Uncertainty Estimation for Semantic Segmentation of Hyperspectral Imagery

Aneesh Rangnekar[0000−0002−0079−9495], Emmett Ientilucci, Christopher Kanan, and Matthew Hoffman

Rochester Institute of Technology, Rochester, NY, USA
aneesh.rangnekar@mail.rit.edu, emmett@cis.rit.edu, kanan@rit.edu, mjhsma@rit.edu

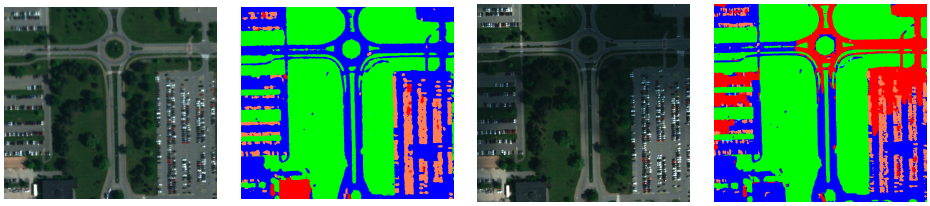
Abstract. As a step in a Dynamic Data-Driven Applications Systems (DDDAS) method to characterize the background in a vehicle tracking problem, we extend the application of deep learning to a hyperspectral dataset (the AeroRIT dataset) to evaluating network uncertainty. Expressing uncertainty information is crucial for evaluating what additional information is needed in the DDDAS algorithm and where more resources are required. Hyperspectral signatures tend to be very noisy, when captured from an aerial flight and a slight shift in the atmospheric conditions can alter the signals significantly, which in turn may affect the trained network's classifications. In this work, we apply Deep Ensembles, Monte Carlo Dropout and Batch Ensembles and study their effects with respect to achieving robust pixel-level identifications by expressing the uncertainty within the trained networks on the task of semantic segmentation. We modify the U-Net-m architecture from the AeroRIT paper to account for the frameworks and present our results as a step towards accounting for sensitive changes in hyperspectral signals.

Keywords: hyperspectral · uncertainty · segmentation

1 Introduction

Instead of solely modeling vehicle movement or focusing on vehicle appearance for a vehicle tracking problem, we are working on adaptively modeling the background in a DDDAS [4] framework using hyperspectral data. This allows the possibility of identifying potential confusers and modifying the detection or tracking strategy. In this paper we describe efforts to characterize the background and the uncertainties in a classification problem. A fair amount of effort has been invested in applying deep learning methodologies to hyperspectral imagery for the purpose of learning scene representations towards aerial object detection and tracking [13, 17, 18]. In this paper, we use the AeroRIT data set released in with SegNet and U-Net networks [1, 13, 16] trained for the task of semantic segmentation. However, AeroRIT also comes with the limitation of being a single flight line captured under clear atmospheric conditions. If the same set of trained networks are used to run inference on a similar dataset but under

different atmospheric conditions, the outputs will, more than likely, vary. We visually verify this claim by applying one of the networks established in the paper (U-Net-m, discussed further in Sec. 3.2) to another flight line captured under cloudy atmospheric conditions in Fig. 1. We would ideally pre-process the images to ensure no atmospheric occlusions are present in the scene - for example, a cloud shadow removal algorithm, however we use this snapshot as a particular example to illustrate our goal in this paper. We observe that the network fails to recognize the correct set of classes in key areas of interest - for example, the region around the circular roundabout is predicted to be a building instead of a road. This can affect the flow of down-streaming tasks dependent on decision trees - do we want to look for vehicles at pixels classified as buildings? While the straight forward answer is a No, the approach can be altered if we could also be privy to information about the network's confidence (viz-a-viz, uncertainty) of the pixel's classification. This information may help in creating more robust inferences as other networks in down-streaming tasks would be aware of the prediction's uncertainty and can dynamically adapt to account for variations.



(a) (b) (c) (d)

Fig. 1: Roundabout section from AeroRIT under sunny and cloudy atmospheric conditions. We observe the output of a network trained on the clear flight line to its cloud-occluded counterpart, (c)-(d), (e)-(f) respectively. The labels are roads (blue), cars (ivory), buildings (red) and vegetation (green).

We train deep networks by minimizing the difference between the networks' prediction and the true distribution of labels and during evaluation, use the learnt set of weights for classification by selecting the class label corresponding to the maximum probability. However, this approach does not provide any information about the network's uncertainty of the predictions. Kendall *et al.* applied Bayesian deep learning to obtain the network's uncertainty for depth regression and semantic segmentation tasks [9, 10]. We adopt their approach in this paper and analyze the effect of uncertainty quantification towards AeroRIT scene understanding.

2 Related Works

There are many areas of research that can be used to estimate the network's uncertainty, the most popular being: 1) forming ensembles [1, 7, 8, 19], 2) varia-

tional inference [2], and 3) K-FAC Laplace approximation [15]. We focus on the first type of approach - forming ensembles as it is relatively simpler to follow and easier to implement compared to the other areas. The core idea is to train a bunch of networks with different initializations on the same set of data and at test time, evaluate the final predictions as an average of the ensemble networks predictions. Gal and Ghahramani showed that using dropout across layers of the convolutional neural network (CNN) can act as approximate Bayesian interpretation [7]. This facilitates training a single network and using dropout at test time to create model ensembles. Kendall *et al.* further demonstrated that applying dropout at selective layers of the network instead of all layer further improves the predictions [9]. Lakshminarayanan *et al.* trained different networks separately for forming ensembles [11], and Huang *et al.* obtained sets of networks by taking *snapshots* at different intervals using cyclic learning rate schedule [12]. Recently, Wen *et al.* proposed to use multiple rank-1 matrices along with the core weight matrix to form ensembles as an alternative to existing methods [19]. Uncertainty estimation approaches [3, 6, 14] have already been applied in other areas of remote sensing. In this paper, we adopt deep ensembles [11], Monte-Carlo dropout based ensembles [1, 7] and batch ensembles [19] for estimating network uncertainty.

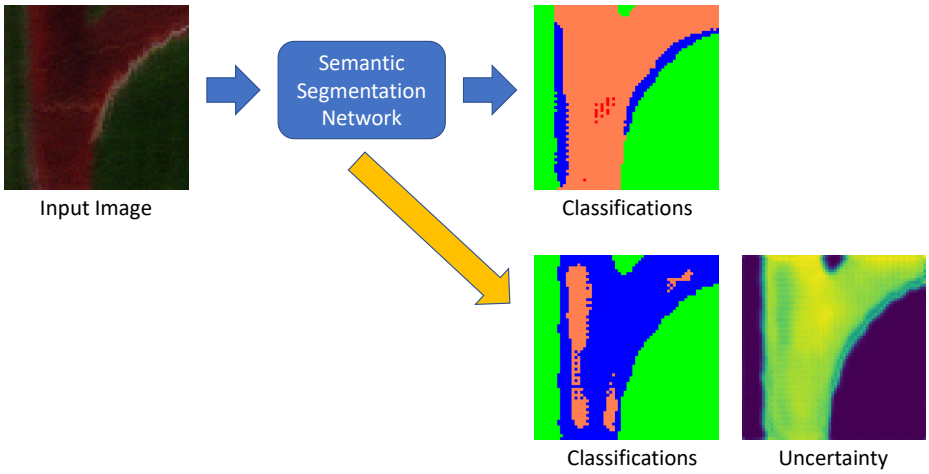


Fig. 2: Schematic overview of the uncertainty based pipeline. The standard flow is shown with blue arrows where the trained network predicts the pixel-wise labels. We augment the flow with a ensemble learning framework (orange) that eventually accounts for the uncertainty within the network. Brighter areas correlate to larger uncertainty - and as image chips corresponding to the racetrack are not present in the training set, the network is overall highly uncertain of its prediction.

3 Estimating Uncertainty

3.1 Types of Uncertainties

Kendall and Gal expressed uncertainty into two subtypes - Aleatoric and Epistemic, in accordance with Kiureghian and Ditlevsen [5]. Aleatoric uncertainty corresponds to noise that is data-independent, for example, sensor noise, environmental noise, and cannot be reduced even if more data is collected. Epistemic uncertainty can be expressed as more data-dependent and model-based, and hence is widely modelled using ensembles. In our paper, we focus on epistemic uncertainty and use ensembles for estimation. Fig. 2 outlays the overall framework.

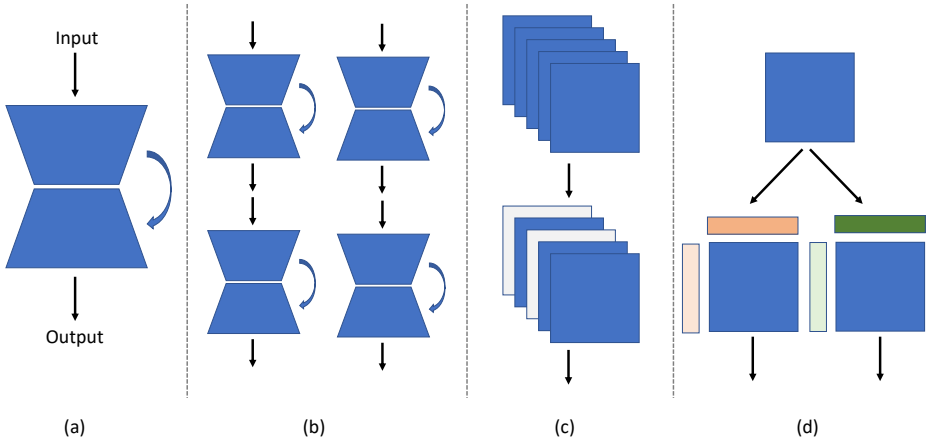


Fig. 3: All settings used in the paper: (a) U-Net-m, (b) 4 deep ensembles for [11], (c) MC-Dropout applied on the convolutional maps of (a) with *Spatial Dropout*, (d) Batch Ensembles with two sets of rank-1 matrices on weights of (a).

3.2 Network review

We use the U-Net-m architecture developed in AeroRIT [13] for its better performance among other networks. It contains 2 downsampling convolutional blocks, followed by a bottleneck layer and 2 upsampling blocks with skip connections. Each convolutional block contains two sets of convolutional kernels of 3×3, a Batch-Normalization layer and ReLU activation. We represent this structure in Fig. 3 (a).

3.3 Deep Ensembles (DE)

This approach, proposed by Lakshminarayanan *et al.* [11], averages the predictions across networks trained independently starting from different initializations

(Fig. 3 (b)). Every member of the deep ensembles is trained with the same hyperparameters as discussed in Sec. 4.1. At test time, we average the predictions to obtain the final set of predictions. Following all approaches that estimate uncertainty, we use entropy of the resulting distribution as the measure of uncertainty and use it in all the figures throughout this paper.

3.4 MC-Dropout (MCD)

Monte-Carlo Dropout is the less training-time alternative to Deep Ensembles. Instead of training separate copies of networks multiple times, Gal *et al.* [7, 9] proposed to inject Bernoulli noise in form of Dropout over the activations of the network weights. In practice, we observed that applying spatial dropout instead of conventional dropout produced more better uncertainty estimates (Figs. 3 (c), X). Spatial dropout randomly drops an entire feature map from the list of feature maps as compared to individual elements in conventional dropout. We use the same set of hyperparameters as discussed in Sec. 4.1. At test time, we average the predictions obtained across a fixed set of runs with dropout enabled to obtain the final set of predictions.

3.5 Batch Ensembles (BE)

This approach was proposed by Wen *et al.* and works as an alternative to using Dropout for ensembles (Fig. 3 (d)). The core idea is to have a single *slow* matrix (W), which corresponds to the 2-D convolution kernel weight and two corresponding rank-1 matrices (r_i, s_i) that act as *fast* matrices:

$$\overline{W_i} = W \circ F_i, \text{ where } F_i = r_i s_i^\top, \tag{1}$$

and hence, we obtain $\overline{W_i}$ as the corresponding weight for ensemble i. The number of ensembles is equal to the number of sets of rank-1 matrices used and is very efficient in terms of model storage. During evaluation, similar to above, we repeat the mini-batch to correspond with total number of ensemble members and average the predictions.

4 Experiments and Results

4.1 Hyperparameters

We use all 51 bands available in the AeroRIT dataset chips in this paper - 31 visible and 20 infrared bands. All chips are clipped to a maximum of 2^{14}, and normalized between 0 and 1, before forward passing through the networks. All networks are initialized with Kaiming init, and the rank-1 matrices for BE are initialized to have a mean of 1 and standard deviation of 0.5 in accordance with the original paper. We use an initial learning rate of $1e^{-2}$: for DE and MCD, we train for 60 epochs with drops of 0.1 at 30, 40, 50^{th} and for BS, we train for 120 epochs with drops of 0.1 at 50, 80, and 100^{th} epoch. We train with standard cross-entropy loss (CE) for DE and MCD and use weighted CE only for the BE approach.

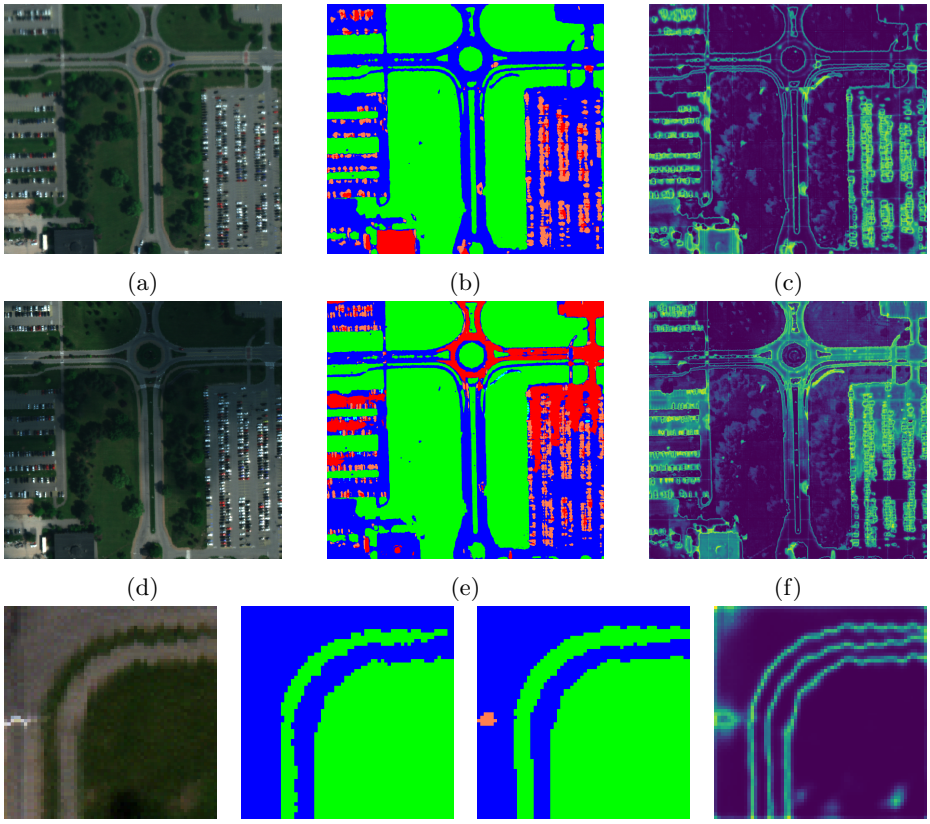


Fig. 4: Visualization of uncertainty estimates on the sunny and cloudy roundabout scenes from Fig. 1. (a) and (d) are the RGB rendered images, (b) and (e) are the corresponding network predictions while, (c) and (f) are the uncertainty maps. We also visualize an instance from (g) the AeroRIT test set with (h) corresponding ground truth label, (i) network predictions and (j) uncertainty map.

4.2 Results

We observe visual improvement over the scenes presented in Fig. 1 using 10 runs with MCD ensembles. The network predictions for the roundabout area show high uncertainty (Fig. 4 (f)) which is desired in this setting. This information can be used by down-steam tasks which can dynamically adapt to ensure continuity. Further, we also observe that the network uncertainty estimates are high for row Fig. 4 (g) as the road crossing has been incorrectly classified as belonging to the vehicle category. We also observe uncertainty around the boundaries of classes - this can possibly be due to the presence of mixed pixels. Table 1 shows us

Table 1: Results of techniques discussed in Sec. 3 compared to the baseline network from AeroRIT [13].

	Standard Network	Deep Ensembles	Monte Carlo Dropout	Batch Ensembles
mIOU	70.62	71.41 ± 2.48	72.45 ± 1.56	69.05 ± 3.45

that all ensemble techniques are able to achieve near-par or higher performance than the conventional counterpart. We use mean IOU (mIOU) as the metric of interest (following [13]) and do not discuss metrics pertaining to uncertainty estimations (for example, Expected Calibration Error) for the scope of this paper. mIOU is the class-wise mean of the area of intersection between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. To generate the results, we ran all ensembles 10 different times, with varying number of models for DE and MCD. We found 4 to be a sufficient set of models for DE and BE and 10 for MCD in our ablation studies.

5 Conclusion

We presented the extension of uncertainty estimation to hyperspectral remote sensing imagery as a first step towards dynamic scene adaptation under varying atmospheric conditions. Our next set of questions are as follows: 1) can we reduce uncertainty in mixed pixel areas to obtain a much precise map that can be passed to down-stream tasks ? and 2) can we decrease the inference speed to get as close to a single forward pass of a network ? 3) is it possible to design an end to end framework to adaptively shift between sensor modalities using uncertainty as an input?

Acknowledgements

This work was supported by the Dynamic Data Driven Applications Systems Program, Air Force Office of Scientific Research under Grant FA9550-19-1-0021. We gratefully acknowledge the support of NVIDIA Corporation with the donations of the Titan X and Titan Xp Pascal GPUs used for this research.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561 (2015)
2. Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D.: Weight uncertainty in neural networks. arXiv preprint arXiv:1505.05424 (2015)

3. Cobb, A.D., Himes, M.D., Soboczenski, F., Zorzan, S., O'Beirne, M.D., Baydin, A.G., Gal, Y., Domagal-Goldman, S.D., Arney, G.N., Angerhausen, D., et al.: An ensemble of bayesian neural networks for exoplanetary atmospheric retrieval. The Astronomical Journal **158**(1), 33 (2019)
4. Darema, F.: Dynamic data driven applications systems: A new paradigm for application simulations and measurements. In: Bubak, M., van Albada, G.D., Sloot, P.M.A., Dongarra, J. (eds.) Computational Science - ICCS 2004. pp. 662–669. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)
5. Der Kiureghian, A., Ditlevsen, O.: Aleatory or epistemic? does it matter? Structural safety **31**(2), 105–112 (2009)
6. Fletcher, S., Lickley, M., Strzepek, K.: Learning about climate change uncertainty enables flexible water infrastructure planning. Nature communications **10**(1), 1–11 (2019)
7. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059 (2016)
8. Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J.E., Weinberger, K.Q.: Snapshot ensembles: Train 1, get m for free. arXiv preprint arXiv:1704.00109 (2017)
9. Kendall, A., Badrinarayanan, V., Cipolla, R.: Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. arXiv preprint arXiv:1511.02680 (2015)
10. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: Advances in neural information processing systems. pp. 5574–5584 (2017)
11. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In: Advances in neural information processing systems. pp. 6402–6413 (2017)
12. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
13. Rangnekar, A., Mokashi, N., Ientilucci, E.J., Kanan, C., Hoffman, M.J.: Aerorit: A new scene for hyperspectral image analysis. IEEE Transactions on Geoscience and Remote Sensing (2020)
14. Rao, V., Sandu, A.: A posteriori error estimates for dddas inference problems. Procedia Computer Science **29**, 1256–1265 (2014)
15. Ritter, H., Botev, A., Barber, D.: A scalable laplace approximation for neural networks. In: 6th International Conference on Learning Representations, ICLR 2018-Conference Track Proceedings. vol. 6. International Conference on Representation Learning (2018)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
17. Uzkent, B., Rangnekar, A., Hoffman, M.J.: Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 233–242. IEEE (2017)
18. Uzkent, B., Rangnekar, A., Hoffman, M.J.: Tracking in aerial hyperspectral videos using deep kernelized correlation filters. IEEE Transactions on Geoscience and Remote Sensing **57**(1), 449–461 (2018)
19. Wen, Y., Tran, D., Ba, J.: Batchensemble: An alternative approach to efficient ensemble and lifelong learning (2020)