Supplementary Material: Uncertainty Estimation and Sample Selection for Crowd Counting

Viresh Ranjan[1], Boyu Wang[1], Mubarak Shah[2], and Minh Hoai[1]

[1] Department of Computer Science, Stony Brook University, Stony Brook, NY 11790
[2] University of Central Florida, Orlando, FL 32816

1 Overview

Earlier works [1] have shown the usefulness of both Aleatoric and Epistemic Uncertainty estimates for various Computer Vision tasks. In the main paper, we presented our CTN architecture for estimating aleatoric uncertainty pertaining to crowd density prediction. In Sec. 2 in the Supplementary, we present Monte Carlo CTN(MC-CTN), our architecture for estimating epistemic uncertainty. Subsequently in Sec. 3, we compare epistemic and aleatoric uncertainty estimates, and show that aleatoric uncertainty is more correlated with the prediction error. In the main paper, we focused on using aleatoric uncertainty estimates for sample selection since our preliminary experiments showed Aleatoric uncertainty to be more effective at sample selection compared with epistemic uncertainty.

In the main paper, we pointed out that our sampling strategy can be used for picking out images from a large pool of unlabeled images, and it can also be used to pick out informative crops from an image. Depending on the annotation budget, it might be useful to get partial annotations for an image by picking out informative crops from an image and getting human annotations for the informative crops rather than annotating the entire image. The experiments on image level sample selection are discussed in the main paper. In Sec. 4 of the supplementary material, we present experiments pertaining to selecting most informative crops from an image for human annotation.

Finally, for image level sample selection, in addition to experiments performed on Shanghaitech Part A [2] and NWPU [3] datasets as shown in the main paper, we present more experiments on Shanghaitech Part B in Sec. 5.

2 Epistemic Uncertainty Estimation

The CTN architecture described in the main paper captures the aleatoric uncertainty, i.e., the uncertainty inherent in the input data. In this section, we present a variant of CTN which can capture the model uncertainty, which is also known as Epistemic Uncertainty and arises due to the use of finite training data. Epistemic uncertainty, can be captured by Bayesian Neural Networks [4]. Such networks assume a prior distribution $\mathbb{P}(\theta)$ over the parameters of the network, and find the posterior probability $\mathbb{P}(\theta|\{X, Y\})$ conditioned on the training data $\{X, Y\}$. However, it is computationally expensive to perform inference

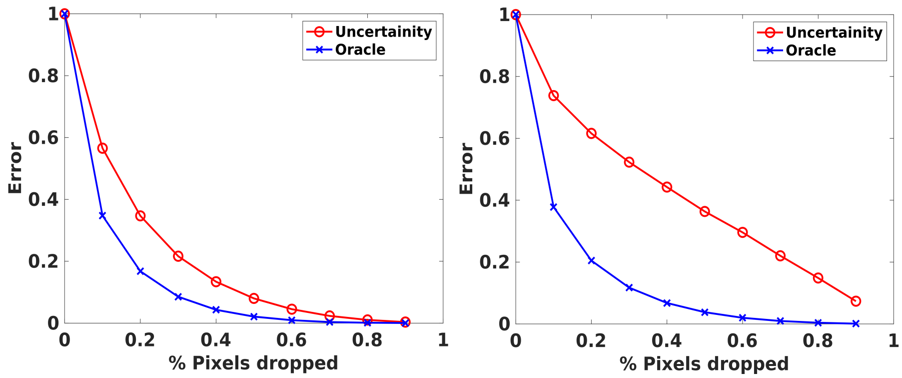

Fig. 1: The sparsification plots for aleatoric (left) and epistemic (right) uncertainty estimation on the test set of Shanghaitech Part A. The closer the uncertainty curve to the oracle, the higher the correlation between the uncertainty and prediction error. The area between the uncertainty curve and the oracle curve for the two cases is 0.07 and 0.25. The aleatoric uncertainty is more correlated with the prediction error.

with Bayesian Neural Networks with large number of parameters. To address this issue, [5] present a variational approximation which corresponds to adding a dropout [6] layer before every weight layer. The uncertainty in such a network can be obtained by doing multiple forward passes of the input image through the network, and computing the sample mean as the prediction output. The sample variance is the epistemic uncertainty of the prediction. This approach requires running the network multiple times with different dropout instantiations, so we will refer to it as Monte Carlo Uncertainty.

We will refer to a counting network where Predictive Uncertainty is replaced by Monte Carlo Uncertainty as Monte Carlo CTN (MC-CTN). MC-CTN is similar to CTN presented in the main paper with two major differences: 1) we remove the Predictive Uncertainty Estimation branch, and 2) we add a dropout layer before every convolution layer except for the ones in density prediction branch.

Performance of MC-CTN on Crowd Counting MC-CTN led to MAE/RMSE of 102/180 on UCF-QNRF dataset. CTN, as reported in the main paper, outperforms MC-CTN and results in MAE/RMSE of 86/146.

3 Comparing Aleatoric and Epistemic Uncertainties

Sparsification plots provide a way to ascertain whether the uncertainty estimate is correlated with the prediction error [7]. Such plots are obtained by removing pixels with the highest uncertainty and measuring the error for the remaining pixels. If the predicted uncertainty is correlated with the error, the overall error should monotonically decrease as we remove more uncertainty pixels. We show the sparsification plots for the Aleatoric and Epistemic uncertainty estimates

Selection approach	Shtech Part A #Crops	MAE	RMSE
None (Pretrained)	NA	69.2	113.5
Random	1 of 16	68.6	113.5
Count	1 of 16	68.2	113.9
Aleatoric Uncertainty	1 of 16	65.7	103.5
Density based Ensemble Disagre.	1 of 16	66.4	112.4
KL-Ensemble Disagreement	1 of 16	66.4	109.7
Full dataset (previous best method)	full image	62.8	99.4
Full dataset (CTN)	full image	61.5	103.4

Table 1: **Comparing different strategies for selecting informative crops from an image for annotation.** We train the network on the UCF-QNRF dataset, and use it to select informative crops Shanghaitech Part A train data for acquiring annotation. The selected crops are used to adapt the network to the target domain. We compare the random selection and Count based sample selection baselines with the proposed uncertainty-guided selection strategies. For the experiment, each image is divided into 16 crops and a single crop from each image is chosen for annotation. We compute the informativeness score for all 16 crops, and pick out the crop with the highest score. For the random baseline, we pick out a crop at random from each image. For the count baseline, we pick out the crop with the highest count. Our proposed sample selection strategies outperforms the random selection and count based selection baselines.

in Fig. 1. The optimal uncertainty estimate, would be perfectly correlated with the error, and the characteristic curve for the best possible uncertainty estimate can be obtained by dropping pixels with the highest errors first. We refer to this curve as *Oracle* in Fig. 1, and the closer the uncertainty sparsification curve to the Oracle curve, the better. The area between the sparsification curve and the oracle curve for the Aleatoric and Epistemic uncertainty estimations are 0.07 and 0.25, respectively. This shows that the prediction error has higher correlation with the aleatoric uncertainty estimation than the epistemic uncertainty estimation. Similar evaluation strategies have been used for evaluating the uncertainty estimates pertaining to optical flow estimation [7].

4 Crop-level Sample Selection

In the main paper, we pointed out that our sampling strategy can be used for picking out images from a large pool of unlabeled images, and it can also be used to pick out informative crops from an image. Depending on the annotation budget, it might be useful to get partial annotations for an image by picking out informative crops from an image and getting human annotations for the informative crops rather than annotating the entire image. The experiments on image level sample selection are discussed in the main paper. In Tab. 1, we present experiments pertaining to selecting informative crops from an image for human annotation. For the experiment, each image is divided into 16 non-overlapping crops and a single crop from each image is chosen for annotation. We compute the informativeness score for all 16 crops, and pick out the crop with the highest score. The informativeness score of crops are computed using the three sample selection strategies proposed in the main paper. We compare our proposed approach with two baselines: random crop selection and count based crop selection. For the random baseline, we pick out a crop at random from each image. For the count baseline, we pick out the crop with the highest count. Our proposed sample selection strategies outperforms the random selection and count selection baselines.

5 Image-level Sample Selection on Shanghaitech Part B

In the main paper, we presented image level sample selection experiments on Shanghaitech Part A [2] and NWPU [3] datasets. In Tab. 2, we present image level sample selection experiments on Shanghaitech Part B.

	Shtech Part B		
Selection approach	#Train	MAE	RMSE
None (Pretrained)	NA	13.2	21.7
Random	50	9.4	16.6
Count	50	8.9	15.1
Aleatoric Uncertainty	50	8.7	14.3
Density based Ensemble Disagr.	50	8.5	13.6
KL-Ensemble Disagreement	50	8.8	15.4
Random	100	9.1	15.1
Count	100	8.0	13.5
Aleatoric Uncertainty	100	8.8	15.0
Density based Ensemble Disagreement	100	8.5	14.5
KL-Ensemble Disagreement	100	8.5	13.7
Full dataset (previous best method) [8]	400	7.6	11.8
Full dataset (CTN)	400	7.5	11.9

Table 2: **Comparing different strategies for selecting images for annotation on Shanghaitech Part B dataset** We train the network on the UCF-QNRF dataset, and use it to select images from the Shanghaitech Part B train data for acquiring annotation. We compare the random selection baseline with the proposed uncertainty-guided selection strategy. Our sample selection strategies outperform the two baselines when 50 images are sampled, and outperforms the random baseline when sampling 100 images.

Bibliography

[1] Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: NIPS. (2017)

[2] Zhang, Y., Zhou, D., Chen, S., Gao, S., Ma, Y.: Single-image crowd counting via multi-column convolutional neural network. In: Proc. CVPR. (2016)

[3] Wang, Q., Gao, J., Lin, W., Li, X.: Nwpu-crowd: A large-scale benchmark for crowd counting. arXiv preprint arXiv:2001.03360 (2020)

[4] Neal, R.M.: Bayesian learning for neural networks. Volume 118. Springer Science & Business Media (2012)

[5] Gal, Y., Ghahramani, Z.: Bayesian convolutional neural networks with bernoulli approximate variational inference. arXiv preprint arXiv:1506.02158 (2015)

[6] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15** (2014) 1929–1958

[7] Ilg, E., Cicek, O., Galesso, S., Klein, A., Makansi, O., Hutter, F., Brox, T.: Uncertainty estimates and multi-hypotheses networks for optical flow. In: Proc. ECCV. (2018)

[8] Shi, M., Yang, Z., Xu, C., Chen, Q.: Revisiting perspective information for efficient crowd counting. In: Proc. CVPR. (2019)