Active Learning for Deep Detection Neural Networks

Hamed H. Aghdam, Abel Gonzalez-Garcia, Joost van de Weijer, Antonio M. L´opez

ICCV 2019

UCF

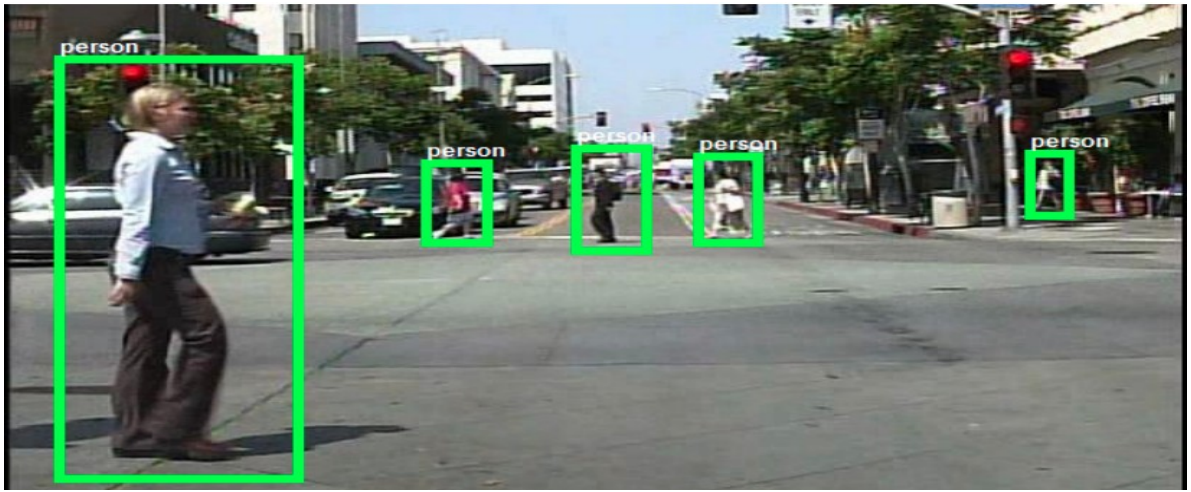
Overview

- Introduction to Active Learning
- Related work
- Proposed Method
- Datasets
- Evaluation
- Results
- Conclusion
- References



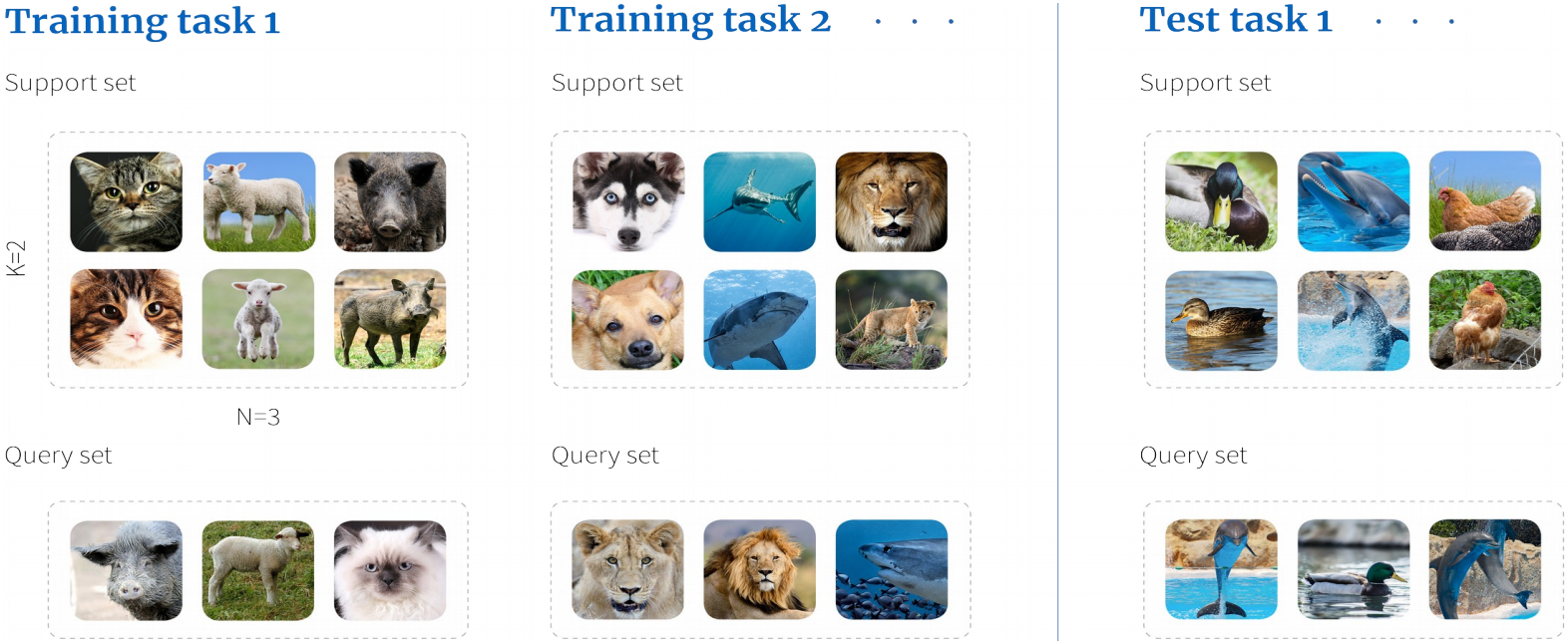
Motivation

- 35 seconds / image for pedestrian labeling
- Semi-supervised learning



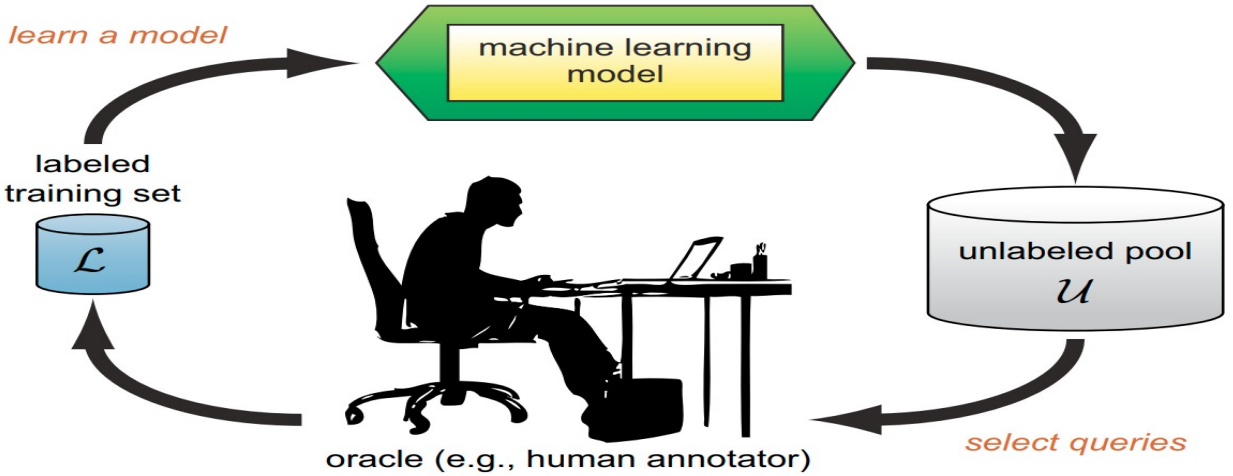
Labeling pedestrians for Caltech Pedestrian Dataset[11].

Few-shot learning



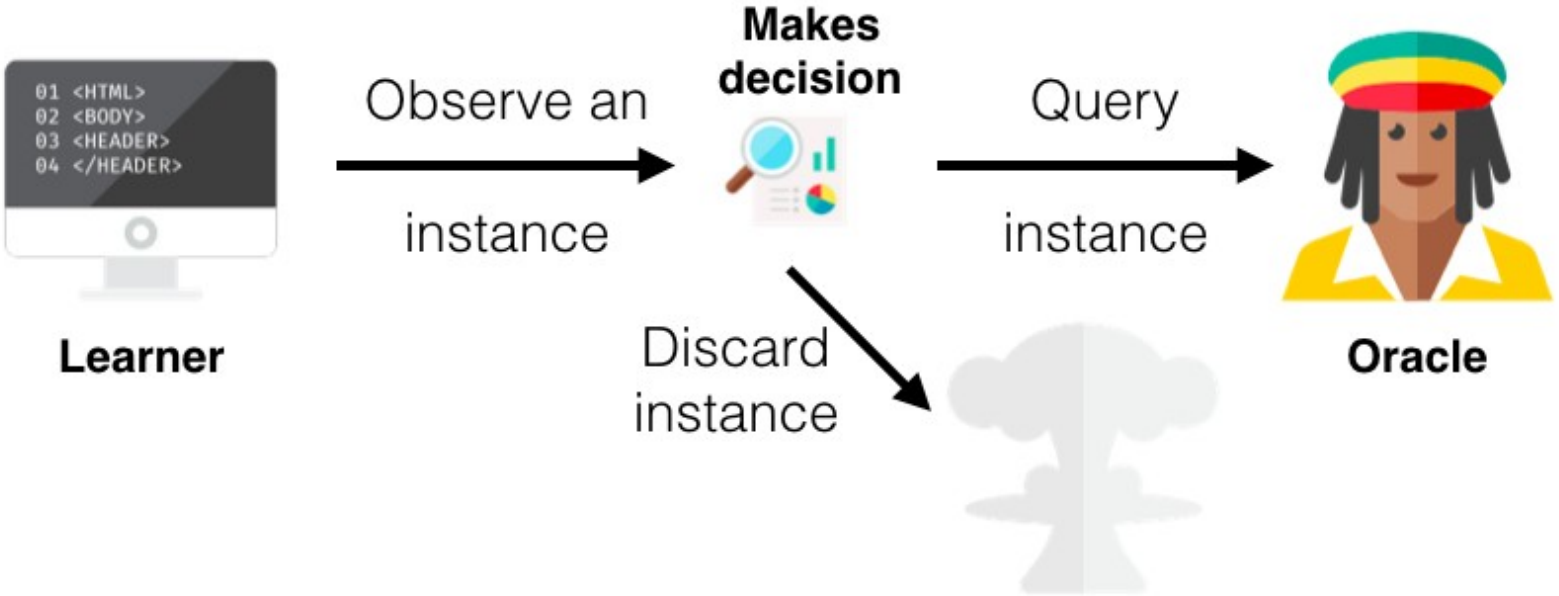
Few-shot learning using meta-learning[13].

Active learning



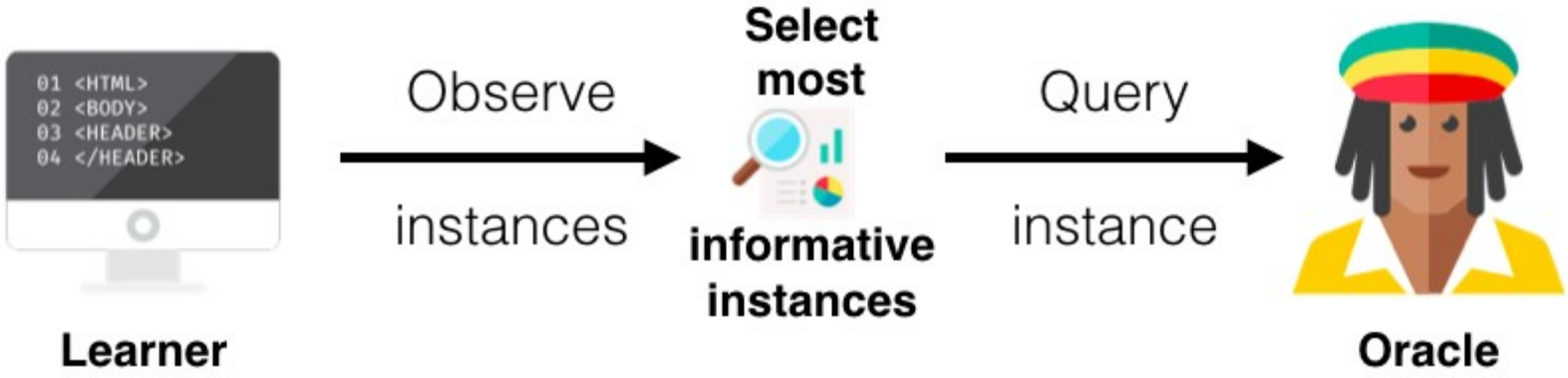
Active learning conceptual diagram[14].

Types of Active Learning – Stream based



Stream based Active learning[14].

Types of Active Learning – Pool based



Pool based Active learning[14].

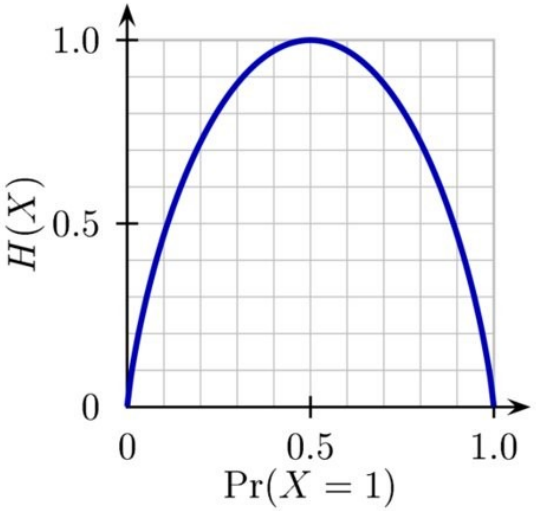
Types of Active Learning – Query synthesis based



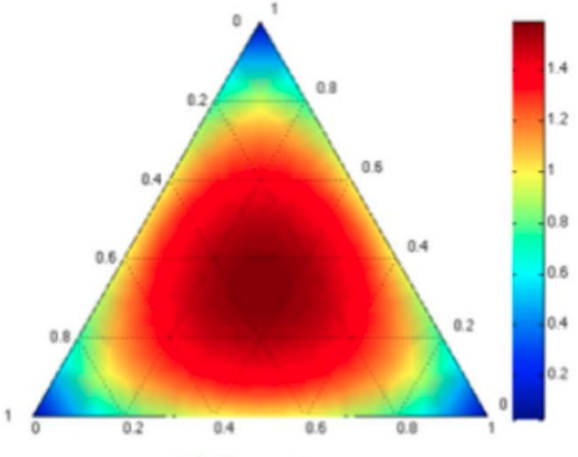
Query synthesis based Active learning[14].

Related Work

Uncertainity sampling – Maximum Entropy [17]



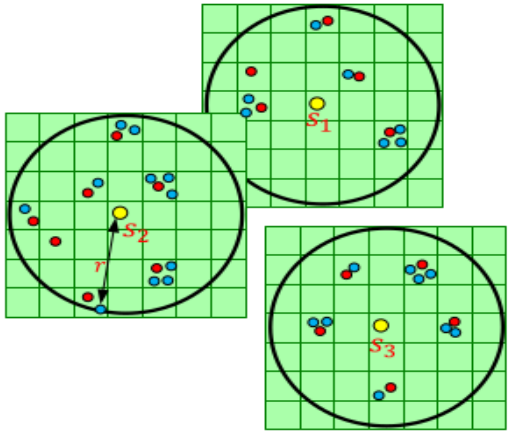
Binary Cross Entropy [15]



Cross Entropy for 3 variables [16]

Related Work

Representiveness – Coresets (Rohan et al. [4])



Coresets as k-center [15]

Related Work

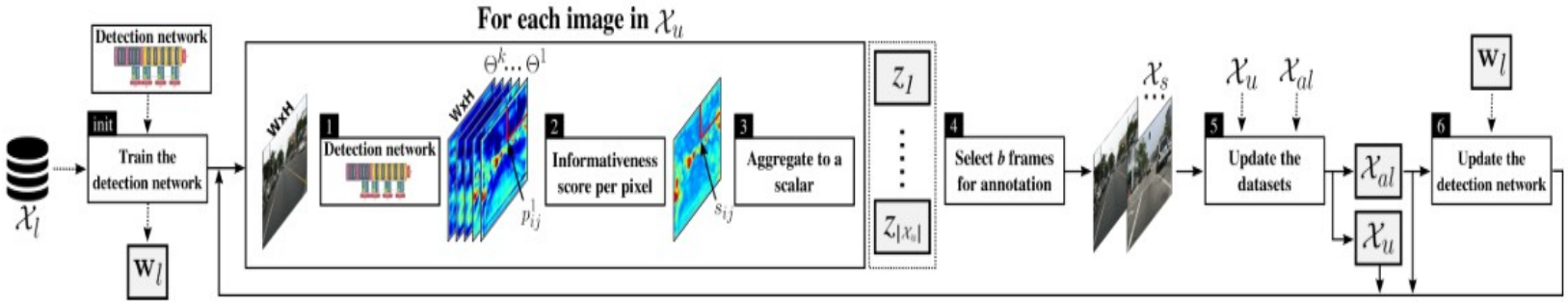
Active Learning for Object Detection (Kao et al.[5])



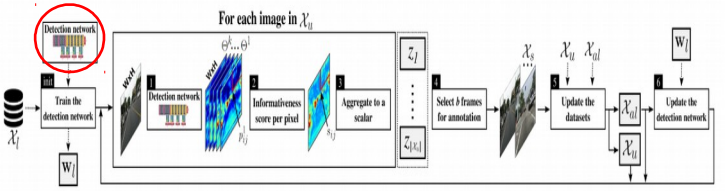
Calculation of stability scores [5]

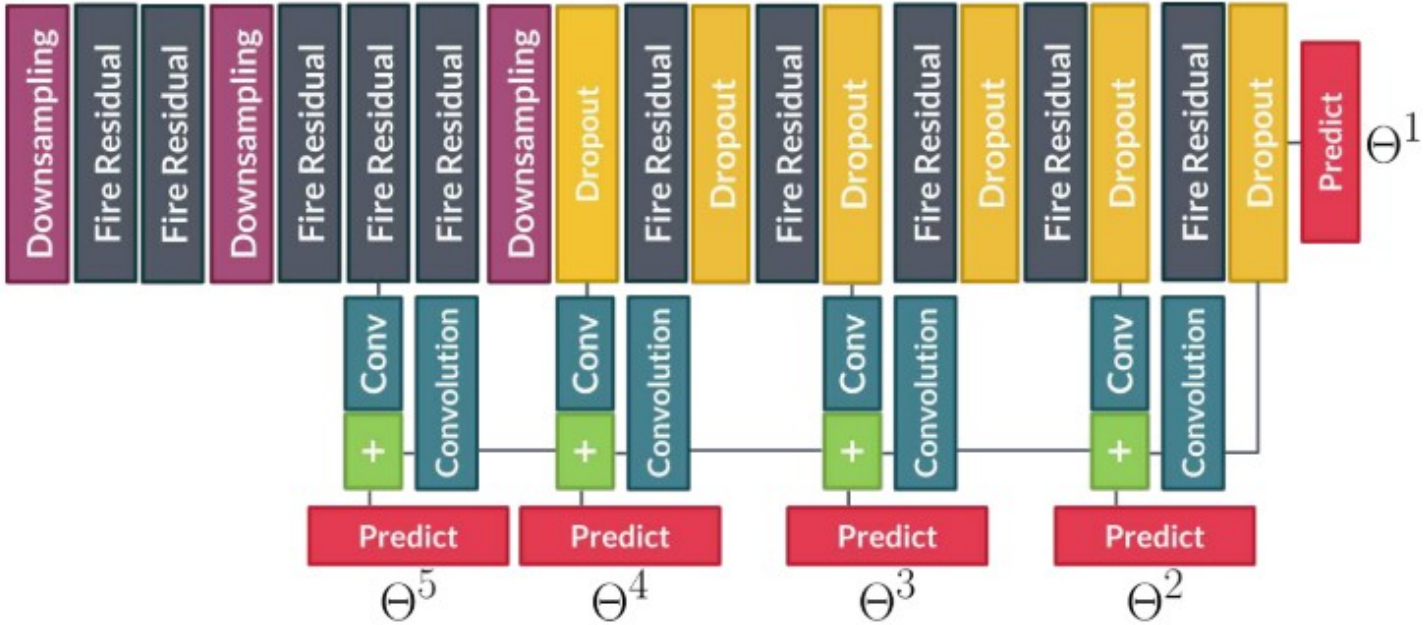
Proposed Method

- Use X_u to assign prediction probability for each pixel.
- Calculate per pixel score.
- Aggregate the score from pixel to image level.
- Select top-b images and create X_S .
- $X_{al} = X_{al} \cup X_S$.
- Train the model on X_{al} for T epochs.

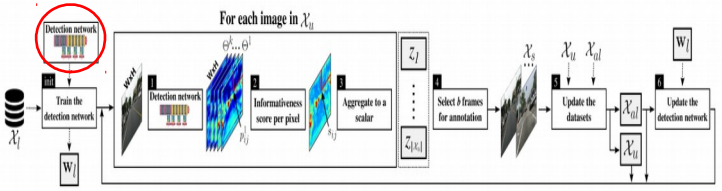


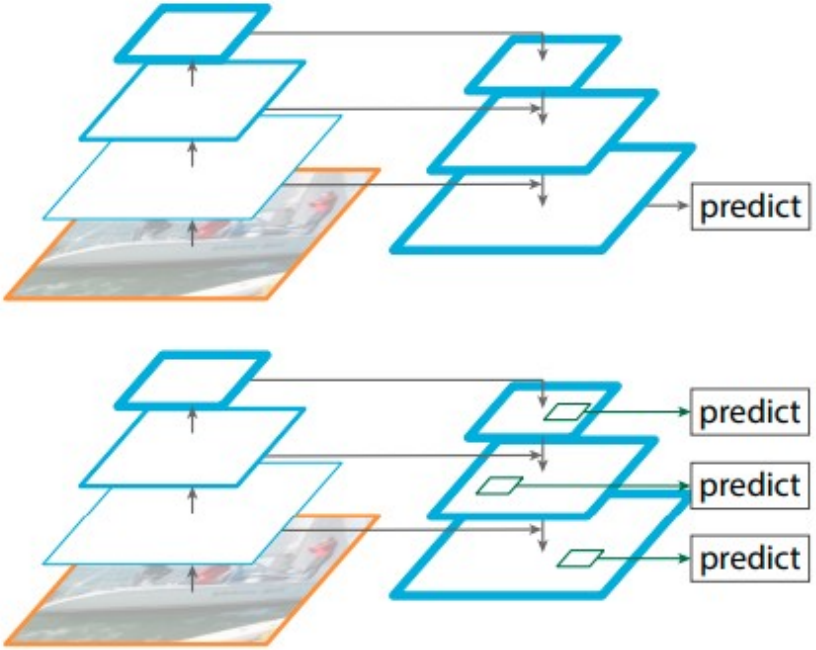
Network Architecture[1]



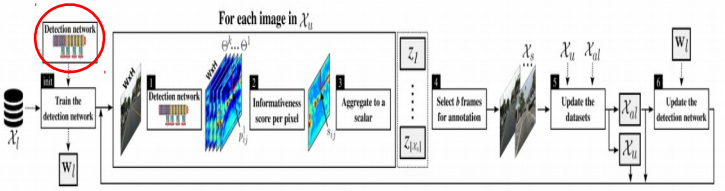


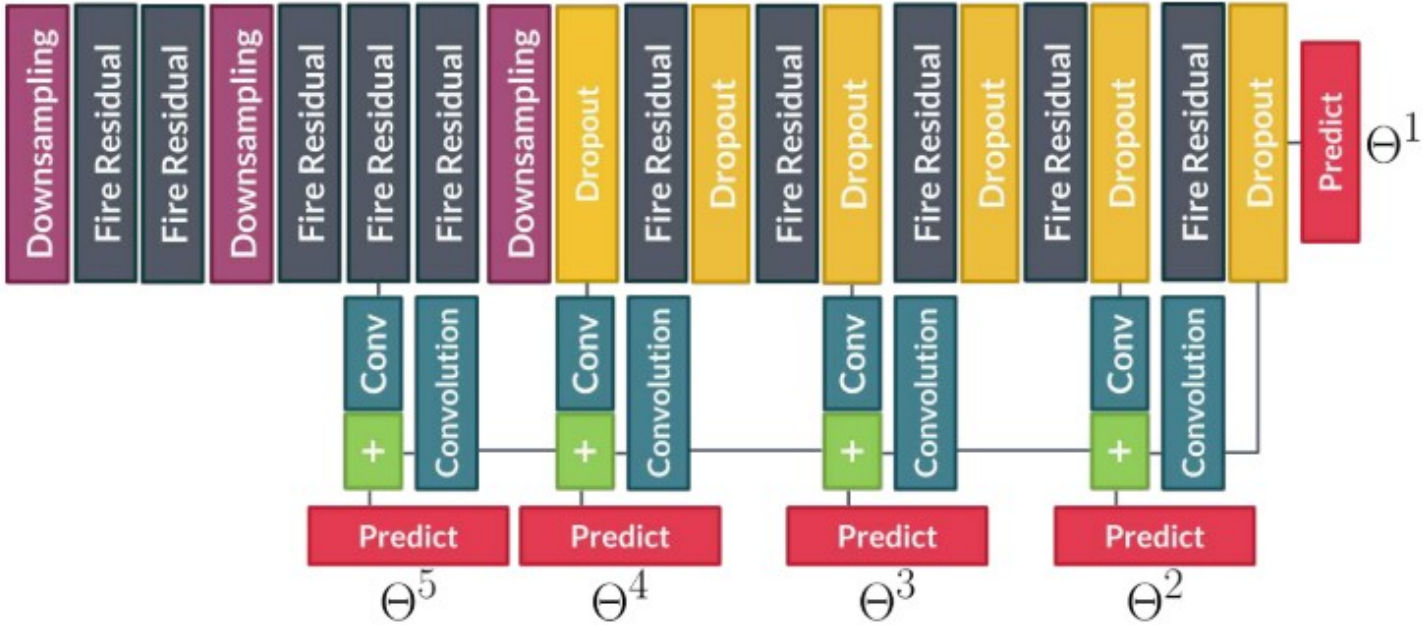
Feature Pyramid Network[8]



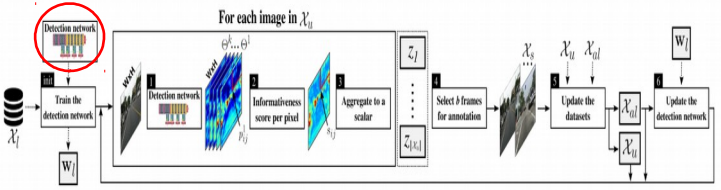


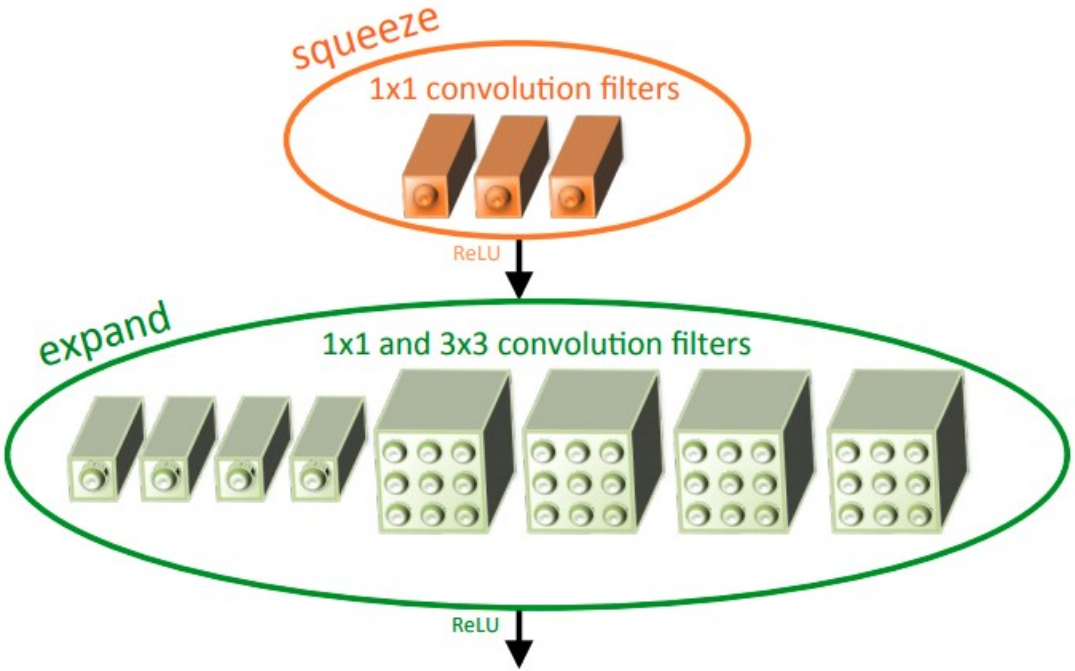
Network Architecture[1]



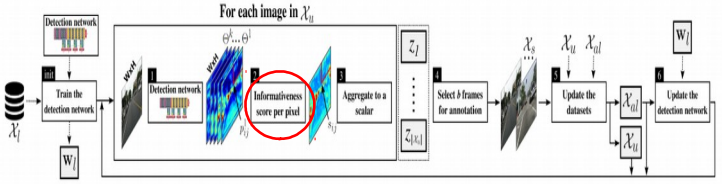


Fire Module(SqueezeNet)[9]



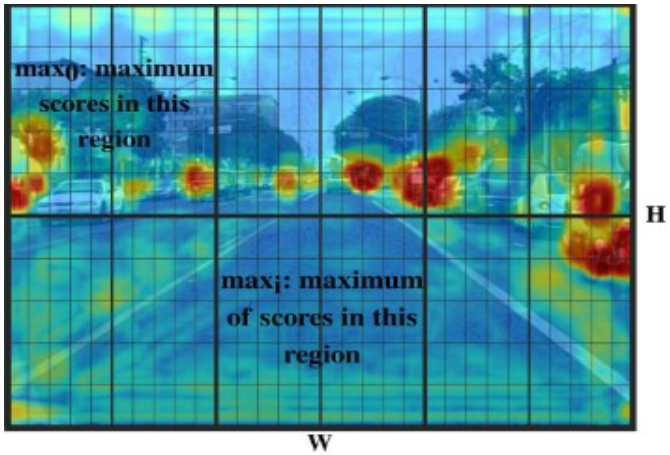


Pixel-level scores

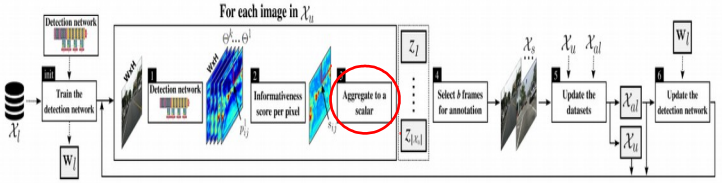


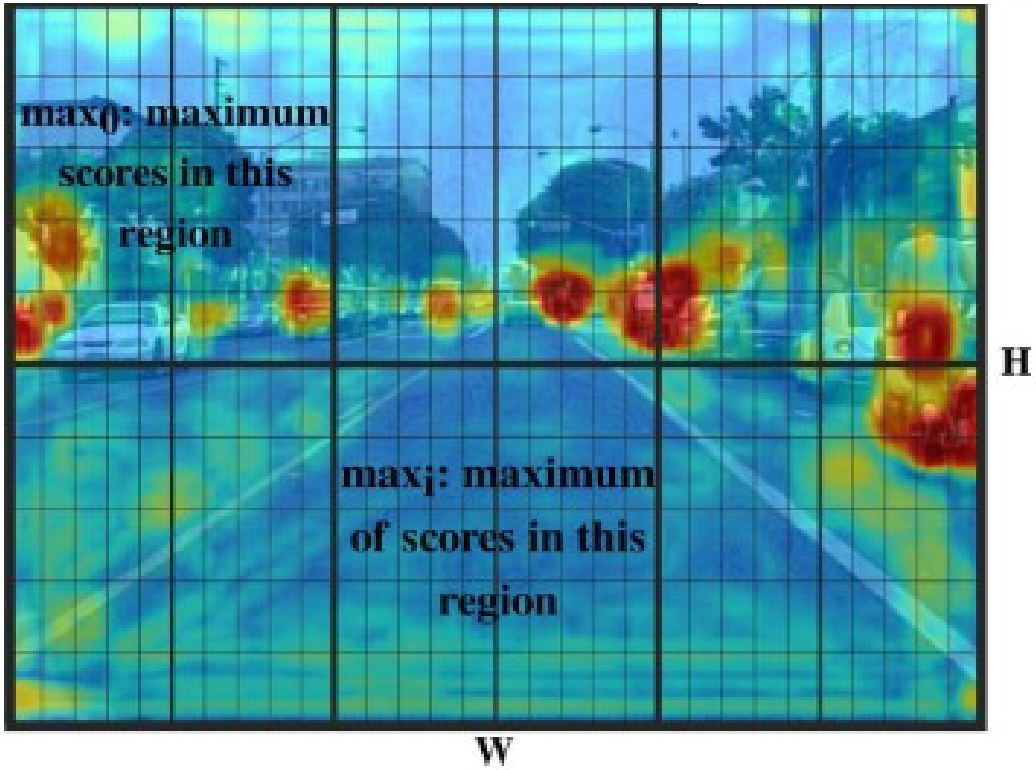
$$\hat{p}^k_{mn} = \frac{1}{(2r+1)^2} \sum_{i=m-r}^{m+r} \sum_{j=n-r}^{n+r} p^k_{ij}.$$

$$s^k_{mn} = \mathbb{H}(\hat{p}^k_{mn}) - \frac{1}{(2r+1)^2} \sum_{i=m-r}^{m+r} \sum_{j=n-r}^{n+r} \mathbb{H}(p^k_{ij})$$

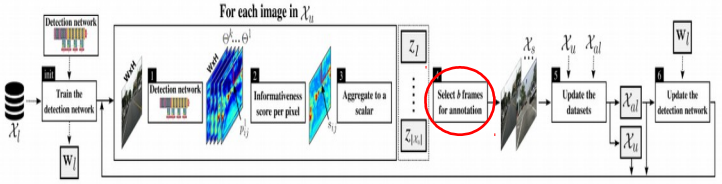


Aggregating scores

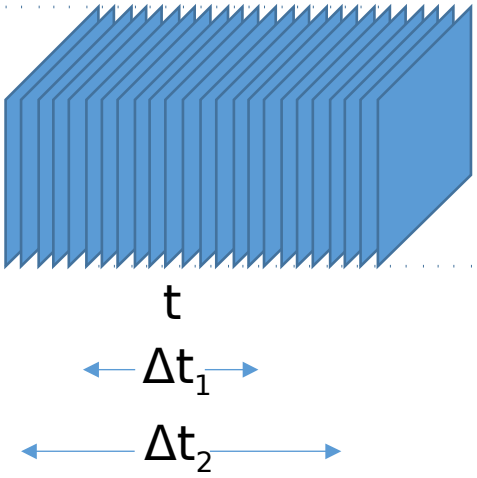




Selecting top-b images



$$\hat{z}_t = \frac{1}{\sum_i w_i} \sum_{i=t-\triangle t}^{t+\triangle t} w_{i+\triangle t} z_i.$$



Datasets

CityPersons[10]

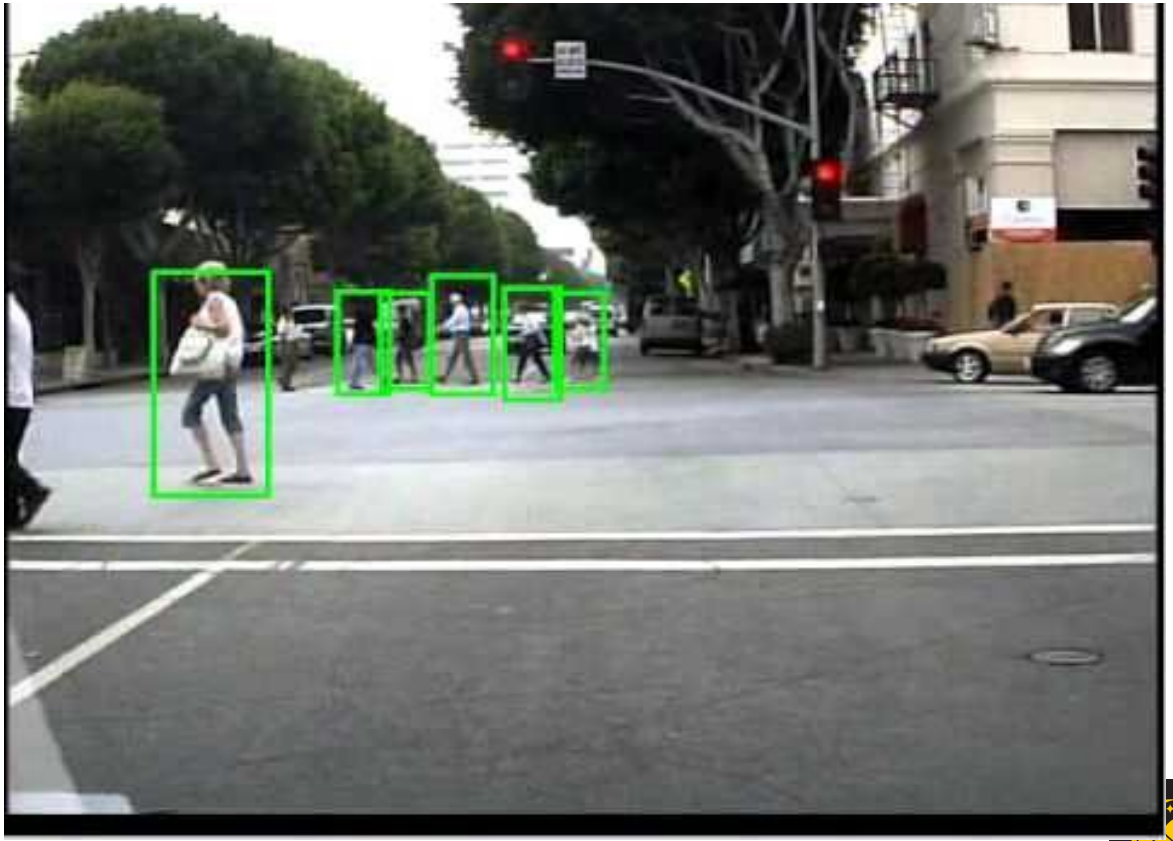
- Subset of CityScapes
- 2048x1024 resolution
- 1835 Images
- 7735 Boxes



Datasets

CaltechPed[11]

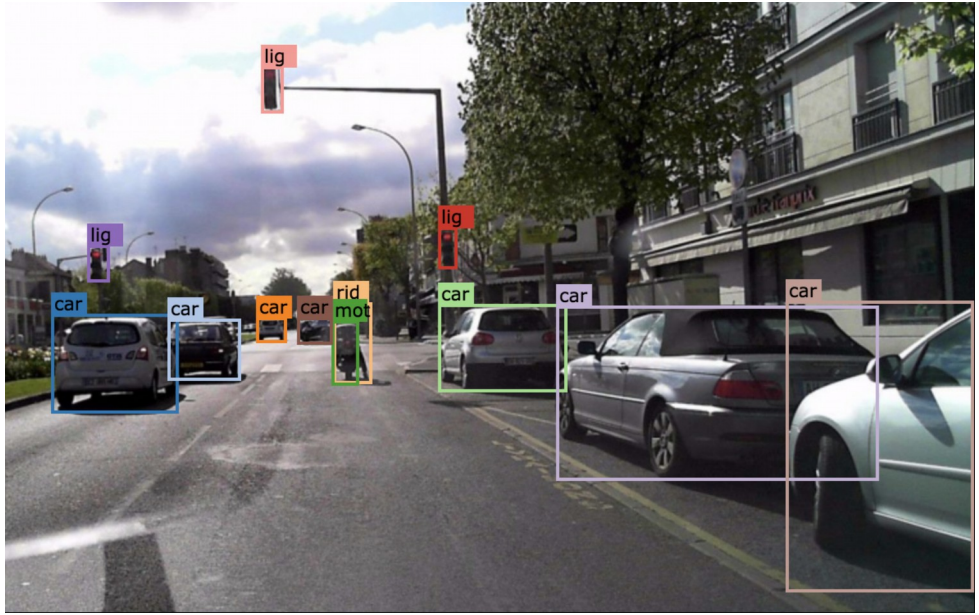
- 640x480 resolution
- 51363 Images
- 20062 Boxes
- Videos



Datasets

BDD100K[12]

- 1280x720 resolution
- 10 classes
- 69836 Images
- 86047 Boxes

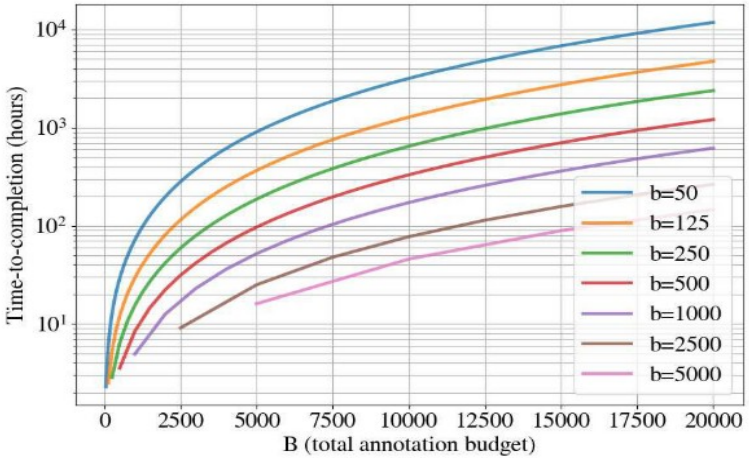


Evaluation

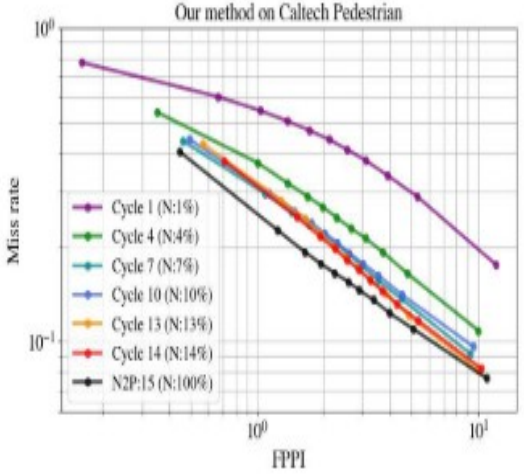
- Miss rate vs False positives per image
- Uniform random vs proposed method

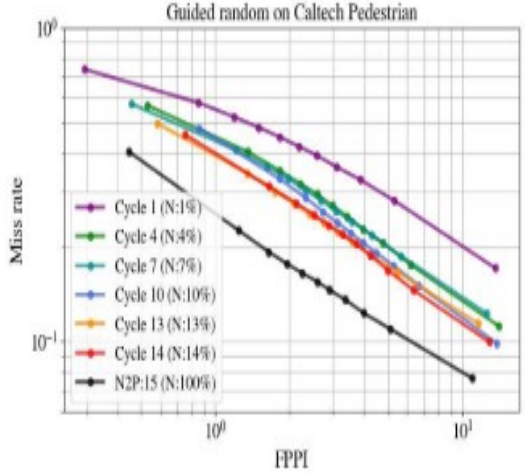
Results

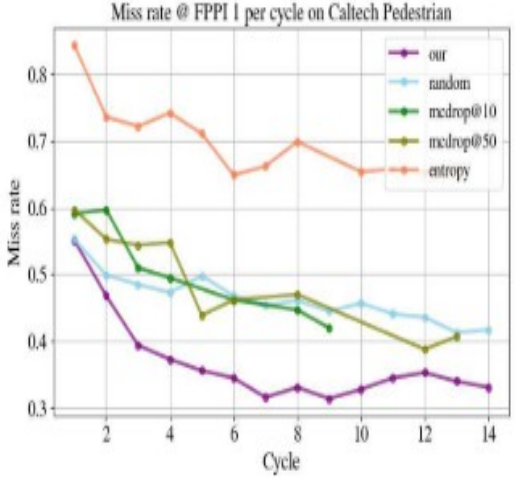
Time-to-completion



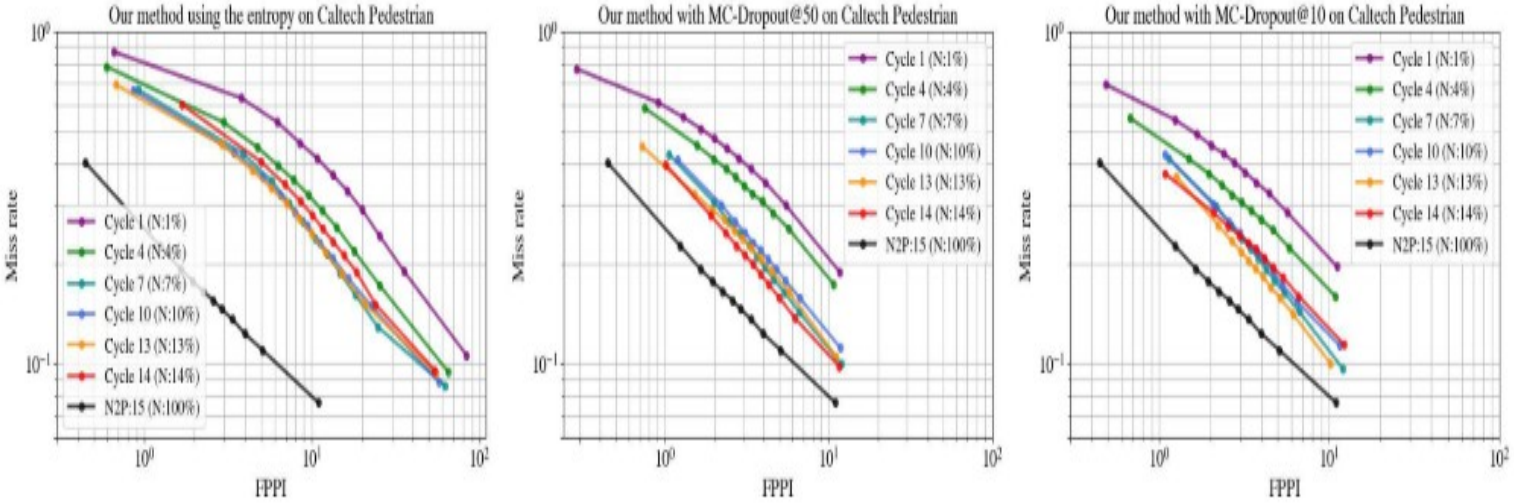
Random vs Proposed



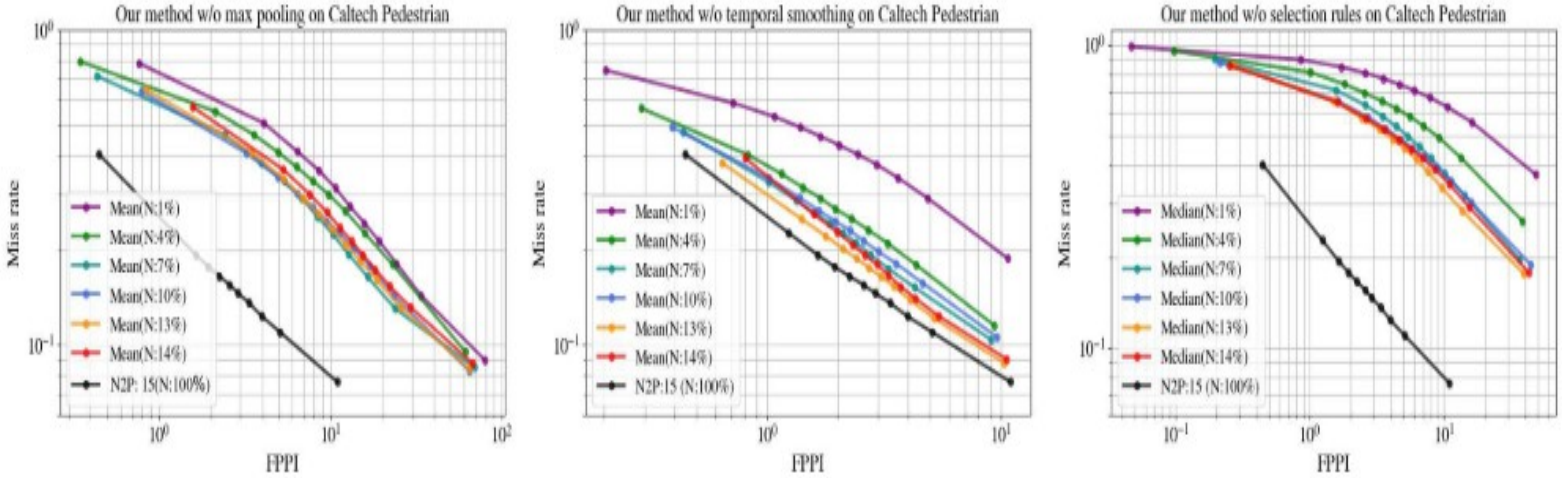




Scoring Function



Ablation study



Conclusion

- Good scoring function
- Aggregation method
- Works on videos and images
- Generalize to multiple classes



References

- [1]- Hamed H. Aghdam1, Abel Gonzalez-Garcia1, Joost van de Weijer1, Antonio M. L´opez. Active Learning for Deep Detection Neural Networks. In ICCV, 2019.
- [2] - Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep Bayesian Active Learning with Image Data. 2017.
- [3] - Ehsan Elhamifar, Guillermo Sapiro, Allen Yang, and S Shankar Sasrty. A convex optimization framework for active learning. In ICCV, 2013.
- [4] - Rohan Paul, Dan Feldman, Daniela Rus, and Paul Newman. In ICRA, 2014.
- [5] - Chieh-Chi Kao, Teng-Yok Lee, Pradeep Sen, and Ming-Yu Liu. Localization-aware active learning for object detection. CoRR, abs/1801.05124, 2018.
- [6] - Soumya Roy, Asim Unmesh, and Vinay P. Nambood- iri. Deep active learning for object detection. In BMVC, 2018.
- [7] - Clemens-Alexander Brust, Christoph K¨ading, and Joachim Denzler. Active learning for deep object de- tection. CoRR, abs/1809.09875, 2018.
- [8] - Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, Serge J. Belongie. Feature Pyramid Networks for Object Detection. CoRR, abs/1612.03144

UCF

References(Continued)

- [9] - Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. CoRR, abs/1602.07360, 2016.
- [10] - Shanshan Zhang, Rodrigo Benenson, and Bernt Schiele. Citypersons: A diverse dataset for pedestrian detection. In CVPR, 2017.
- [11] - Piotr Doll´ar, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian detection: An evaluation of the state of the art. TPAMI, 2012.
- [12] - Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, and Trevor Dar- rell. BDD100K: A diverse driving video database with scalable annotation tooling. CoRR, abs/1805.04687, 2018.
- [13] - https://www.borealisai.com/en/blog/tutorial-2-few-shot-learning-and-meta-learning-i/
- [14] - https://www.datacamp.com/community/tutorials/active-learning
- [15] - https://en.wikipedia.org/wiki/Binary_entropy_function
- [16] - https://www.cs.cmu.edu/~tom/10701_sp11/recitations/Recitation_13.pdf
- [17] - Ido Dagan and Sean P. Engelson. 1995. Committee-based sampling for training probabilistic classifiers. In Proceedings of the Twelfth International Conference on International Conference on Machine Learning (ICML'95). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 150–157.

UCF



Thank you.

UCF