Uncertainty-Guided Edge Learning for Deep Image Regression in Remote Sensing

Supplementary Material

8. Derivations

We have the likelihood of a beta distribution given the mean μ and the precision ν:

$$p(y|\mu,\nu) = \frac{\Gamma(\nu)}{\Gamma(\mu\nu)\Gamma((1-\mu)\nu)} y^{\mu\nu-1}(1-y)^{(1-\mu)\nu-1}, \tag{12}$$

where $\Gamma(\cdot)$ is the gamma function. We can compute the negative log likelihood loss, $\mathcal{L}_{NLL}$, as:

$$\begin{aligned}\mathcal{L}_{NLL}(\mu,\nu,y) &= -\log p(y|\mu,\nu) \\ &= -[\log \frac{\Gamma(\nu)}{\Gamma(\mu\nu)\Gamma((1-\mu)\nu)} + \log y^{\mu\nu-1} \\ &\quad + \log (1-y)^{(1-\mu)\nu-1}] \\ &= \log \frac{\Gamma(\mu\nu)\Gamma((1-\mu)\nu)}{\Gamma(\nu)} - (\mu\nu-1)\log y \\ &\quad + ((1-\mu)\nu-1)\log(1-y)\end{aligned} \tag{13}$$

The differential entropy of the beta distribution, given shape parameters $\alpha, \beta > 0$, is commly expressed as:

$$\begin{aligned}h = &\log Beta(\alpha,\beta) - (\alpha-1)[\psi(\alpha) - \psi(\alpha+\beta)] \\ &- (\beta-1)[\psi(\beta) - \psi(\alpha+\beta)]\end{aligned} \tag{14}$$

where $\psi(\cdot)$ is the digamma function. In this work, we use an alternative parameterisation with the mean $\mu = \alpha/(\alpha+\beta) \in (0,1)$ and the precision $\nu = \alpha + \beta > 0$. Therefore, the parameters α and β can be represented in terms of the mean μ and precision ν:

$$\begin{aligned}\alpha &= \mu\nu, \\ \beta &= (1-\mu)\nu\end{aligned} \tag{15}$$

Thereby, the differential entropy of the beta distribution is calculated as:

$$\begin{aligned}h = &\log Beta(\mu\nu,(1-\mu)\nu) \\ &- (\mu\nu-1)[\psi(\mu\nu) - \psi(\mu\nu+(1-\mu)\nu))] \\ &- ((1-\mu)\nu-1)[\psi((1-\mu)\nu) - \psi(\mu\nu+(1-\mu)\nu)] \\ = &\log \frac{\Gamma(\mu\nu)\Gamma((1-\mu)\nu)}{\Gamma(\nu)} \\ &+ (\nu-2)\psi(\nu) - (\mu\nu-1)\psi(\mu\nu) \\ &- ((1-\mu)\nu-1)\psi((1-\mu)\nu) \\ = &\log\Gamma(\mu\nu) + \log(\Gamma((1-\mu)\nu)) - \log\Gamma(\nu) \\ &- (\mu\nu-1)[\psi(\mu\nu) - \psi(\nu)] \\ &- ((1-\mu)\nu-1)[\psi((1-\mu)\nu) - \psi(\nu)] \\ = &\log\Gamma(\mu\nu) + \log(\Gamma((1-\mu)\nu)) - \log\Gamma(\nu) \\ &+ (\nu-2)\psi(\nu) - (\mu\nu-1)\psi(\mu\nu) \\ &- ((1-\mu)\nu-1)\psi((1-\mu)\nu)\end{aligned} \tag{16}$$

9. Benchmarking on RS applications

Fig. 8, Fig. 9, and Fig. 10 present the benchmarking results of UGEL against representative methods on `RSRC-S2` using the ResNet18, MobileNetV3, and MobileNetV4 backbones, respectively. These results are consistent with those observed on `RSRC-L8` and `RSLC`, as shown in Fig. 6 in the main paper. Additionally, Fig. 11 and Fig. 12 present the benchmarking results of UGEL against representative methods on `RSRC-L8` and `RSLC` using the MobileNetV4 backbone.

Tab. 4 and Tab. 7 provide the Wilcoxon signed-rank test p-values comparing UGEL with DBR to other methods using the MobileNetV3 and MobileNetV4 backbones, respectively.

10. UGEL ablation studies

10.1. [T1] Effects of different uncertainty estimation methods on UGEL.

The experimental setup, hyperparameter configurations, and evaluation metric were kept consistent with those described in the main paper. We evaluated the impact of different uncertainty estimation methods within the UGEL framework. In addition to Fig. 7 in the main paper, Fig. 13 provides an empirical comparison of their strengths and limitations in cloud and land coverage prediction tasks.

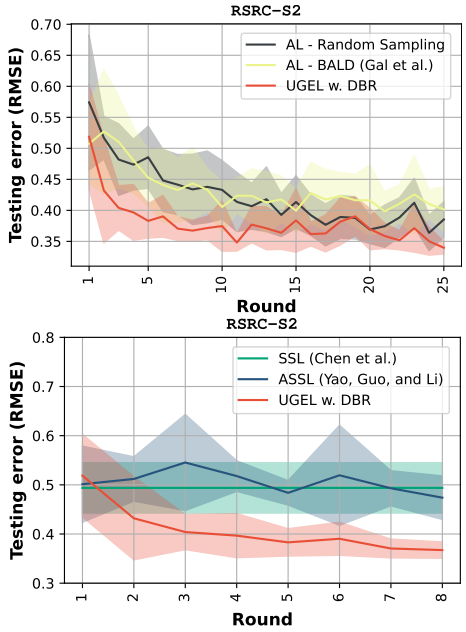

Figure 8. Benchmarking UGEL with DBR against AL (BALD and random sampling), SSL and ASSL on RSRC−S2 using the ResNet18 backbone. Bolded curves and shaded regions respectively indicate mean RMSE testing error and std dev over 10 runs.

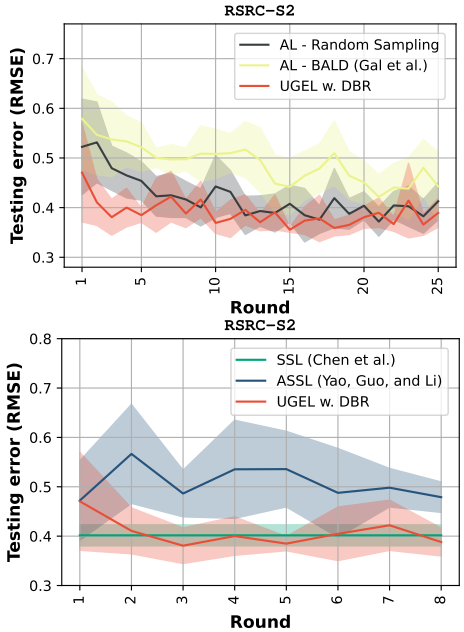

Figure 10. Benchmarking UGEL with DBR against AL (BALD and random sampling), SSL and ASSL on RSRC−S2 using the MobileNetV4 backbone. Bolded curves and shaded regions respectively indicate mean RMSE testing error and std dev over 10 runs.

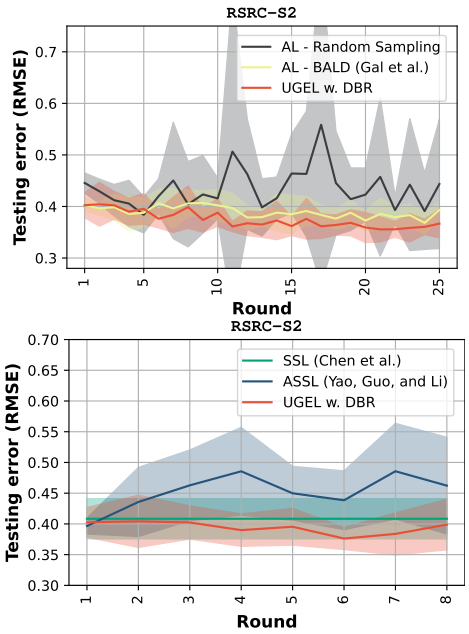

Figure 9. Benchmarking UGEL with DBR against AL (BALD and random sampling), SSL and ASSL on RSRC−S2 using the MobileNetV3 backbone. Bolded curves and shaded regions respectively indicate mean RMSE testing error and std dev over 10 runs.

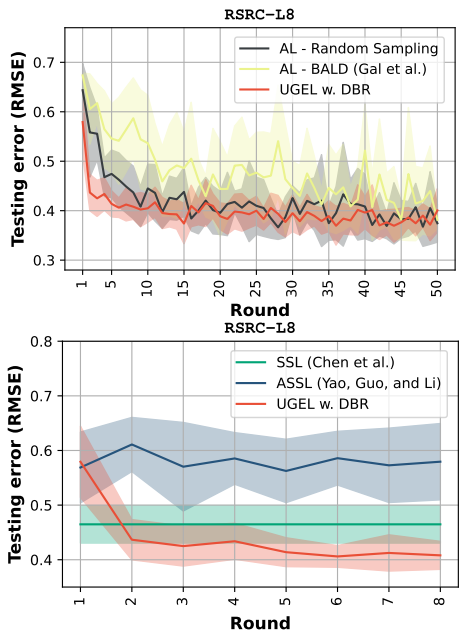

Figure 11. Benchmarking UGEL with DBR against AL (BALD and random sampling), SSL and ASSL on RSRC−L8 using the MobileNetV4 backbone. Bolded curves and shaded regions respectively indicate mean RMSE testing error and std dev over 10 runs.

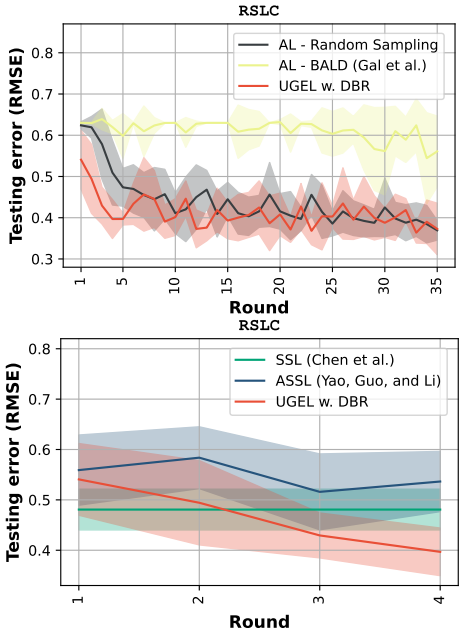

Figure 12. Benchmarking UGEL with DBR against AL (BALD and random sampling), SSL and ASSL on RSLC using the MobileNetV4 backbone. Bolded curves and shaded regions respectively indicate mean RMSE testing error and std dev over 10 runs.

The results demonstrate that DBR is significantly more effective than other uncertainty methods in UGEL, particularly in the early rounds where the performance gap was most evident. DBR's uncertainty estimations facilitated the selection of informative unlabeled samples for model updating. On RSRC−L8 and RSRC−S2 datasets, DBR achieved not only lowest RMSE but also the smallest standard deviation, indicating greater reliability for real-world cloud coverage prediction. For the land coverage prediction using RSLC, while DBR still outperformed in early rounds, DER and RAN quickly reached comparable performance. This outcome may be attributed to the greater diversity of objects considered as coverage in the land coverage prediction task.

10.2. [T2] Effects of different groundtruth regression label distributions on the performance of UGEL.

Real scenes with synthetic clouds (RSSC). We sourced images from NASA's Landsat-9 Level-1 Collection. Ten scenes from different locations and diverse terrains (*e.g.*, sea, mountain, urban, rural, dessert) were selected. We ensured that the collected scenes had no visible clouds. The 12-band raw data of each scene was converted to top-of-atmosphere spectral radiance, corrected for the sun angle using the metadata, and bands 4-3-2 were then selected to generate an RGB base image of dimension of 8000x8000

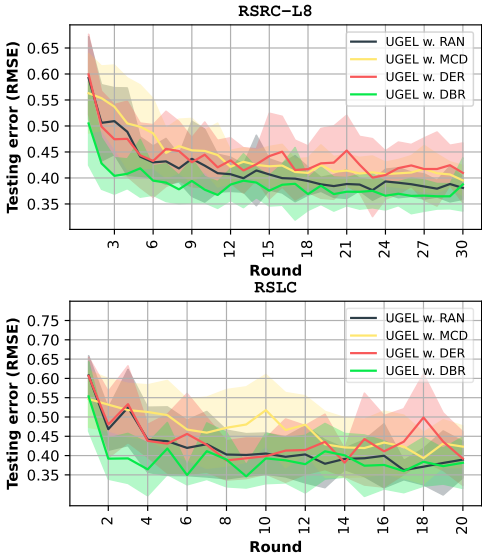

Figure 13. (Top) Comparison of different uncertainty estimation methods in UGEL on RSRC−L8 using the ResNet18 backbone. (Bottom) Same comparison on RSLC using the ResNet18 backbone. Bolded curves and shaded regions respectively indicate mean RMSE testing error and std dev over 10 runs.

	Test	Rd. 2	Rd. 4	Rd. 6	Rd. 8
RSRC−S2	UGEL vs Rand	0.080	0.348	0.014	0.688
	UGEL vs BALD	0.688	0.722	0.032	0.312
	UGEL vs SSL	0.461	0.065	0.042	0.348
	UGEL vs ASSL	0.116	0.002	0.007	0.019
	Test	**Rd. 2**	**Rd. 4**	**Rd. 6**	**Rd. 8**
RSRC−L8	UGEL vs Rand	0.007	0.001	0.065	0.053
	UGEL vs BALD	0.01	0.001	0.246	0.024
	UGEL vs SSL	0.053	0.005	0.005	0.001
	UGEL vs ASSL	0.01	0.002	0.001	0.001
	Test	**Rd. 1**	**Rd. 2**	**Rd. 3**	**Rd. 4**
RSLC	UGEL vs Rand	0.001	0.001	0.001	0.001
	UGEL vs BALD	0.754	0.577	0.161	0.014
	UGEL vs SSL	0.161	0.754	0.423	0.313
	UGEL vs ASSL	0.406	0.156	0.062	0.031

Table 4. Wilcoxon signed-rank test p-values comparing UGEL with DBR against other methods on all datasets using the MobileNetV3 backbone.

pixels. Synthetic clouds with groundtruth cloud masks were generated and added to the base images using SatelliteCloudGenerator [14]; see Fig. 14. Then, we cropped the base images and their ground-truth masks into 128x128 non-overlapping patches (x). In total, we extracted $9,600$ patches from 5 base images for EL and $8,000$ patches from the remaining 5 base images as testing data from unseen before scenes. Using synthetic clouds allowed us to con-

trol the groundtruth cloud coverage percentage (y) in each image and its distribution over the dataset. RSSC was modified into four variations, each corresponding to a different ground truth label distribution.

- RSSC−B: Bimodal distribution,
- RSSC−N: Negatively skewed distribution,
- RSSC−U: Uniform distribution,
- RSSC−G: Gaussian distribution.

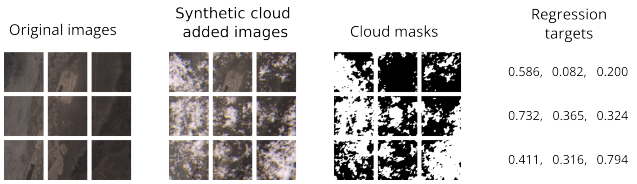

Figure 14. Sample images from RSSC with regression targets.

Fig. 15 shows the performance of different uncertainty estimation methods in UGEL under different ground truth label distributions. Tab. 8 provides Wilcoxon signed-rank test p-values comparing UGEL with DBR to other methods on RSSC using the ResNet18 backbone.

10.3. Runtime comparisons

Method	Uncertain. est.	Model retrain.
UGEL w. RAN	n/a	8.26-12.54
UGEL w. MCD	122.19-127.94	5.93-12.35
UGEL w. DER	18.39-19.04	6.06-10.61
UGEL w. DBR	18.01-18.63	5.96-12.16

Table 5. Runtime (minimum and maximum, in seconds) for different uncertainty estimation methods during the first 10 rounds of UGEL on the RSRC−L8 dataset using the ResNet18 backbone.

Method	Uncertain. est.	Model retrain.
UGEL w. RAN	n/a	7.73-9.61
UGEL w. MCD	1021.00-1095.18	7.83-10.24
UGEL w. DER	176.55-507.84	7.84-9.65
UGEL w. DBR	171.94-507.43	7.93-9.32

Table 6. Runtime (minimum and maximum, in seconds) for different uncertainty estimation methods during the first 5 rounds of UGEL on the RSLC dataset using the ResNet18 backbone.

	Test	Rd. 2	Rd. 4	Rd. 6	Rd. 8
RSRC−S2	UGEL vs Rand	0.031	0.031	0.422	0.156
	UGEL vs BALD	0.016	0.016	0.016	0.016
	UGEL vs SSL	0.891	0.500	0.578	0.422
	UGEL vs ASSL	0.016	0.047	0.078	0.016
	Test	**Rd. 2**	**Rd. 4**	**Rd. 6**	**Rd. 8**
RSRC−L8	UGEL vs Rand	0.007	0.080	0.014	0.065
	UGEL vs BALD	0.001	0.002	0.003	0.001
	UGEL vs SSL	0.097	0.053	0.001	0.002
	UGEL vs ASSL	0.001	0.001	0.001	0.001
	Test	**Rd. 1**	**Rd. 2**	**Rd. 3**	**Rd. 4**
RSLC	UGEL vs Rand	0.031	0.078	0.031	0.031
	UGEL vs BALD	0.016	0.031	0.016	0.016
	UGEL vs SSL	0.953	0.781	0.156	0.047
	UGEL vs ASSL	0.219	0.219	0.047	0.047

Table 7. Wilcoxon signed-rank test p-values comparing UGEL with DBR against other methods on all datasets using the MobileNetV4 backbone.

	Test	Rd. 5	Rd. 10	Rd. 15	Rd. 20
RSSC−B	DBR vs RAN	0.032	0.002	0.007	0.019
	DBR vs MCD	0.002	0.001	0.001	0.001
	DBR vs DER	0.032	0.014	0.116	0.002
	Test	**Rd. 5**	**Rd. 10**	**Rd. 15**	**Rd. 20**
RSSC−N	DBR vs RAN	0.01	0.188	0.461	0.539
	DBR vs MCD	0.001	0.001	0.005	0.024
	DBR vs DER	0.007	0.001	0.348	0.116
	Test	**Rd. 5**	**Rd. 10**	**Rd. 15**	**Rd. 20**
RSSC−U	DBR vs RAN	0.08	0.348	0.116	0.423
	DBR vs MCD	0.001	0.001	0.001	0.003
	DBR vs DER	0.005	0.019	0.014	0.024
	Test	**Rd. 5**	**Rd. 10**	**Rd. 15**	**Rd. 20**
RSSC−G	DBR vs RAN	0.539	0.348	0.278	0.935
	DBR vs MCD	0.001	0.001	0.001	0.003
	DBR vs DER	0.005	0.019	0.014	0.024

Table 8. Wilcoxon signed-rank test p-values comparing UGEL with DBR to other methods on RSSC using the ResNet18 backbone.

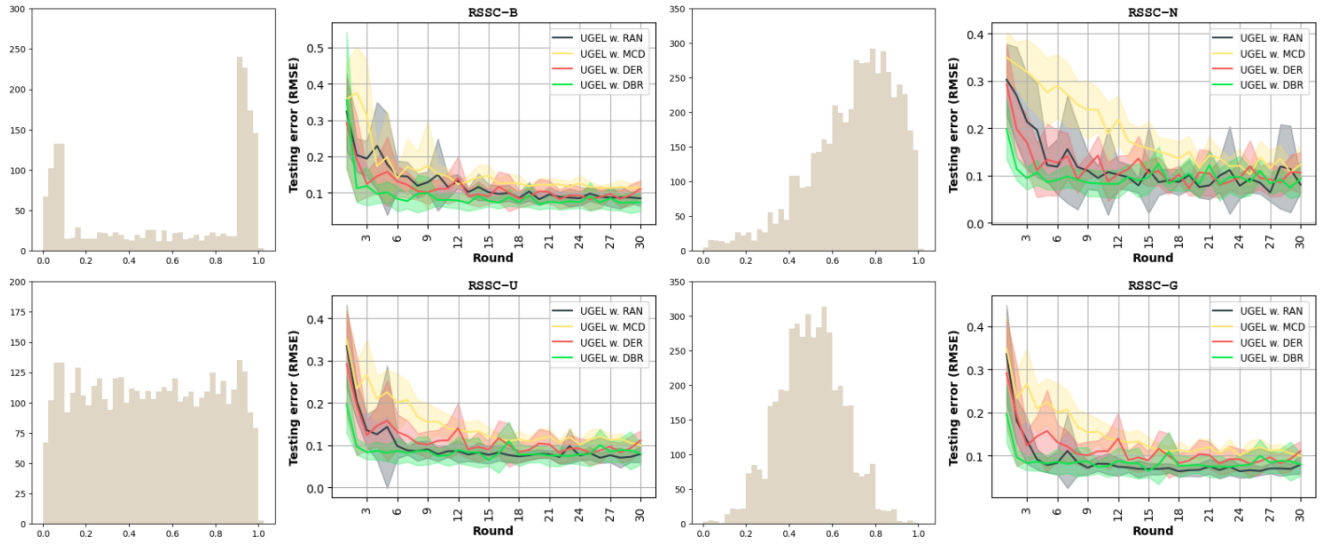

Figure 15. Comparison of different uncertainty estimation methods in UGEL on synthetic datasets: (Top Left) `RSSC−B`, (Top Right) `RSSC−N`, (Bottom Left) `RSSC−U`, (Bottom Right) `RSSC−G`. Bolded curves and shaded regions respectively indicate mean RMSE testing error and std dev over 10 runs.