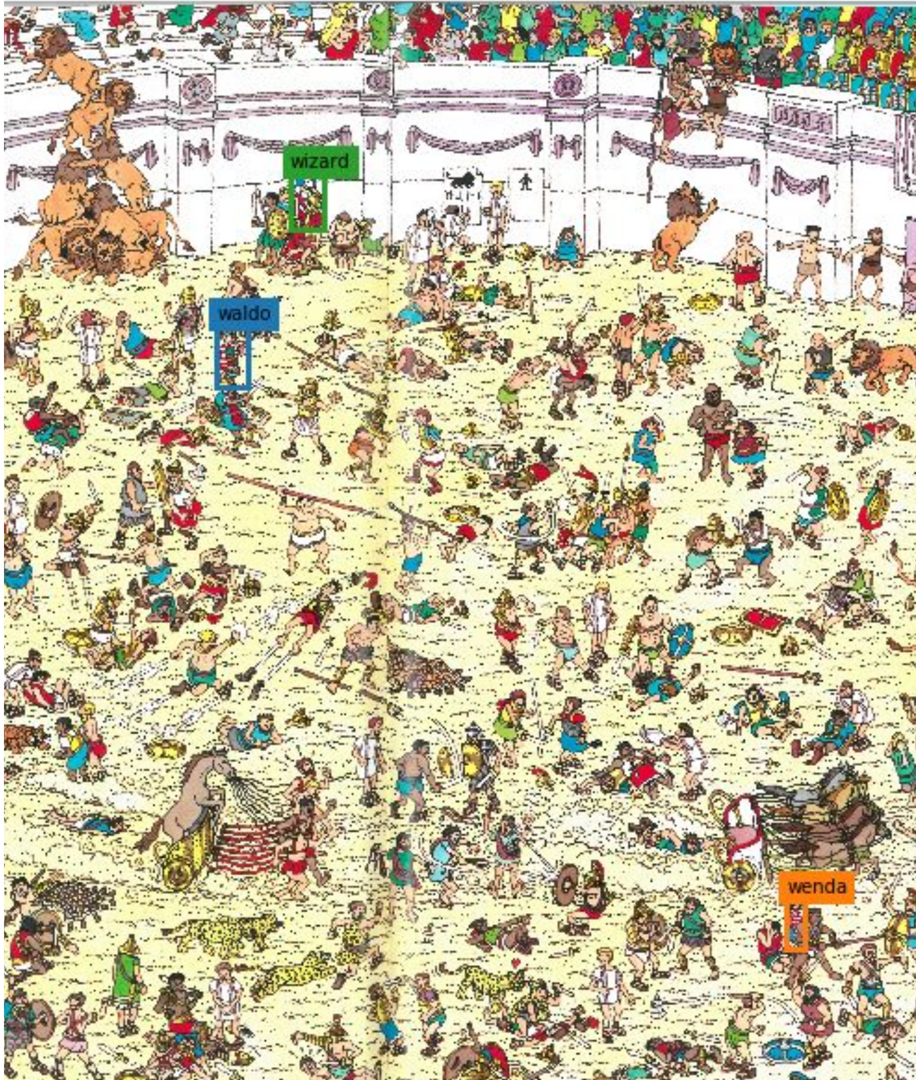
Where’s Waldo ?!

Multiclass object detection in large images

Maximilian Fruehauf, David Drews, Choo Wen Xin
{e0445541, e0454245, e0053347}@u.nus.edu

Problem statement

- Detect bounding boxes for the characters *Waldo, Wenda* and *Wizard* in Where's Waldo search book images
- Can use any "Non-Deep" method
- Training set of ~100 Images for all classes combined

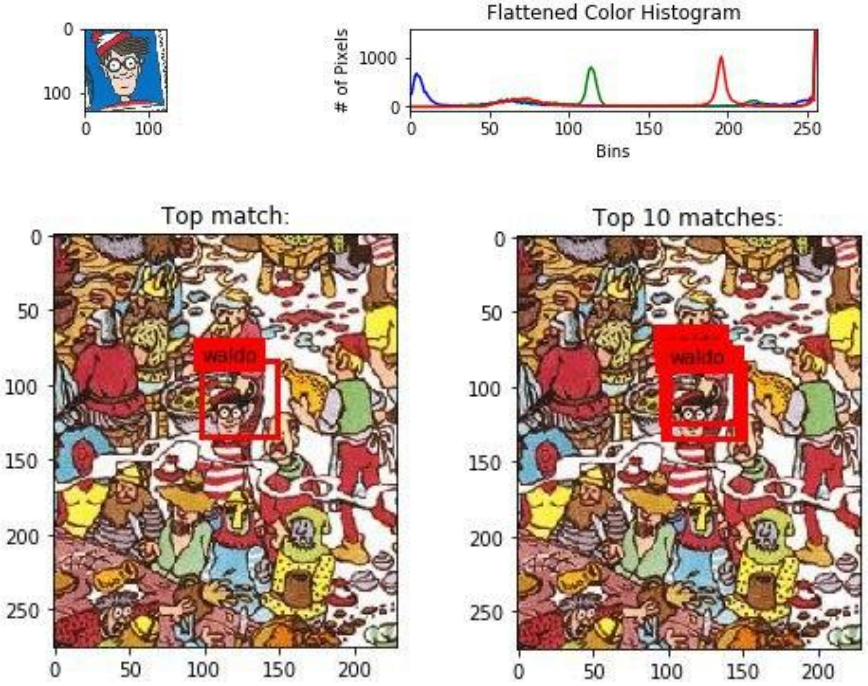


Challenges aka Why it’s hard

- Class instances have a multitude of appearances and can be partially covered by other items
- Ground truth bounding boxes have different aspect ratios
- Illustrated characters not detected well for pre-trained classifiers

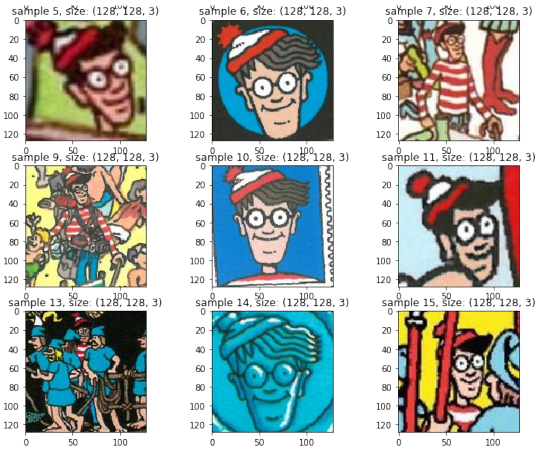
Solutions

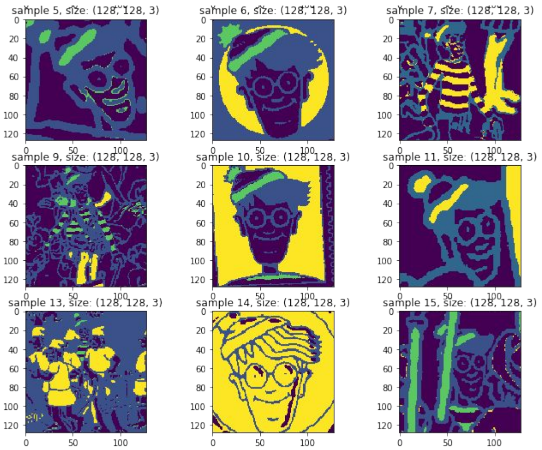
- Histogram correlation
  - RGB histogram correlation with training image
- Haar-cascade:
  - ?? → David
- HOG & SVM:
  - Sample preprocessing
  - window classification based on HOG feature vector with SVM



Sample Preprocessing

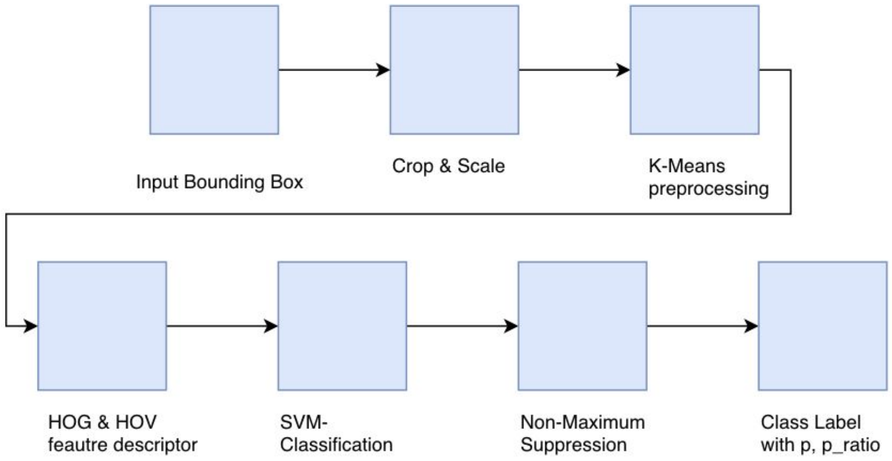
- Goals of preprocessing
  - crop bounding boxes to 128 x 128 px
  - Reduction gradient complexity
  - Grayscale mapping of color samples
- We use k-means to find k = 5 modes corresponding to main colors in positive samples
  (red, blue, black, white, other)
  - Uniform mapping of found colors to grayscale values in [0, 255]
- Random generation of negative samples





HOG / HOV & SVM

- HOG / HOV feature descriptor
  - HOG for gradient based information
  - HOV for color distribution information (based on pyramidal SIFT)
- SVM multiclass classifier:
  - One vs. one approach, as examples per class are highly unbalanced
  - Radial basis function to achieve linear separability
  - Used parameter search to determine optimal hyperparameters (C, γ)
- Non-Maximum Suppression



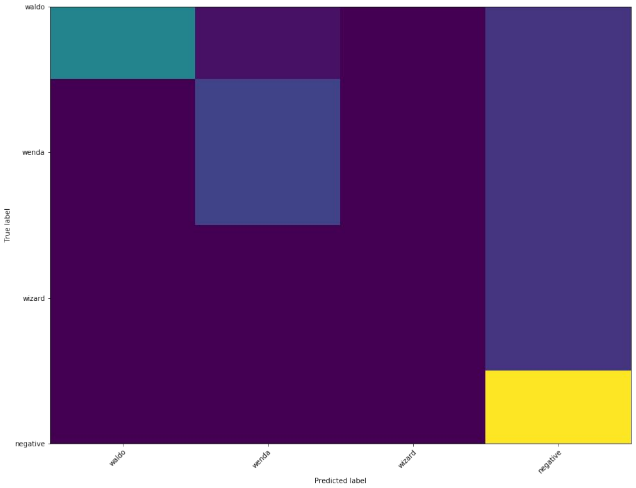
Class name	#samples	percentage
waldo	137	1.01%
wenda	43	0.32%
wizard	27	0.20%
negative	13315	98.47%

Table 1: Dataset statistics for the SVM training

Results

- Generally high F_1 scores
  - Classifier can confidently differentiate between waldo / wenda / negative
- waldo & negative scores opposite
  - Suggests the classifier can mainly differentiate between waldo / negative
  - Assumption backed up by empirical results
- Low F_1 scores for wizard
  - Possible due to few training examples and non-distinctive appearance

class	precision	recall	F_1 score	#samples
waldo	1.00	0.69	**0.82**	13
wenda	0.80	0.57	**0.67**	7
wizard	0.00	0.00	**0.00**	3
negative	0.69	1.00	**0.82**	20



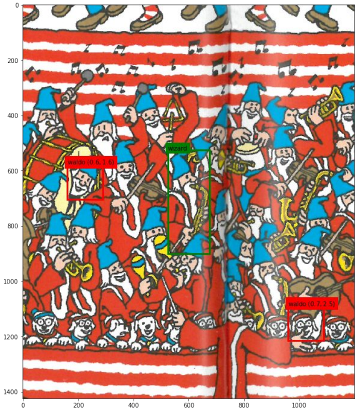
Evaluation

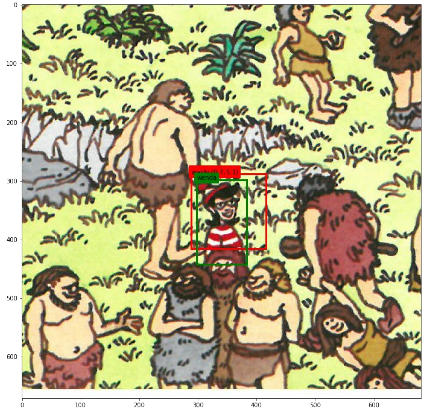
- Empirical evaluation with three hand picked samples (wizard, wenda, wlado)
- HOG & SVM
    - Zero IoU for first two images, giving low mean IoU
    - Faster object detection because of faster feature descriptor
- HOG / HOV & SVM
    - Higher mean IoU, suggesting better accuracy
    - Slower object detection, caused by higher dimensional feature descriptor & unoptimized code

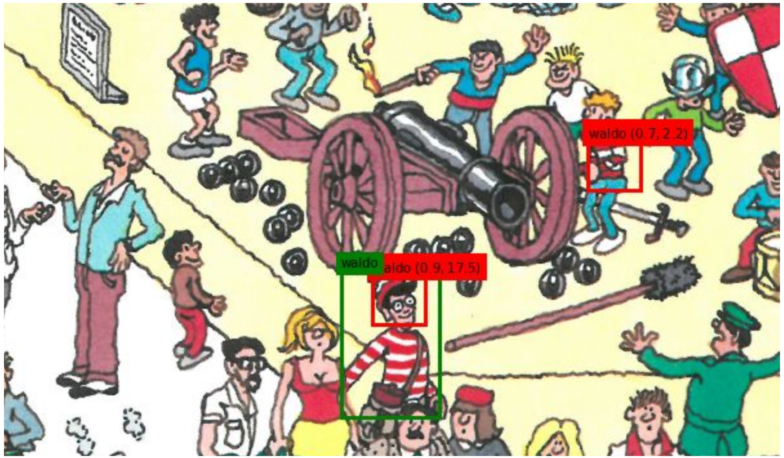
method	I_1 IoU	I_2 IoU	I_3 IoU	Mean IoU	Mean Time
HOG & SVM	0.0	0.0	0.37	**0.13**	12.8 s
HOG/HOV & SVM	0.0	0.5	0.09	**0.21**	21.3 s

Discussion

- Large visual similarity between ground truth and false positive bounding boxes
- Detector strongly picks up on horizontal red / white stripes pattern
- Future work:
  - Larger dataset / sample augmentation
  - Aspect ratio independent feature descriptor for arbitrary bounding boxes
  - Improve wizard classifier

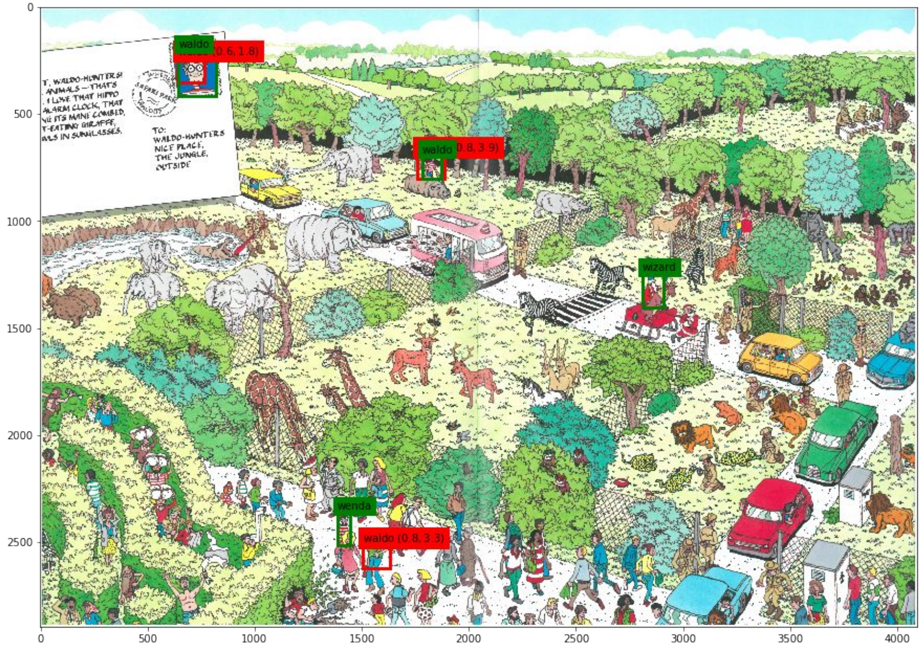






Conclusion

- Further work would aim to increase detection accuracy
- Current mAP (overlap: 0.1)
  - waldo: 0.12
  - wenda: 0.16
  - wizard: 0.08
- Considered different approaches and chose the most promising one
- Team communication worked well and everyone contributed to the final result



Thanks for your attention

Maximilian Fruehauf, David Drews, Choo Wen Xin
{e0445541, e0454245, e0053347}@u.nus.edu