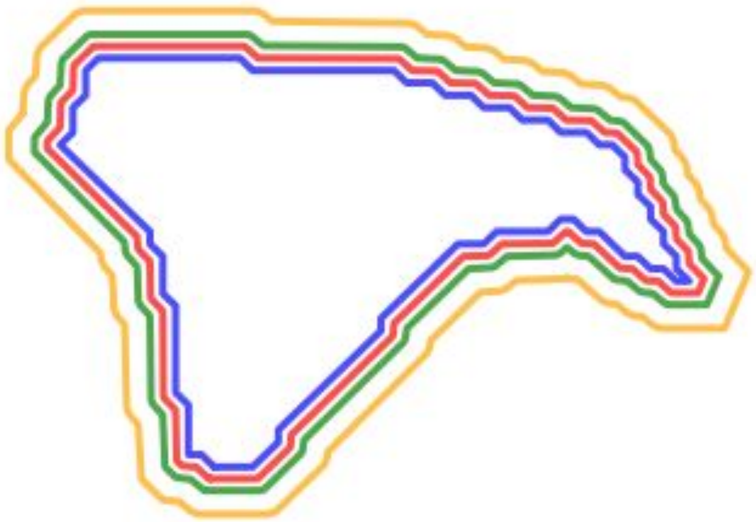
Experimental Setup

How does model performance change with label noise?

Experimental setup

Sama-COCO

- Strictly labeled detection and segmentation dataset treated as the gold standard
- Simulate realistic changes in polygon annotations
- Train RCNN on distorted annotation set
- Evaluate on clean validation set
- Compute mean Average Precision (mAP)



Example for simple dilations of an initial contour

sama

Label Noise and Model Performance

Performance and Results

How does model performance change with label noise?

Results:
Systemic noise leads to systemic bias and degrades qualitative performance.
Model is robust to realistic random noise:

Boundaries that have mixed biases outperformed those with constant bias

Larger instances are more tolerant to changes in boundaries at training time

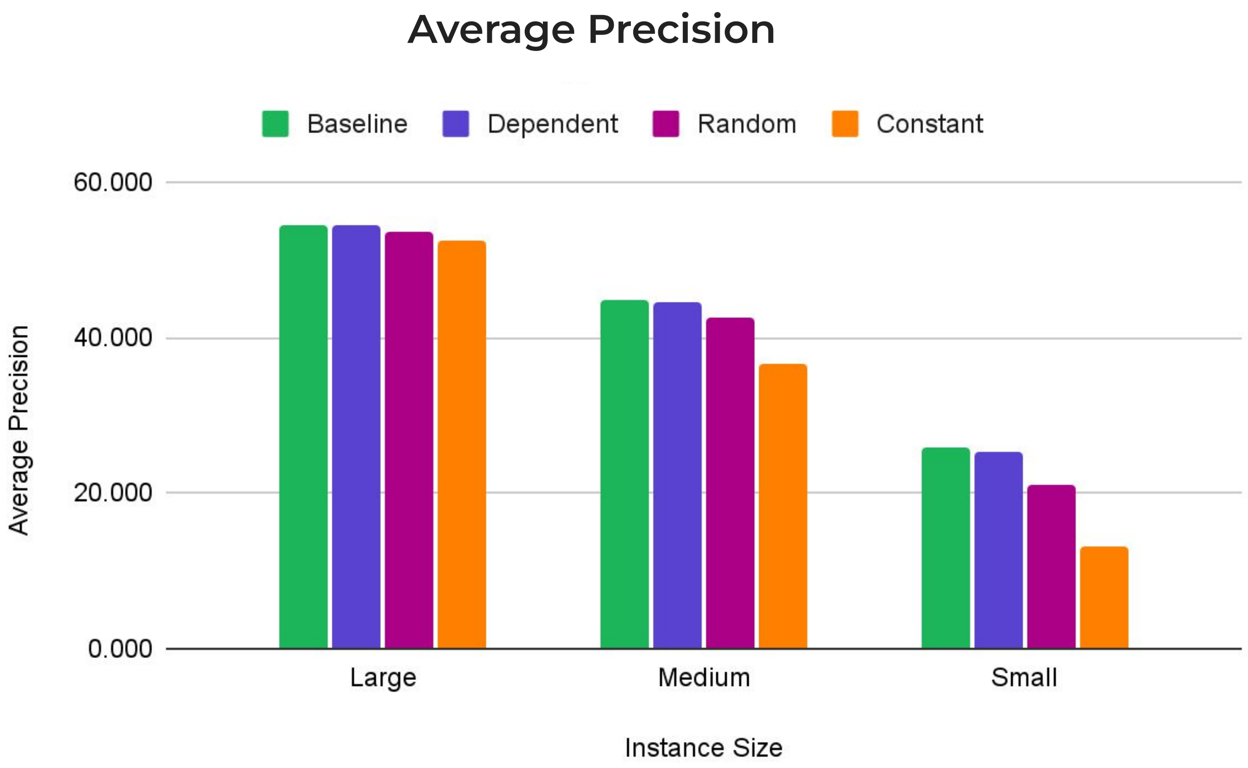
Smaller instances are more sensitive to changes in boundaries at training time

It is possible to specify quality tolerances based on the requirements of the system

sama

Label Noise and Model Performance

Performance and Results



sama

Last Notes & Questions

sama

Open Source Data

We’ve recently released sama-drives-california

Hugging Face

Available on Hugging Face

sama

Open Source Data

We've made available Sama-COCO, a relabelling of the Coco-2017 dataset

VOXEL51

Explore the data

sama

sama

Scan for datasets & periodic updates