Quality Match

Bicycle Project Crowd Evaluation

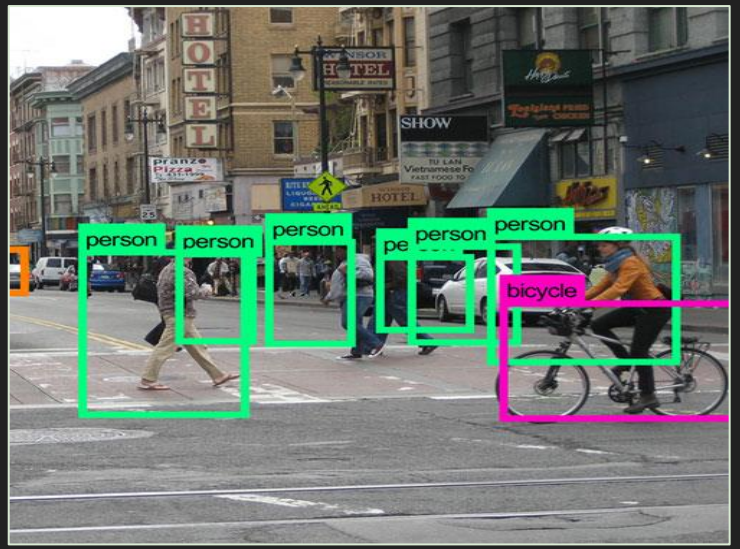
By: Shiwali Kathpal

Contents

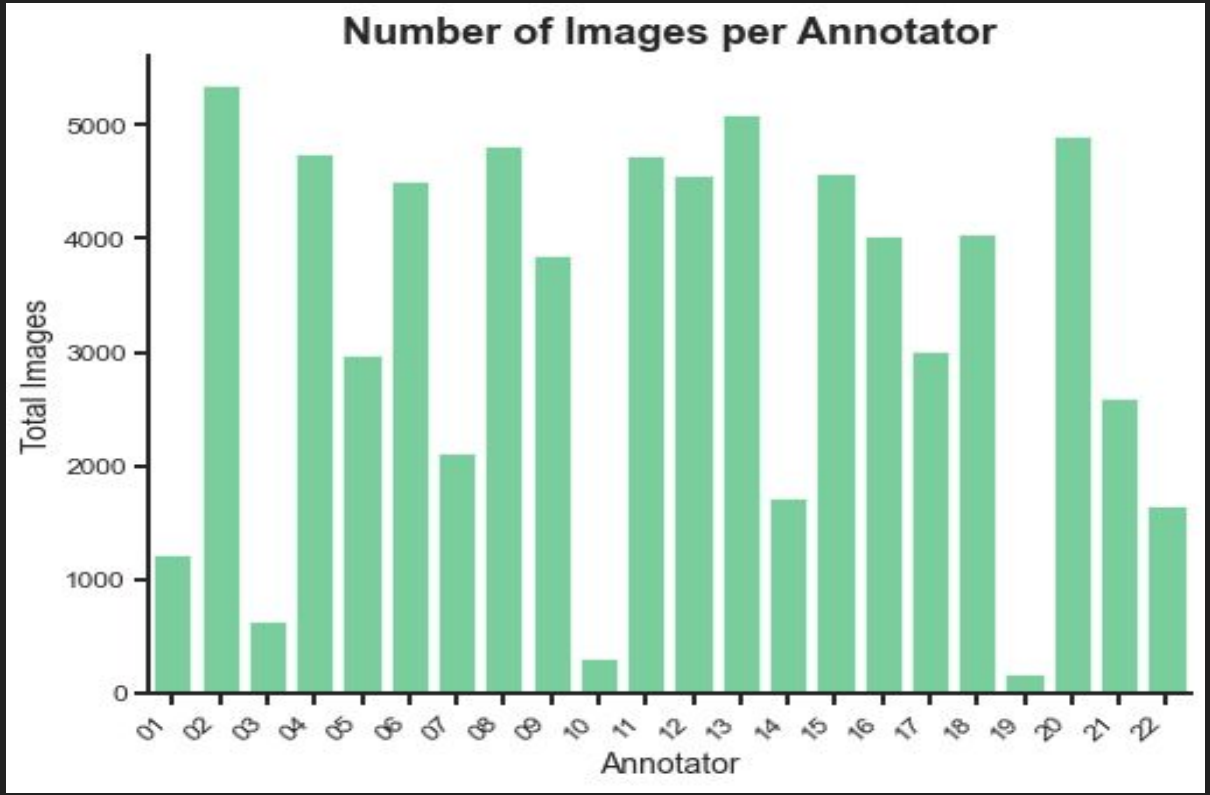
Introduction to data

Question: “Do you see a bicycle?”

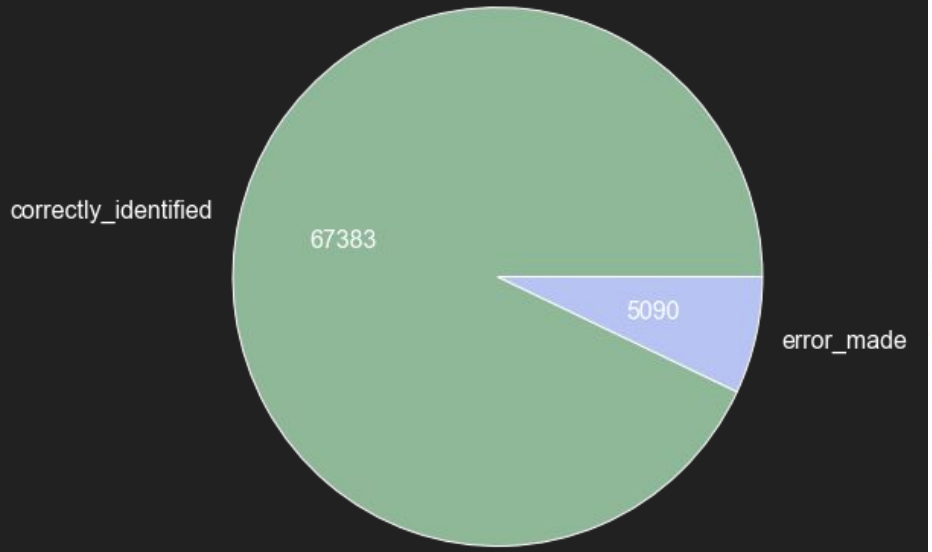
- Number of tasks: 9087
- Number of Annotators: 22
- Each task annotated: 10 times



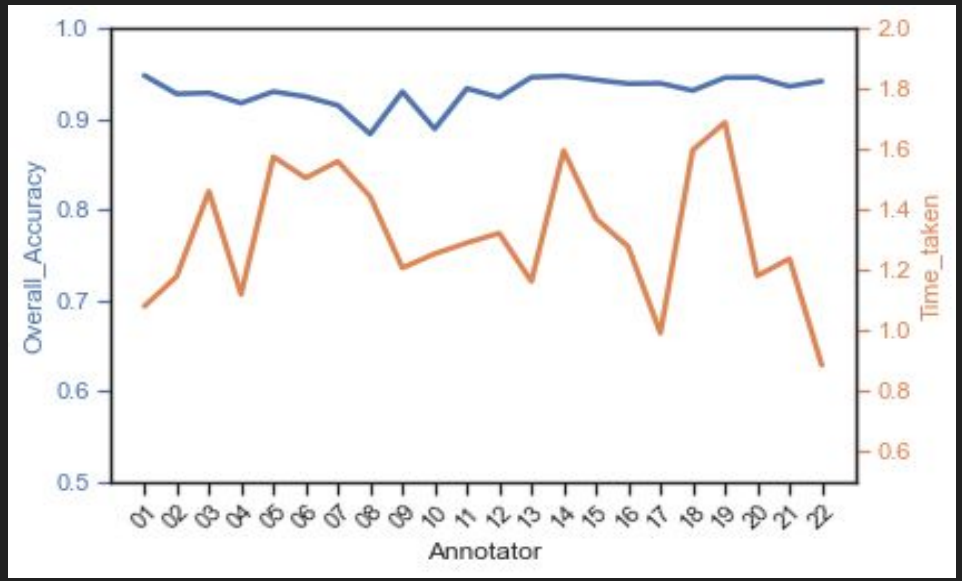
Distribution of Images across Annotators



Overall Annotation Results



Total annotations



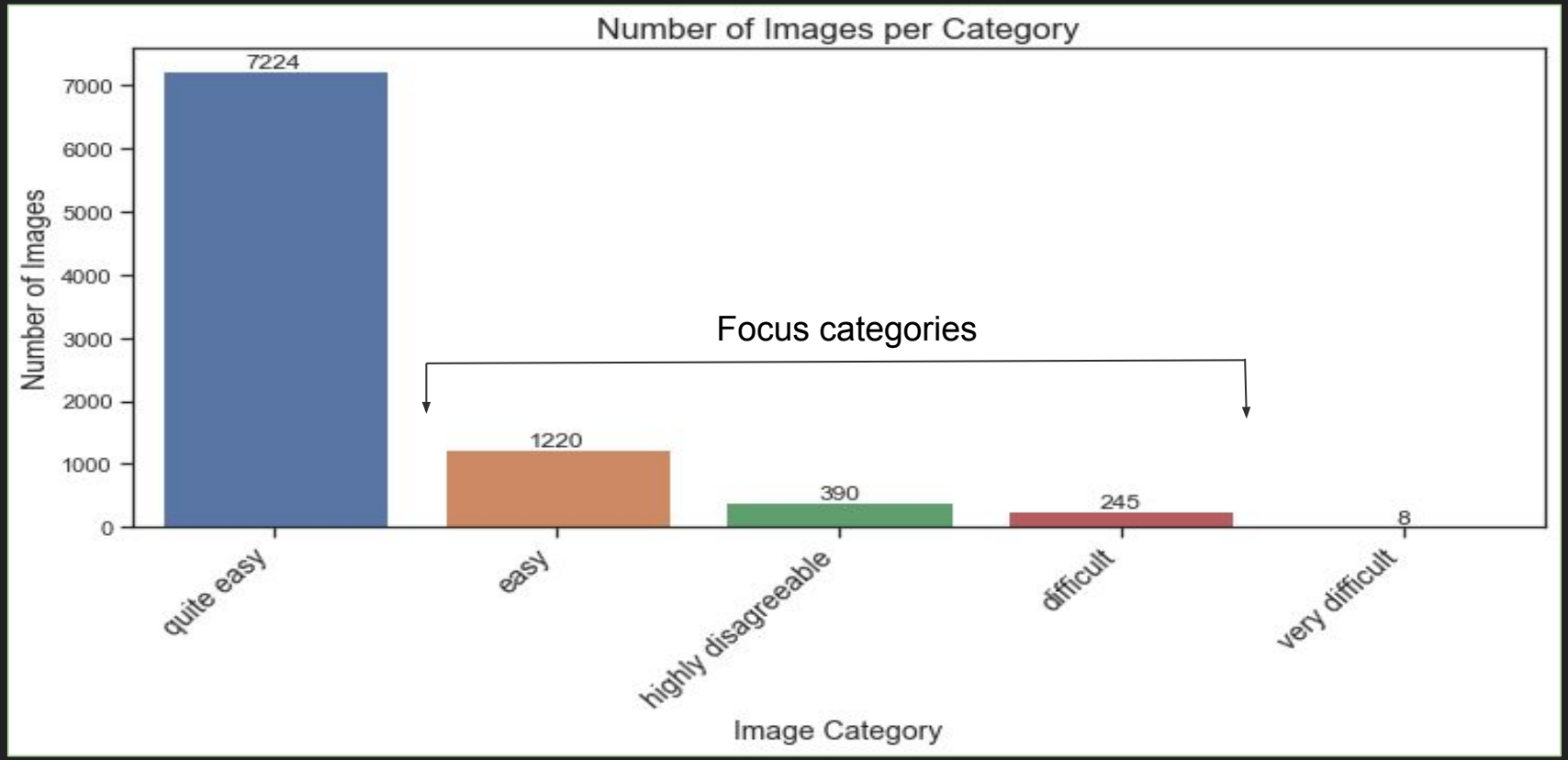
Accuracy vs Speed

Image Classification

Image Classification

Sr. No.	Number of errors	Image category
1.	0	quite easy
2.	1-3	easy
3.	4-6	highly disagreeable
4.	7-9	difficult
5.	10	very difficult

Image Classification



Examples of Images Category wise

Quite Easy (errors = 0)

Reference Output : True

Reference Output : False

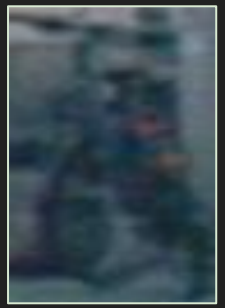
Very Difficult (errors = 10)

Reference Output : False

Reference Output : True

Easy (errors= 1-3)

Reference Output : True



Reference Output : True

Highly Disagreeable (errors = 4-6)

Reference Output : False

Reference Output : False

Difficult (errors = 7-9)

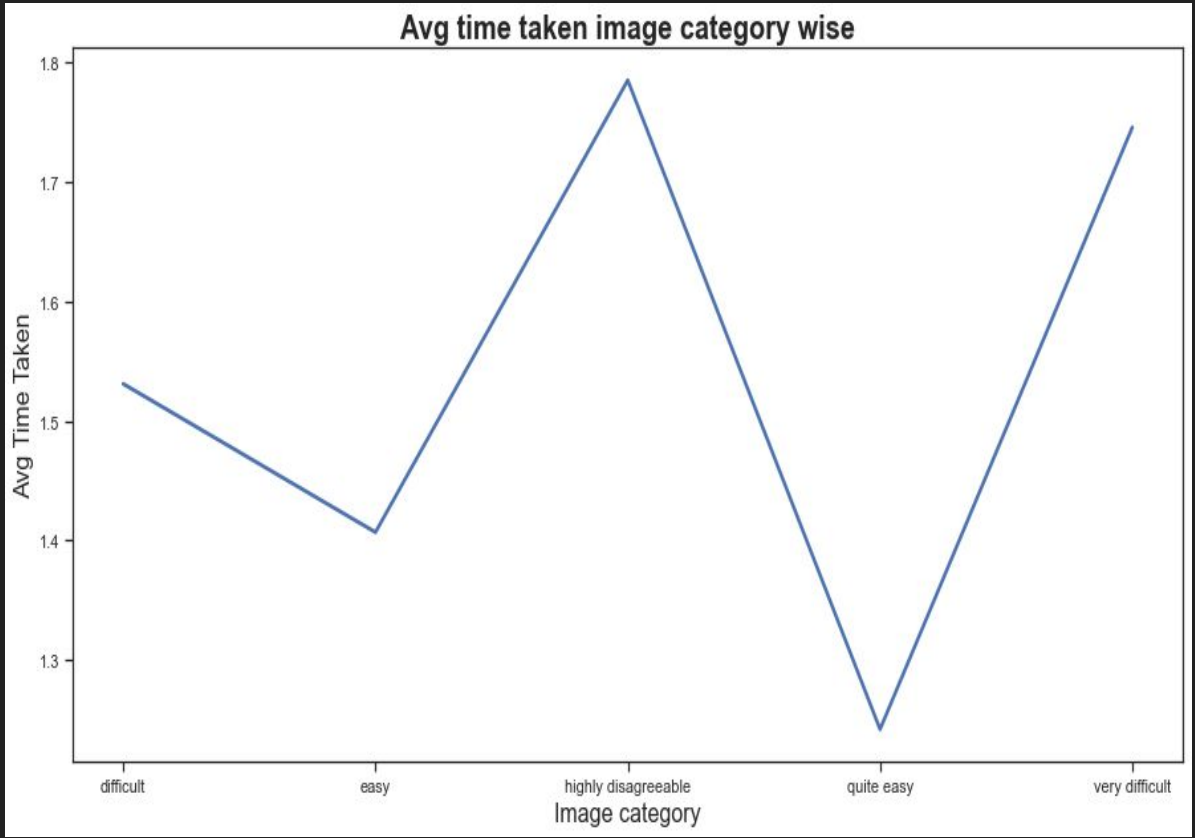
Reference Output : True

Reference Output : True

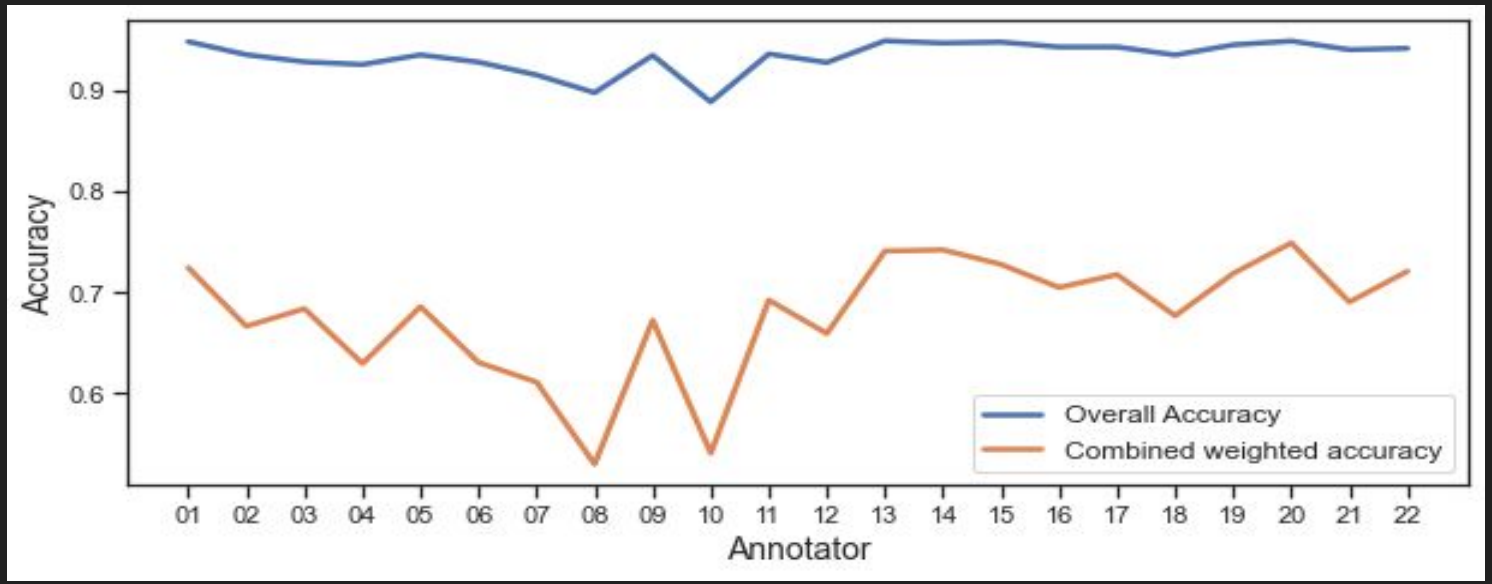
Focus Categories

Metric Formulation

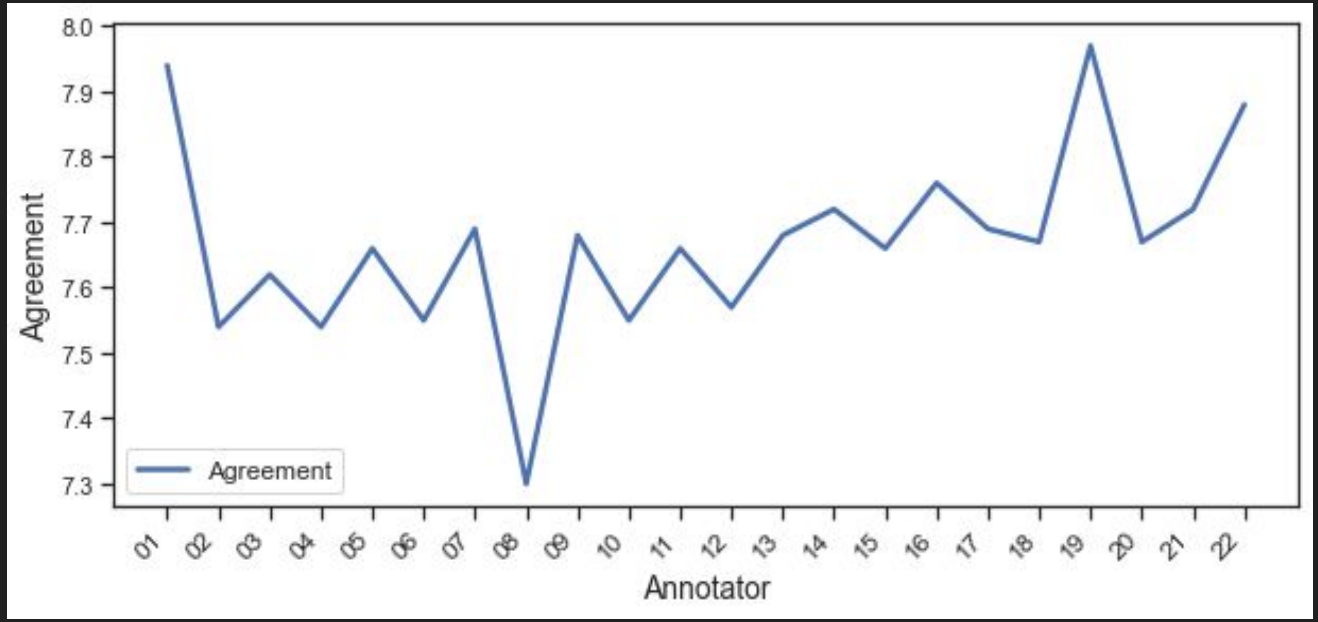
Time Taken by Annotators



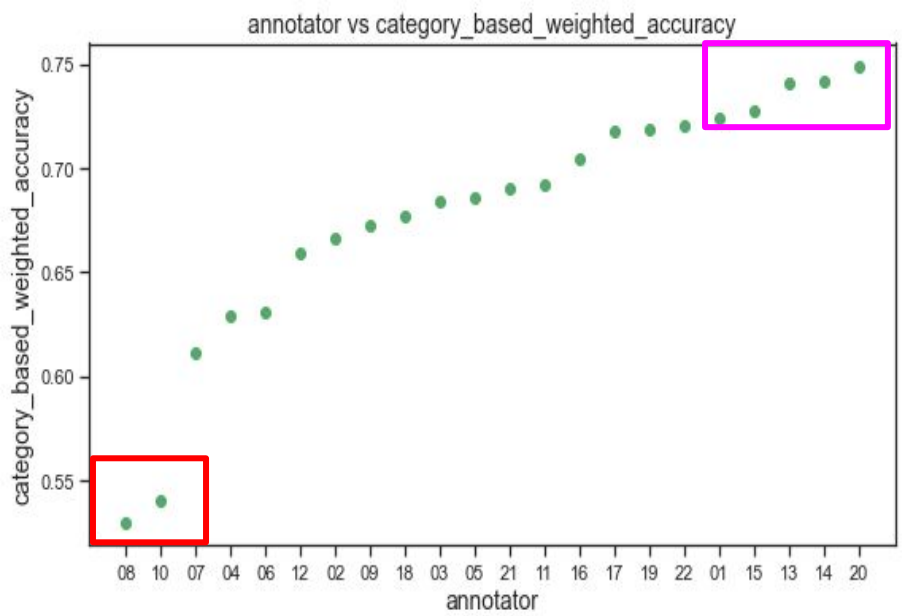
Comparison of Overall accuracy and accuracy in focus categories

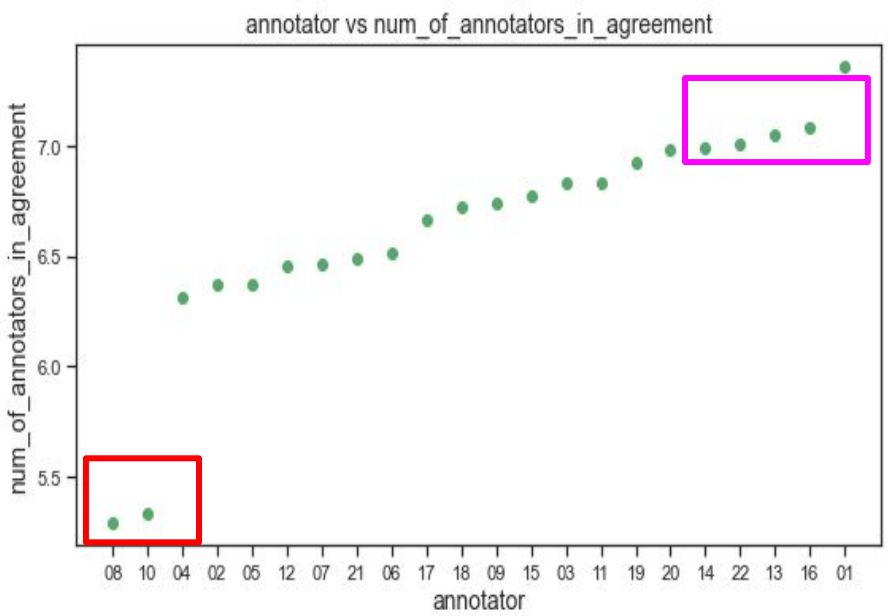


Agreement of annotators



Weighted accuracy & Agreement Distribution





$$\text{Combined Score} = \frac{(W1*\text{Accuracy}) + (W2*\text{Agreement}) + (W3*(1/\text{time}))}{W1+W2+W3}$$

Results

Good annotators

Sr. No.	Annotator
1.	Annotator_19
2.	Annotator_14
3.	Annotator_01

Bad annotators

Sr. No.	Annotator
1.	Annotator_08
2.	Annotator_10
3.	Annotator_04

Clustering results

annotator_num	labels
01	Good
02	Bad
03	Good
04	Bad
05	Good
06	Bad
07	Good
08	Bad
09	Good
10	Bad
11	Good
12	Bad
13	Good
14	Good
15	Good
16	Good
17	Good
18	Good
19	Good
20	Good
21	Good
22	Good

Summary

- 91000 annotations by 22 annotators.
- Filtered duplicate data.
- Filtered further using Image classification.
- Combined score (accuracy, agreement and time).

Conclusion

- Effective Image Classification
- Combined metric of weighted accuracy and level of agreement
- Image scrutiny
- Take on time

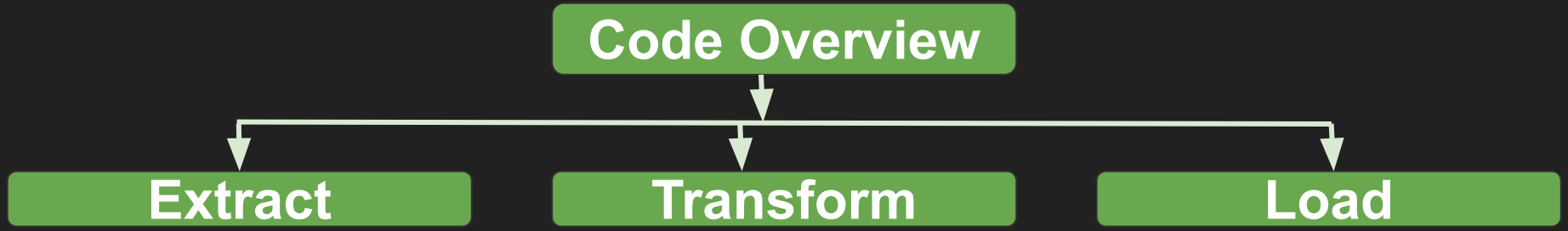
Limitations

- Limited information about the image properties.
- Less data where errors were made.

Further Scope of work

- Better image classification.
- Optimization of metric formula.
- Improved prototype.
- Regular learning for annotators.
- Annotations by GPT-4.

Thank You



- Json parser

- Merger
- Entities extractor
- Data filter
- Metric Builder
- Clustering

- csv