Active Learning for Deep Detection Neural Networks

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

ICCV 2019
Overview

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

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

Uncertainty 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

\[ \hat{p}_{mn}^k = \frac{1}{(2r + 1)^2} \sum_{i=m-r}^{m+r} \sum_{j=n-r}^{n+r} p_{ij}^k. \]

\[ s_{mn}^k = H(\hat{p}_{mn}^k) - \frac{1}{(2r + 1)^2} \sum_{i=m-r}^{m+r} \sum_{j=n-r}^{n+r} H(p_{ij}^k). \]
Aggregating scores

\[ \text{max}_i: \text{maximum scores in this region} \]

\[ \text{max}_j: \text{maximum of scores in this region} \]
Selecting top-b images

\[ \hat{z}_t = \frac{1}{\sum_i w_i} \sum_{i=t-\Delta t}^{t+\Delta t} w_i + \Delta t \hat{z}_i. \]
Datasets

CityPersons[10]
- Subset of CityScapes
- 2048x1024 resolution
- 1835 Images
- 7735 Boxes
Datasets

- 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

The graph illustrates the relationship between time-to-completion and the total annotation budget (B). Each line represents a different value of the parameter b, as indicated by the legend. The logarithmic scale on the y-axis and linear scale on the x-axis show how increasing the budget affects the time required to complete the task.
Random vs Proposed
Scoring Function
Ablation study
Conclusion

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

References (Continued)

Thank you.