EnsembAudit: Autonomous Error Detection and Rectification

Reducing Labeling Effort: Identifying Labeling Errors Without Access To Ground Truth

Author
Zeryab Alam (Z.M.Alam-2@student.tudelft.nl)

Responsible Professor
Jan van Gemert (j.c.vangemert@tudelft.nl)

Supervisor
Osman Kayhan (o.s.kayhan@tudelft.nl)

1. Introduction

- Accurate object detection relies on detailed annotations.
- Obtaining these annotations is costly and time-consuming.
- Annotations typically contain varying levels of noise.
- Even minor imperfections in annotations can have a significant impact on model accuracy.
- We introduce **EnsembAudit (EA)**, a novel approach to detecting and automatically fixing annotation errors.

2. Method: EnsembAudit (EA)

- EA combines a disagreement monitoring system using **Threshold Voting** (set at 80%) with an ensemble method employing **Non-Max Suppression (NMS).**

2.1 Concepts

Threshold Voting: Requires 80% model agreement for a prediction to be accepted; otherwise, flagged as disagreement.
NMS refines predictions by prioritizing high-confidence, non-overlapping boxes.

2.2 How does it work?

- Multiple object detector models are trained on the same dataset.
- Each model independently predicts on each image post-training.
- EA utilizes Threshold Voting to identify instances where models disagree.
- EA applies NMS to refine predictions on flagged instances.
- New predictions replace old labels to reduce overall noise.

3. Results: EA Application on VOC Corrupted Dataset

3.1 Classification Noise

Model	Actual Errors	Detected Errors	Verified Errors	Performance		
				Precision	Recall	mAP@50
10% Noise	1750	371	87	0.6346	0.7734	0.3611
25% Noise	4375	690	288	0.6753	0.7585	0.2948
50% Noise	8751	2130	1397	0.8099	0.6638	0.2818

EA detects and corrects 16% of errors in a dataset with 50% classification noise, achieving 80% precision and 66% recall.

3.2 Localization Noise

Model	Actual Errors	Detected Errors	Verified Errors	Performance		
				Precision	Recall	mAP@50
10% Noise	1750	248	40	0.5536	0.6596	0.2867
25% Noise	4375	286	104	0.4889	0.5906	0.2809
50% Noise	8751	423	262	0.5316	0.6359	0.4128

EA struggles to detect errors with minor localization noise, achieving a success rate of about 2%.

3.3 Both Noise Types

Model	Actual Errors	Detected Errors	Verified Errors	Performance		
				Precision	Recall	mAP@50
10% Noise	1750	391	124	0.6289	0.7394	0.3015
25% Noise	4375	778	393	0.6391	0.6868	0.2422
50% Noise	8751	2600	1833	0.7731	0.5844	0.2699

EA achieves 20% error detection and correction with 50% noise presence, showing good metric scores.

4. Discussion & Conclusions

- Results demonstrate **EA's effectiveness** in handling datasets with labeling errors, autonomously detecting and rectifying approximately **20% of errors.** This highlights EA's potential to enhance object detection accuracy in practical applications.
- A notable **limitation** observed is EA's reduced efficacy with datasets containing minimal noise typical in practical scenarios. It shows limited ability to detect errors due to the resilience of pretrained models used. EA also shows high false positives, requiring precise threshold tuning to prevent unnecessary label changes.
- Future research can enhance EA's ensembling techniques and disagreement systems, improving error detection across diverse datasets.