Loss Rescaling by Uncertainty Inference for Single-stage Object Detection

*Yan Li,[1] Xiaoyi Chen,[1] Li Quan,[1] Ni Zhang[1]

[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

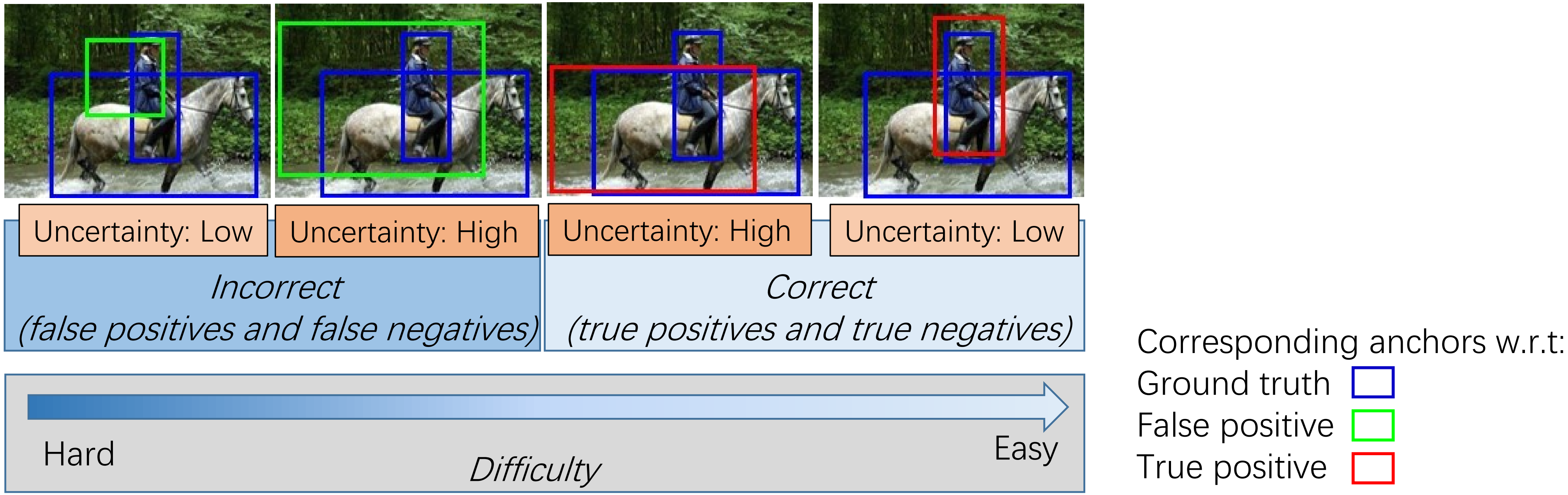
- Approximation to the Bayesian uncertainty estimation
  - Monte-Carlo Dropout
  - Variational Bayesian inference

- 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.

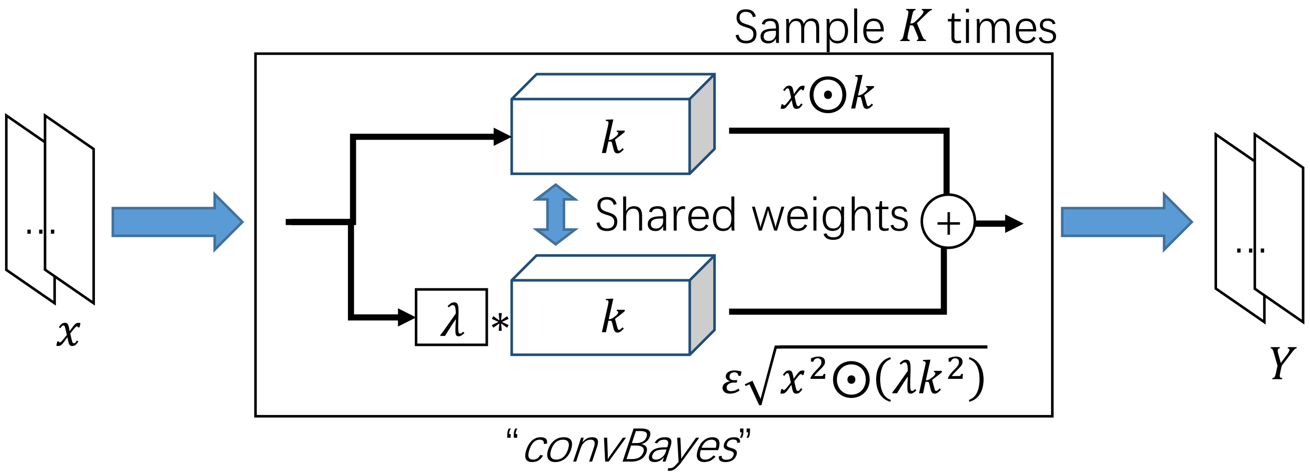
Relating Difficulty to Uncertainty Estimates and Prediction Correctness

- Learning difficulty of positive anchors for single-stage object detectors



Bayesian Convolution Layer (“*convBayes*”)

- Assume the weights of the convolution kernel: $w \sim N(k, \lambda k_2)$.



The output follows a Gaussian distribution:

$$\begin{cases} \mu = x \odot k \\ \sigma = \sqrt{x^2 \odot (\lambda k^2)} \end{cases}$$

- Formulaic as,

Sampled from N(0, 1)

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

Convolution opt.

Bayesian Active Learning by Disagreement (BALD)

- Using the estimated mean μ and variance σ, the uncertainty score U_{BALD} is calculated by method proposed by Neil Houlsby et al,.

Binary entropy function

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

Gaussian cumulative density

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

Difficulty Scoring

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

$$D = \begin{cases} 0.5 * U & if\ correct \\ 0.5 * (1.0 - U) + 0.5 & otherwise, \end{cases} \quad (3)$$

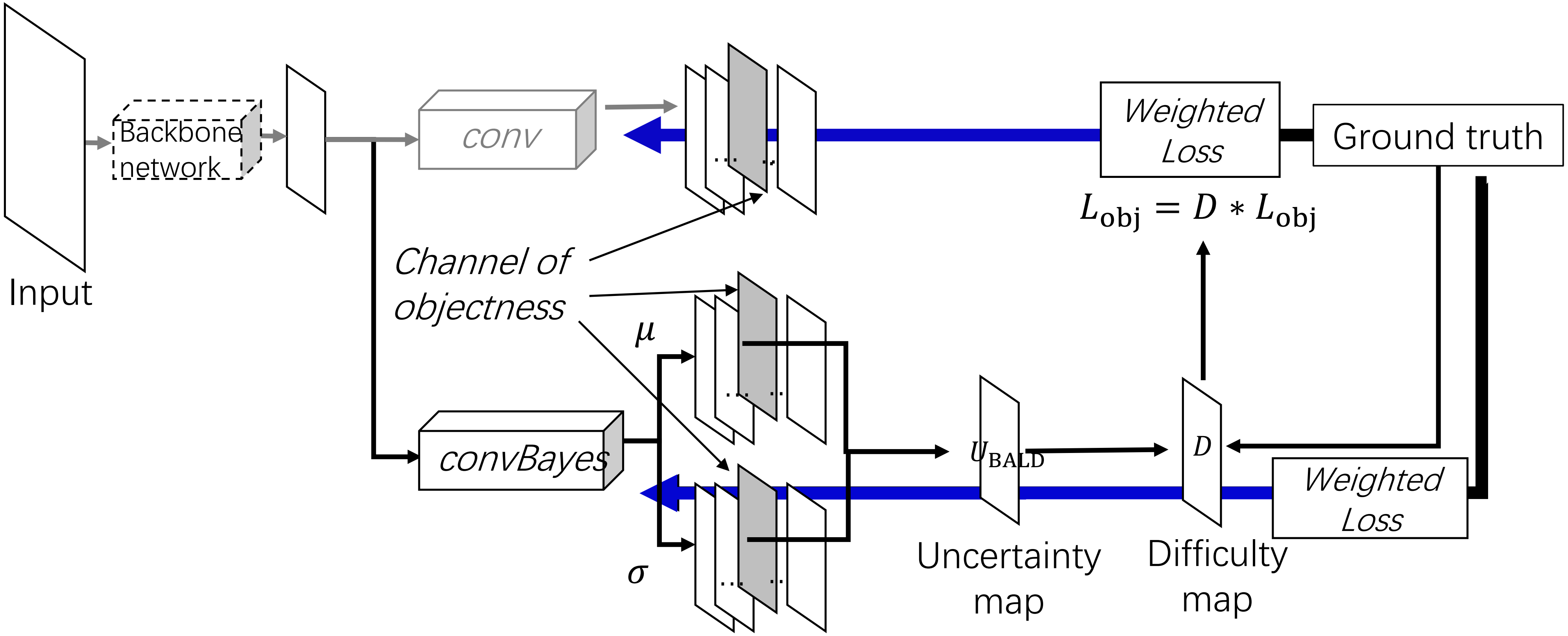
where $U = sigm(U_{\mathrm{BALD}})$.

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**)

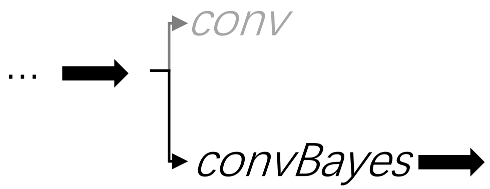
- Training details
  - Data augmentation: HSV random distortion, random cropping and flipping
  - Input size: 608x608 with zero padding
  - Backbone network: Darknet53

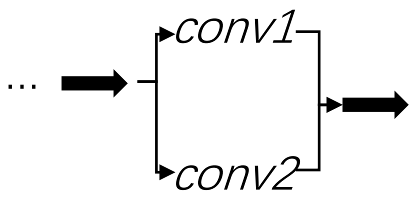
- Evaluation metric
  - Mean average precision (mAP)

Comparison of Results & Ablation Study

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

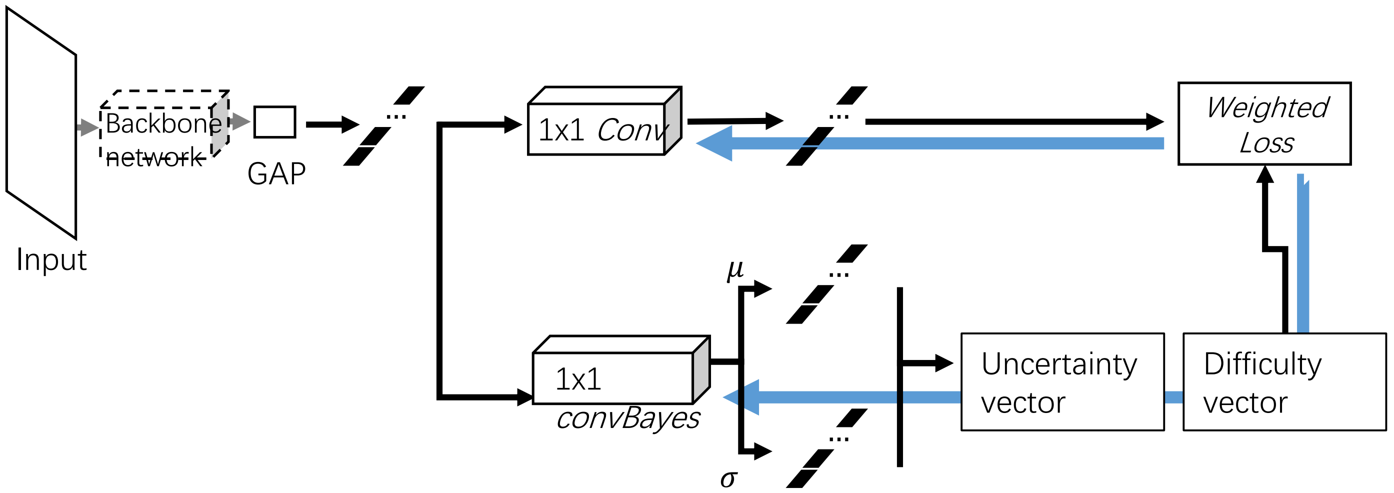
Name	**mAP**
Baseline	76.62%
Obj_Focal loss	75.18%
Obj_Anchor loss	76.43%
Uncertainty weighted loss (Ours)	**77.34%**
For ablation study	
Baseline w/ *convBayes*	76.38%
Baseline w/ additional *conv*	76.01%





Additional Experiments on Image Classification Task

- Architecture slightly changed for image classification task

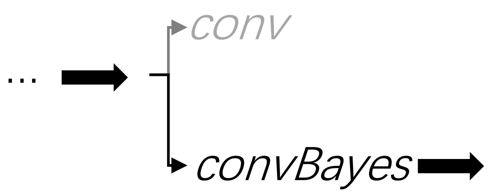


- Training details
  - CIFAR-10 with 10 categories of image level annotations
  - Backbone network: Resnet50
- Evaluation metric
  - Top-1 classification accuracy

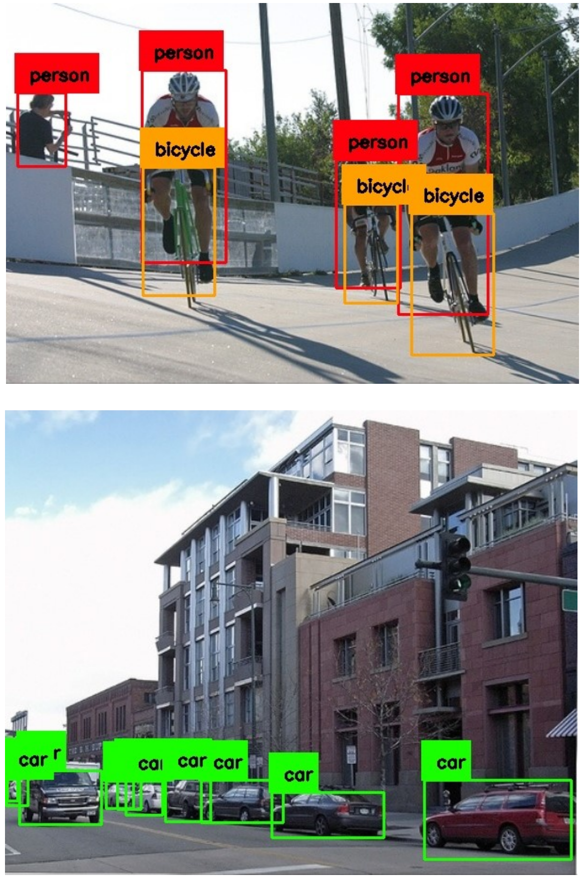
Comparison of Results & Ablation Study

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

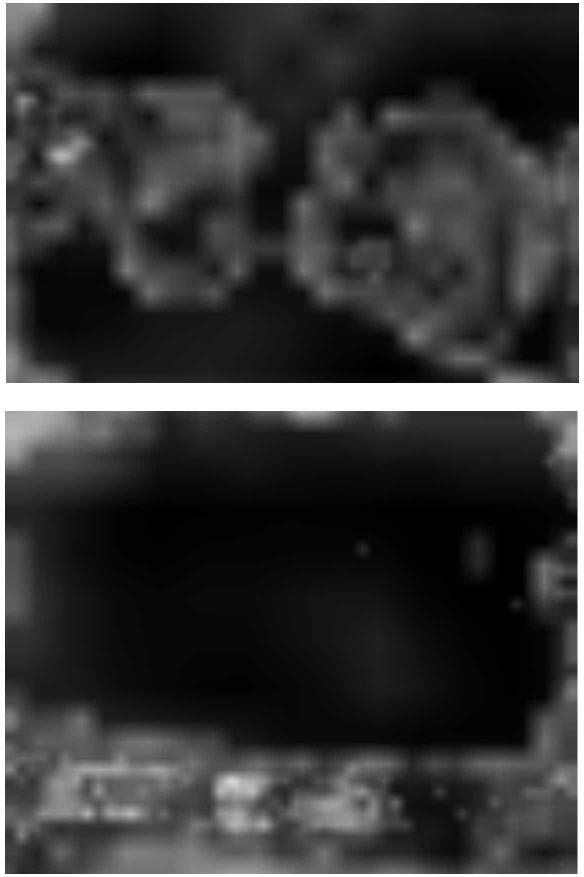
Name	Top 1 Accu.
CE	92.88 ± 0.33%
Focal loss	93.07 ± 0.49%
Anchor loss	93.13 ± 0.99%
Uncertainty weighted loss (Ours)	**93.56 ± 0.37%**
For ablation study	
Baseline w/ *convBayes*	92.73 ± 1.13%



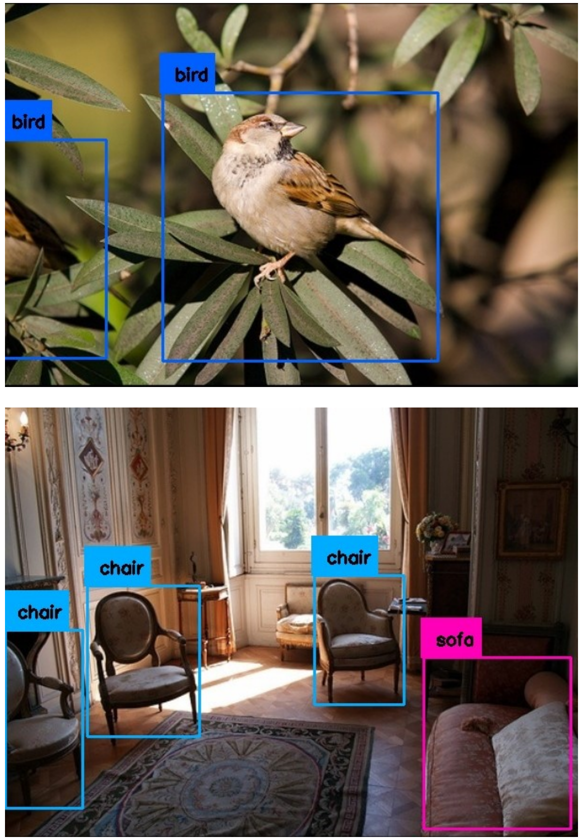
Some Visualized Results



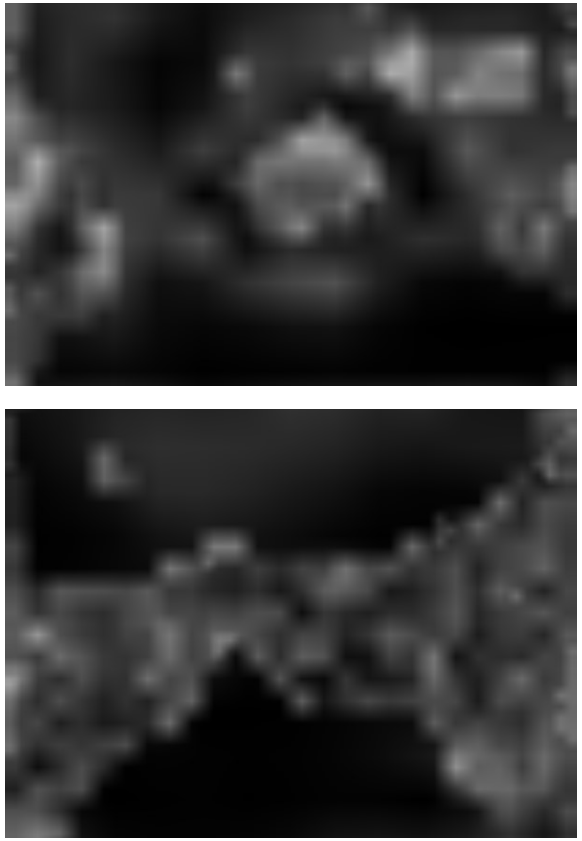
Detection results



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.*

Thanks.

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)