Loss Rescaling by Uncertainty Inference for Single-stage Object Detection

*Yan Li\textsuperscript{1}, Xiaoyi Chen\textsuperscript{1}, Li Quan\textsuperscript{1}, Ni Zhang\textsuperscript{1}

\textsuperscript{1}NEC Laboratories, China
*Presenter

The 27th IEEE International Conference on Image Processing (ICIP 2020)
Session: TEC-04 -- Machine Learning for Image and Video Processing II
Paper: TEC-04.6
Learning Convolutional Object Detectors with Biases

• Deep learning model may fit biasedly to examples that appear more frequently or are easier to be distinguished.

• In object detection tasks, negative samples often dominate in training because the background areas are larger than the foreground objects, and most of them have similar appearances.

• To guide the convolutional object detectors learning with different difficulty levels:
  • Hard example mining
  • Focal loss
  • ...

Is learning difficulty involves only with the prediction probability?
Modeling Uncertainty for Deep Neural Network

- Approximation to the Bayesian uncertainty estimation
  - Monte-Carlo Dropout
  - Variational Bayesian inference

- Relations between uncertainty and model learning
  - Bayesian Segnet demonstrates strong inverse relationship between segmentation accuracy and model uncertainty.
  - Salman Khan et al re-adjust classification boundaries for biased data by uncertainty estimates.

*Uncertainty estimates are correlated with the difficulty level of input samples.*
Relating Difficulty to Uncertainty Estimates and Prediction Correctness

- Learning difficulty of positive anchors for single-stage object detectors
Bayesian Convolution Layer ("convBayes")

- Assume the weights of the convolution kernel: \( w \sim N(k, \lambda k^2) \).

- Formulaic as,

\[
Y = \frac{1}{K} \sum_{i=1}^{K} \left( x \odot k + \varepsilon \sqrt{x^2 \odot (\lambda k^2)} \right), \quad (1)
\]

The output follows a Gaussian distribution:

\[
\begin{align*}
\mu &= x \odot k \\
\sigma &= \sqrt{x^2 \odot (\lambda k^2)}
\end{align*}
\]
Bayesian Active Learning by Disagreement (BALD)

• Using the estimated mean $\mu$ and variance $\sigma$, the uncertainty score $U_{\text{BALD}}$ is calculated by method proposed by Neil Houlsby et al.,

$$U_{\text{BALD}} = h \left( \rho \left( \frac{\mu}{\sqrt{\sigma^2 + 1}} \right) - \left( \frac{C}{\sqrt{\sigma^2 + C^2}} \right) \exp \left( - \frac{\mu^2}{2(\sigma^2 + C^2)} \right) \right), \quad (2)$$

where $C = \sqrt{\pi \ln 2}/2$. 

Binary entropy function

Gaussian cumulative density
Difficulty Scoring

- Difficulty map is defined by relations between uncertainty score and prediction correctness

\[
D = \begin{cases} 
0.5 \times U & \text{if correct} \\
0.5 \times (1.0 - U) + 0.5 & \text{otherwise,} 
\end{cases}
\]  

where \( U = \text{sigm}(U_{\text{BALD}}) \).

For object detection tasks, we assume that:
1. Correct predictions (TP and TN) are always less difficult than incorrect predictions (FP and FN).
2. For correct predictions, higher uncertainties reveal higher difficulty.
3. For incorrect predictions, lower uncertainties reveal higher difficulty.
Network Architecture

- Using “convBayes” layer as auxiliary branch to estimate model uncertainty during training.
Experiments on VOC datasets

• Dataset
  • Pascal VOC with 20 categories annotations of foreground objects
  • Training subset: VOC07 trainval + VOC12 trainval (16,551)
  • Testing subset: VOC07 test (4,952)

• Training details
  • Data augmentation: HSV random distortion, random cropping and flipping
  • Input size: 608x608 with zero padding
  • Backbone network: Darknet53

• Evaluation metric
  • Mean average precision (mAP)
Comparison of Results & Ablation Study

- Evaluation results of object detection on VOC07test. (Darknet53 as backbone network)

<table>
<thead>
<tr>
<th>Name</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>76.62%</td>
</tr>
<tr>
<td>Obj_Focal loss</td>
<td>75.18%</td>
</tr>
<tr>
<td>Obj_Anchor loss</td>
<td>76.43%</td>
</tr>
<tr>
<td>Uncertainty weighted loss (Ours)</td>
<td>77.34%</td>
</tr>
</tbody>
</table>

For ablation study

<table>
<thead>
<tr>
<th>Name</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline w/ convBayes</td>
<td>76.38%</td>
</tr>
<tr>
<td>Baseline w/ additional conv</td>
<td>76.01%</td>
</tr>
</tbody>
</table>

Evaluation results of object detection on VOC07test. (Darknet53 as backbone network)
Additional Experiments on Image Classification Task

• Architecture slightly changed for image classification task

![Architecture Diagram]

• Training details
  • CIFAR-10 with 10 categories of image level annotations
  • Backbone network: Resnet50

• Evaluation metric
  • Top-1 classification accuracy
Comparison of Results & Ablation Study

- Evaluation results of image classification on CIFAR-10. (Resnet50 as backbone network)

<table>
<thead>
<tr>
<th>Name</th>
<th>Top 1 Accu.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
<td>92.88 ± 0.33%</td>
</tr>
<tr>
<td>Focal loss</td>
<td>93.07 ± 0.49%</td>
</tr>
<tr>
<td>Anchor loss</td>
<td>93.13 ± 0.99%</td>
</tr>
<tr>
<td>Uncertainty weighted loss (Ours)</td>
<td>93.56 ± 0.37%</td>
</tr>
</tbody>
</table>

For ablation study

| Baseline w/ convBayes                                    | 92.73 ± 1.13%        |
Some Visualized Results

Detection results

Uncertainty map
(each pixel indicates the BALD of corresponding anchor)

Detection results

Uncertainty map
(each pixel indicates the BALD of corresponding anchor)
Conclusion & Discussion

- The uncertainty estimates of deep neural network are correlated with prediction difficulty and can be used to guide model learning for better performance.

- The limited improvements brought by proposed uncertainty weighted loss module might because there is only one Bayesian convolution layer in current architecture. In future works, a multi-Baysian convolution layer design might bring more accurate uncertainty estimates.

- The uncertainty inference module should be supported by a pertinently designed loss and merged with loss rescaling mechanism so as to further enlarge the variance of each predictions and reduce computation redundancies.
Loss Rescaling by Uncertainty Inference for Single-stage Object Detection

Yan Li, Xiaoyi Chen, Li Quan, Ni Zhang
NEC Laboratories, China
The 27th IEEE International Conference on Image Processing (ICIP 2020)

Thanks.