Tampere University

Iterative Bounding Box Annotation for Object Detection

Bishwo Adhikari & Heikki Huttunen

Background

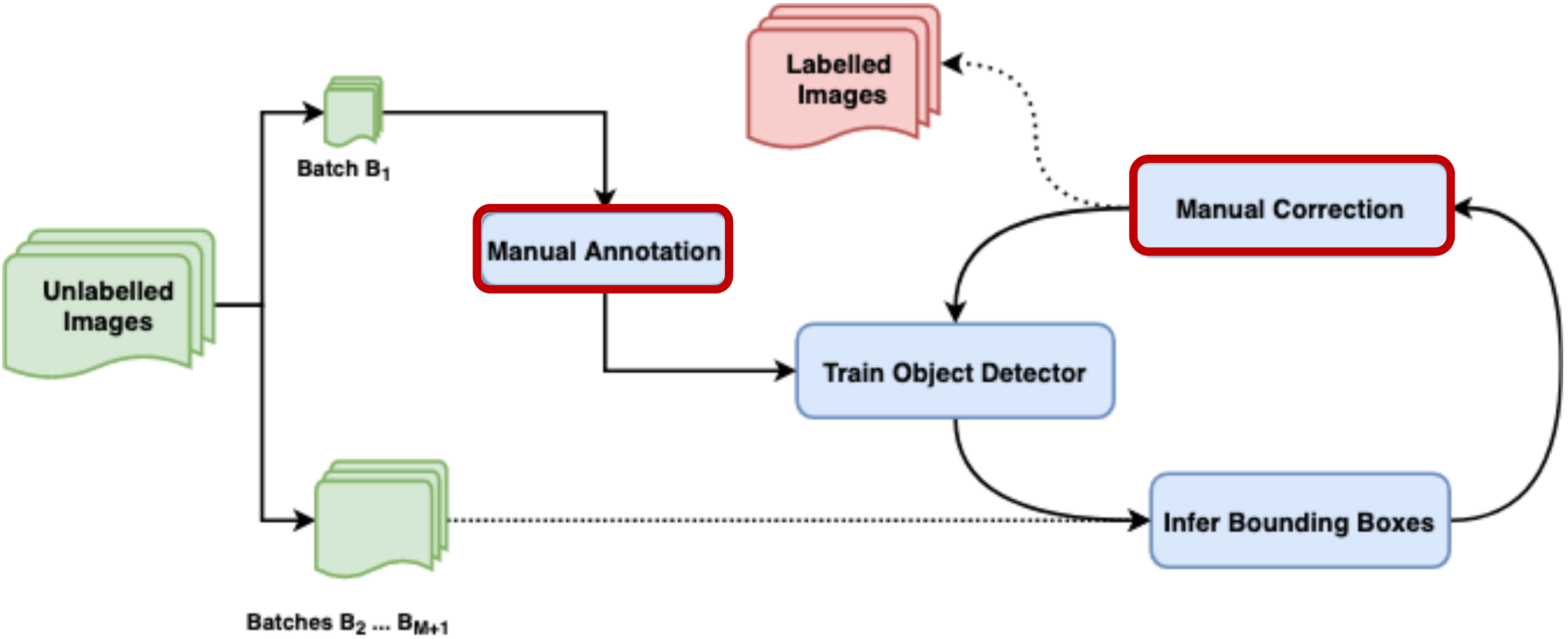
- Supervised object detection requires large amount of labelled data for training.
- Even though many fully labelled dataset is already available, for robust detection some amount of labelled data is required from the test environment.
- Labelling object class and location in image dataset is tedious, error prone and time consuming.

Our approach

- We present an iterative *train-annotate* approach for the bounding box annotation.
- Our method uses freshly trained detector to propose labels for a batch of unlabeled images leaving the annotator inspection and correction work.

Tampere University

Method



Method

Algorithm 1: Iterative annotation

Require: Set of unlabeled images split to $M+1$ distinct annotation batches $B_0, \ldots, B_{M+1}$

annotate images in batch B_0 manually
train object detection model with images from B_0
for $i \in 1, 2, \ldots, M$ **do**
    propose annotations for batch B_i using the current prediction model
    do manual correction for the proposals
    fine-tune the object detection model with batch B_i
end for

return fully labeled dataset

Experiments

- Datasets: Indoor, Pascal VOC 2012 and OpenImages V4 person class
- Networks: SSD MobileNet and Faster RCNN with MSCOCO pre-trained weight
- Sampling strategies:
  - Shuffle – Random shuffle
  - Sorted – Based on the object density
  - Original – Based on the temporal order (Indoor)/file name (VOC & OpenImages)

Results

Table 1: Annotation workload reduction (%) in 3 datasets

Network - Approach	Indoor	PASCAL VOC	OpenImages Person
RCNN - Shuffled	**75.86**	18.40	45.62
RCNN - Sorted	56.97	20.93	**60.05**
RCNN - Original	35.78	**25.23**	45.73
SSD - Shuffled	**47.38**	3.46	20.28
SSD - Sorted	31.58	5.66	**35.13**
SSD - Original	19.24	**7.97**	20.04

Results

Table 1: Annotation workload reduction (%) in 3 datasets

Network - Approach	Indoor	PASCAL VOC	OpenImages Person
RCNN - Shuffled	**75.86**	18.40	45.62
RCNN - Sorted	56.97	20.93	**60.05**
RCNN - Original	35.78	**25.23**	45.73
SSD - Shuffled	**47.38**	3.46	20.28
SSD - Sorted	31.58	5.66	**35.13**
SSD - Original	19.24	**7.97**	20.04

Table 2: Annotation workload reduction (%) in Pascal VOC single class

	Airplane	Bird	Boat	Bottle	Car	Cat	Chair	Dog	Person	Plant	Average
RCNN - Shuffled	56.14	50.30	**35.70**	44.49	51.96	55.34	29.31	57.87	44.61	**38.72**	46.44
RCNN - Sorted	**62.07**	**60.43**	35.65	**46.68**	**56.27**	59.53	**32.44**	63.28	**61.24**	32.75	**51.03**
RCNN - Original	53.87	50.41	32.50	41.54	55.14	**61.58**	29.30	**61.38**	57.16	34.64	47.75

Results

Table 1: Annotation workload reduction (%) in 3 datasets

Network - Approach	Indoor	PASCAL VOC	OpenImages Person
RCNN - Shuffled	**75.86**	18.40	45.62
RCNN - Sorted	56.97	20.93	**60.05**
RCNN - Original	35.78	**25.23**	45.73
SSD - Shuffled	**47.38**	3.46	20.28
SSD - Sorted	31.58	5.66	**35.13**
SSD - Original	19.24	**7.97**	20.04

Table 3: Workload reduction in Indoor dataset with two-stage method [1] & ours

Approach	Reduction (%)
Two-stage (5%) [1]	79.47
Two-stage (6%) [1]	81.21
Two-stage (8%) [1]	78.68
Two-stage (10%) [1]	79.03
Two-stage (20%) [1]	72.46
Ours (iterative)	79.56
Ours (cumulative)	80.56

Table 2: Annotation workload reduction (%) in Pascal VOC single class

	Airplane	Bird	Boat	Bottle	Car	Cat	Chair	Dog	Person	Plant	Average
RCNN - Shuffled	56.14	50.30	**35.70**	44.49	51.96	55.34	29.31	57.87	44.61	**38.72**	46.44
RCNN - Sorted	**62.07**	**60.43**	35.65	**46.68**	**56.27**	59.53	**32.44**	63.28	**61.24**	32.75	**51.03**
RCNN - Original	53.87	50.41	32.50	41.54	55.14	**61.58**	29.30	**61.38**	57.16	34.64	47.75

[1] Adhikari *et al.*, “Faster Bounding Box Annotation for Object Detection in Indoor Scenes,” EUVIP, 2018.

Conclusion

- We present iterative train-annotate approach for the bounding box annotation.
- It is annotator friendly. Single annotator can efficiently create environment specific object detection dataset.
- This method is effective for annotation campaign and it could save up to 75% of manual annotation workload.
- Active learning approach could be useful for the selection of images to be labelled.

Tampere University

Contact us

Bishwo Adhikari
bishwo.adhikari@tuni.fi