We don't need no bounding-boxes: Training object class detectors using only human verification
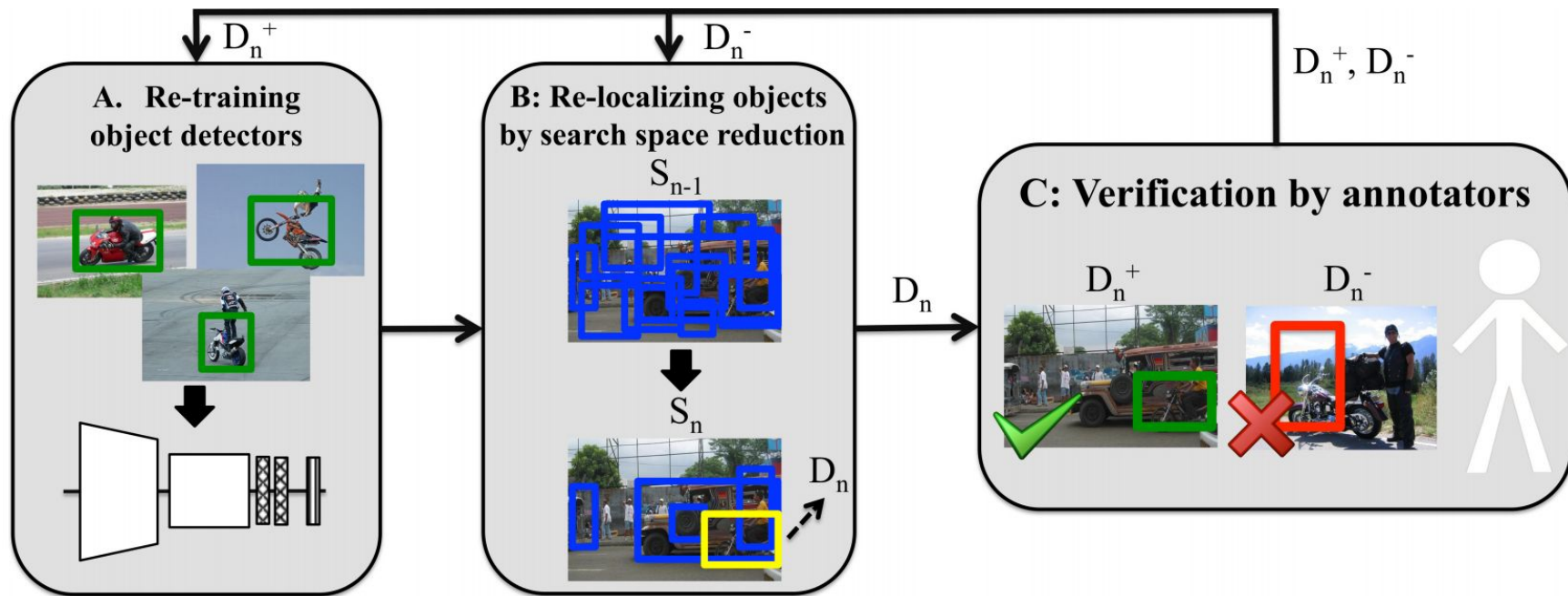
Dim P. Papadopoulos, Jasper R. R. Uijlings, Frank Keller, Vittorio Ferrari

ME780 - Paper Presentation - By Sean Walsh

Presentation Agenda

- System Design
- Issue being addressed
- Framework
    - Object Detector
    - Initial Object Proposals
- Methods
    - Yes/No
    - YPCMM
- Results
- Contributions and Shortcomings

System Design

Issues Being Addressed

- Annotating a datasets is expensive (Time, $)

- What other options exist?
  - Weakly-Supervised Object Localization (WSOL)
  - Using pre-trained networks
  - Active Learning
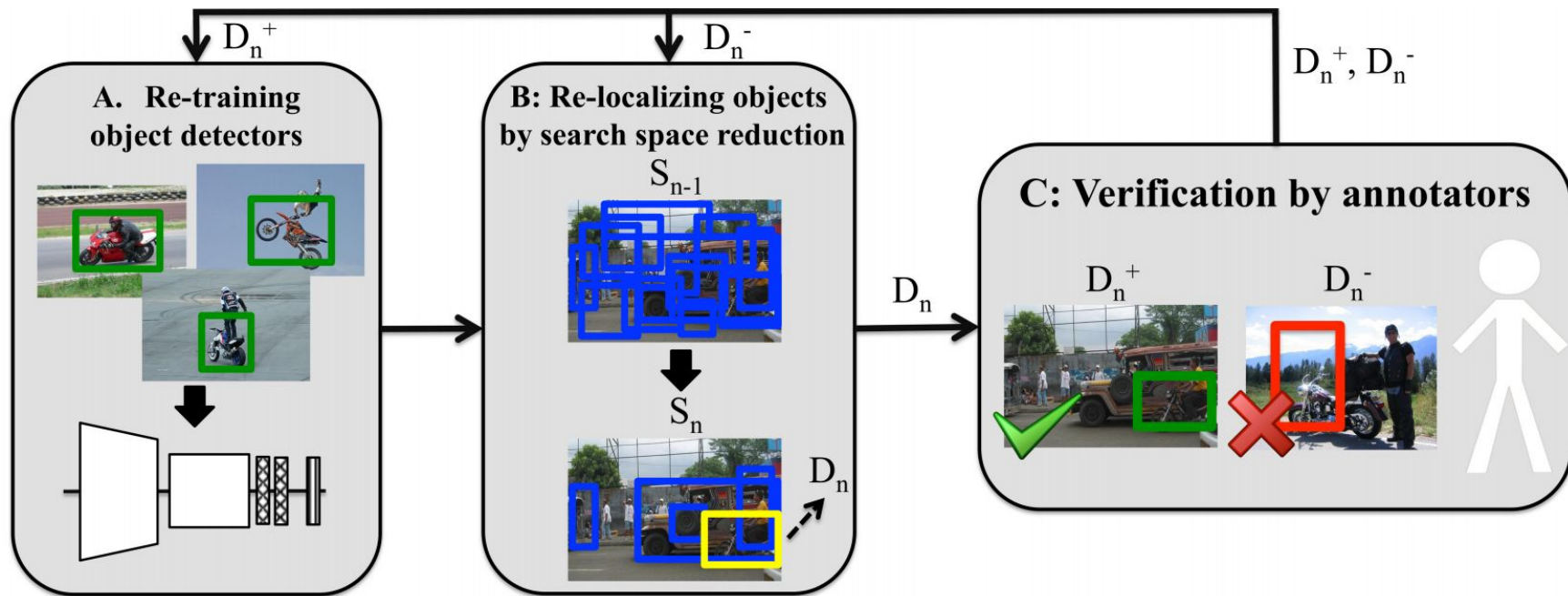 - Others (video based methods, related class labeling, ect.)

Framework

- Object Class Detection
  - Fast RCNN
  - EdgeBoxes over Selective Search
  - No Bounding Box regression
- Object Localization
  - Use EdgeBoxes to gain starting set
  - CNN features fed through SVM
- Initialization
  - Iterate between re-training Object Classification and re-localizing Object proposals
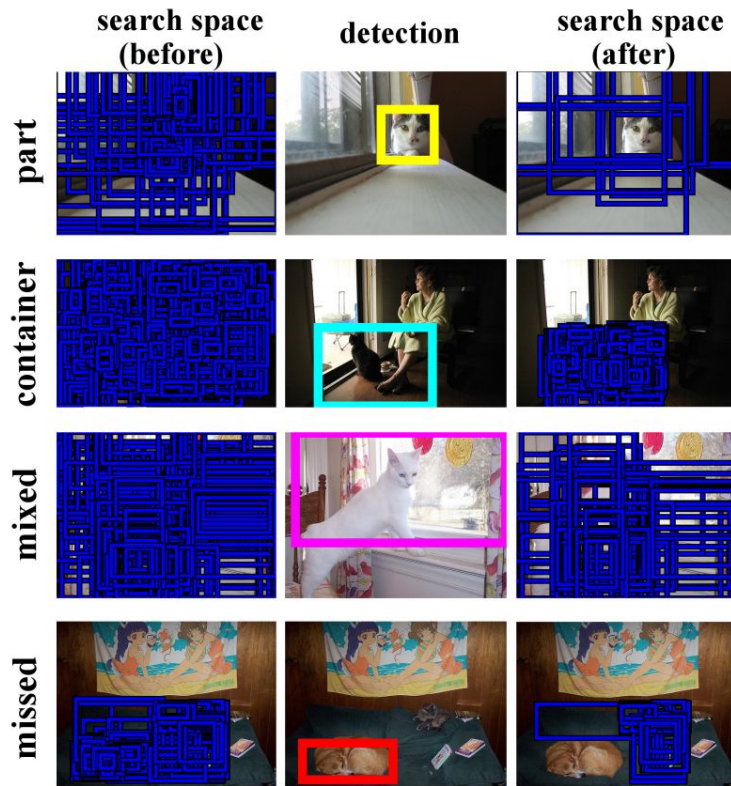 - Stop when Localization stabilizes

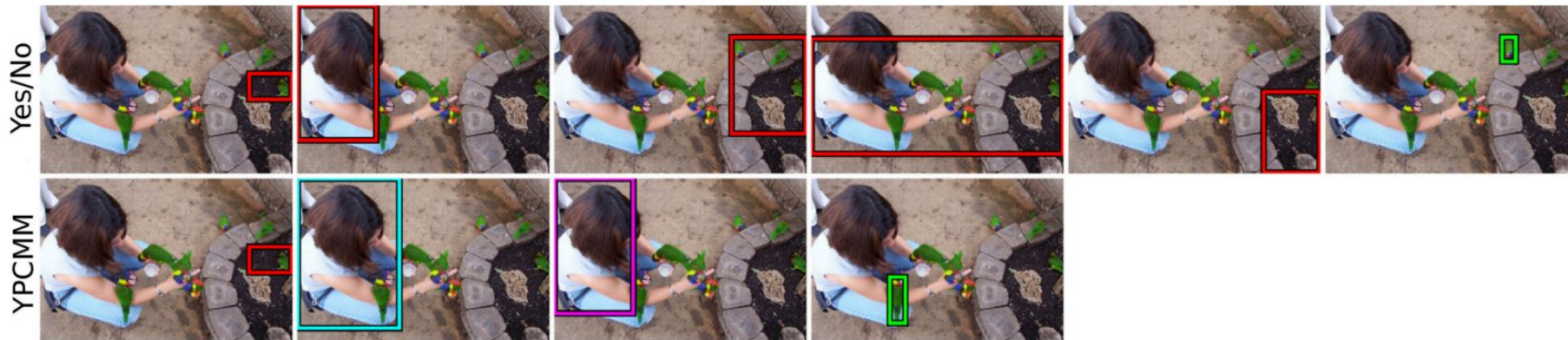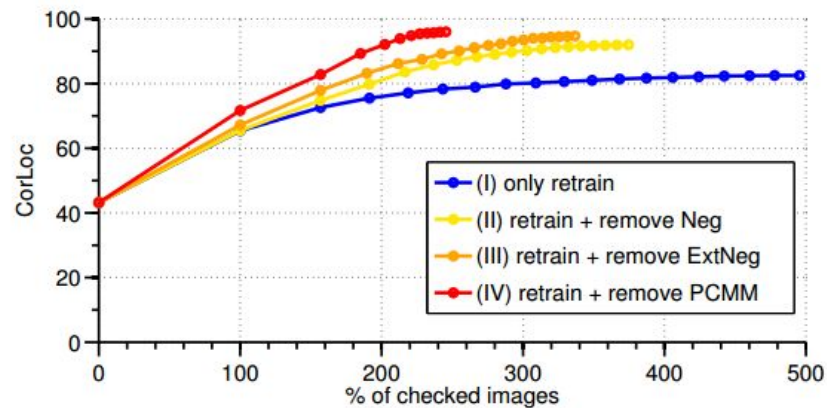# Methods

Re-train Object Detectors

D_n^+

D_n^-

D_n^+, D_n^-

A. Re-training object detectors

B: Re-localizing objects by search space reduction

S_{n-1}

S_n

D_n

D_n

C: Verification by annotators

D_n^+

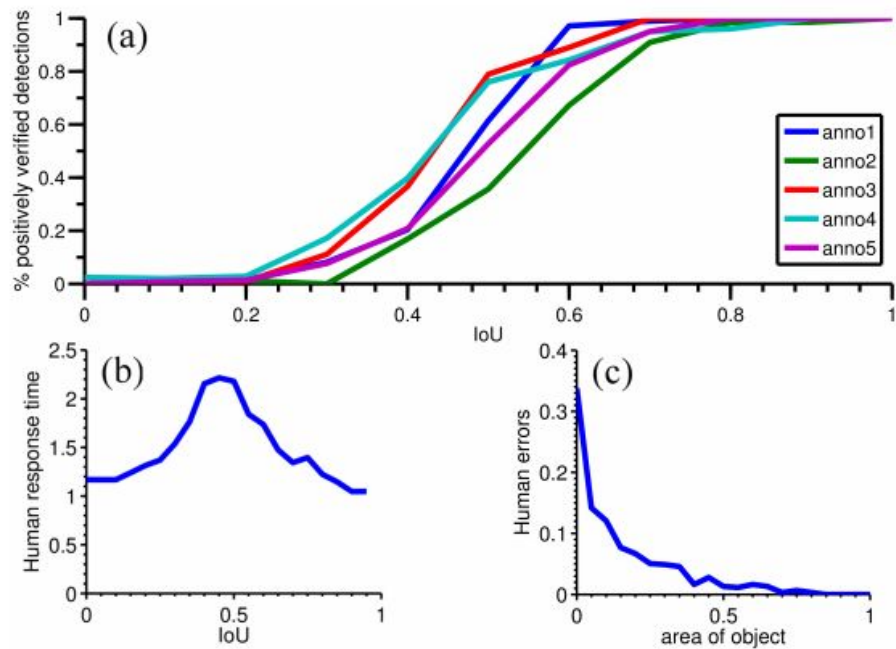D_n^-

Re-localize Object Proposals

- **All images with Yes can be ignored**

- **Yes/No**
  - Remove the No proposal
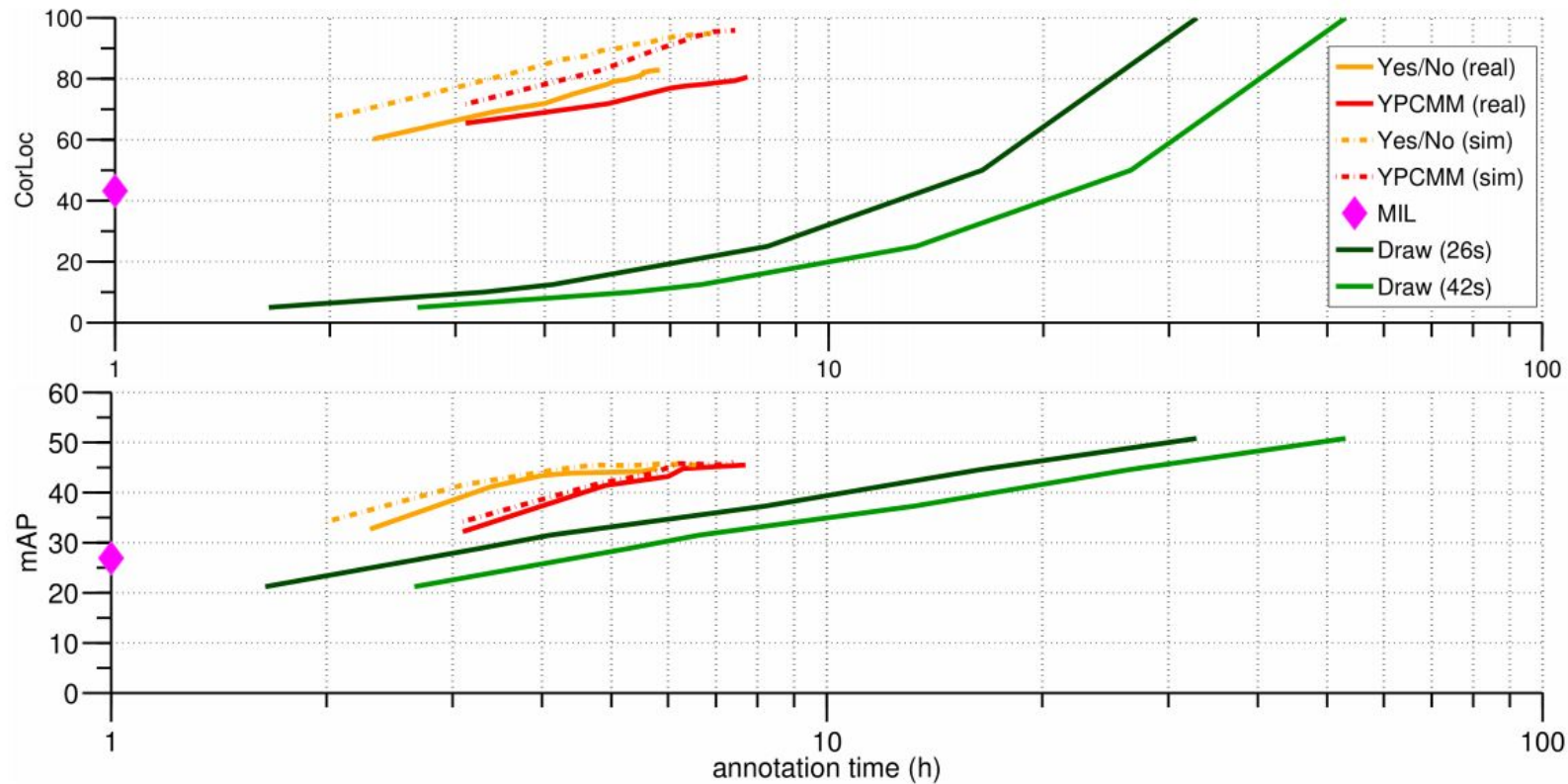 - Additionally those with IOU > 0.5 with No

Results

Results

Results

Results

	reduced training set		complete training set	
	Yes/No	FS	Yes/No	FS
AlexNet	45%	51%	50%	55%
VGG16	55%	61%	58%	66%

Advantages

Drawbacks

- Shows large advantages in combining human verification with deep learning
- 90% of the mAP achieved by fully supervised counterparts
- Significantly faster at localizing objects over drawing bounding boxes

- The timing evaluations were based on training sets with challenging images removed
- Only useful when looking to train with limited resources:
  - Images aren't 100% annotated
  - mAP less desirable for application