Aerial Road Segmentation in the Presence of Topological Label Noise – Supplementary Materials

Corentin Henry
Remote Sensing Technology Institute
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
Email: corentin.henry@dlr.de

Friedrich Fraundorfer
Institute of Computer Graphics and Vision
Graz University of Technology (TUG)
Graz, Austria
Email: fraundorfer@icg.tugraz.at

Eleonora Vig
Amazon
Berlin, Germany
Email: eleonov@amazon.com

I. Fine-tuning the β parameter in losses

We search for the optimal β parameters for our Soft-Bootstrapped Cross-Entropy (BCE SB) and DICE (DICE SB) losses by cross-validating its value on the custom validation sets of the Massachusetts Roads (MA) dataset and the DeepGlobe Roads (DG) dataset. On MA, we find the following best values: $\beta = 0.8$ for BCE SB and $\beta = 0.6$ for DICE SB. On DG, we find: $\beta = 0.7$ for BCE SB and $\beta = 0.4$ for DICE SB. Figure 1 shows the performance fluctuation with varying β. We report our results for the best β values in Tables III and IV.

II. Metrics

A large variety of metrics is in use to evaluate the performance of road segmentation methods. In our study, we consider only pixel-wise metrics, setting aside graph-based metrics as they often evaluate the roads' global connectivity rather than their local topology. Most papers report the IoU, DICE/F1-Score, Precision, and Recall. Some works also report the so-called "relaxed" version of these metrics [1]: within a distance of typically 3 pixels from the labels, any road prediction is counted as an additional True Positive pixel. This idea aims at compensating the lack of accuracy in the ground truth, so we do not report relaxed metrics as we want to study the impact of such noisy labels on the model performance. In the case of the Massachusetts dataset (MA), we rule out the use of the Precision-Recall Breakeven Point as it is inconsistently reported, sometimes relaxed and sometimes not. In 1998, Wiedemann [2] introduced specific metrics for road extraction, called Completeness, Correctness, and Quality (please refer to the original paper for detailed explanations). In the following formulas, $\mathrm{L}()$ is a function calculating the length in pixels of the given road segment:

$$Completeness = \frac{\mathrm{L}(MatchedReference)}{\mathrm{L}(Reference)} \quad (1)$$

$$Correctness = \frac{\mathrm{L}(MatchedExtraction)}{\mathrm{L}(Extraction)} \quad (2)$$

$$Quality = \frac{\mathrm{L}(MatchedExtraction)}{\mathrm{L}(Extraction) + \mathrm{L}(UnmatchedExtraction)} \quad (3)$$

They are especially convenient when evaluating the topology of road centerlines, more so than the traditional pixel-wise metrics, and were already reported in [3] on MA. Although initially designed for vector road centerlines, they are applicable to single-pixel road centerlines, also known as "skeletons": measuring the length of a vector centerline from its coordinates is indeed equivalent to counting the pixels of the rasterized centerline.

III. Detailed results tables

Here, we report a more complete list of metrics and FCNN models for our different experiments. These tables are more exhaustive than most previous works on our four datasets, and hopefully can serve as a comparison basis for later studies. The Massachusetts Roads dataset is referred to below as MA, the Deep Globe Road Extraction Challenge as DG, the CrackTree dataset as CT, and the Electron Microscopy dataset as EM.

References

[1] V. Mnih, "Machine Learning for Aerial Image Labeling," Ph.D. dissertation, University of Toronto, 2013.

[2] C. Wiedemann, C. Heipke, H. Mayer, and O. Jamet, "Empirical Evaluation Of Automatically Extracted Road Axes," in *Empirical Evaluation Techniques in Computer Vision*, 1998, pp. 172–187.

[3] A. Mosinska, P. Marquez-Neila, M. Kozinski, and P. Fua, "Beyond the Pixel-Wise Loss for Topology-Aware Delineation," in *CVPR*, 2018.

[4] Y. Wei, Z. Wang, and M. Xu, "Road structure refined cnn for road extraction in aerial image," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 709–713, 2017.

[5] A. Constantin, J. Ding, and Y. Lee, "Accurate road detection from satellite images using modified u-net," in *IEEE Asia Pacific Conference on Circuits and Systems*, 2018, pp. 423–426.

[6] Z. Zhang and Y. Wang, "Jointnet: A common neural network for road and building extraction," *Remote Sens.*, vol. 11, no. 6, 2019.

[7] M. Yuan, Z. Liu, and F. Wang, "Using the wide-range attention u-net for road segmentation," *Remote Sens. Lett.*, vol. 10, no. 5, pp. 506–515, 2019.

[8] X. Gao *et al.*, "An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network," *IEEE Access*, vol. 6, pp. 39 401–39 414, 2018.

[9] L. Gao, W. Song, J. Dai, and Y. Chen, "Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network," *Remote Sens.*, vol. 11, no. 5, 2019.

[10] T. Sun, Z. Chen, W. Yang, and Y. Wang, "Stacked u-nets with multi-output for road extraction," in *CVPR Workshops*, 2018.

[11] D. Costea, A. Marcu, E. Slusanschi, and M. Leordeanu, "Roadmap generation using a multi-stage ensemble of deep neural networks with smoothing-based optimization," in *CVPR Workshops*, 2018.

TABLE I
BASELINE RESULTS FOR VARIOUS FCNN ARCHITECTURES ON THE CUSTOM TEST SPLIT OF MA (* NOT PRE-TRAINED)

Model	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
DeepLabv3+*	52.95	69.35	78.13	62.34	86.69	74.07	66.74
DenseASPP	46.63	64.44	73.75	57.22	85.62	69.89	62.32
D-LinkNet34	53.93	70.34	82.34	61.40	91.37	72.23	67.65
D-LinkNet50	54.90	71.01	81.15	63.12	90.01	73.75	68.25
D-LinkNet101	54.52	70.86	82.29	62.22	91.14	72.72	67.97
U-Net*	55.92	71.91	80.59	64.92	89.26	75.33	69.25
Res-U-Net18	55.94	72.02	80.74	65.00	88.47	75.51	68.75
Res-U-Net34	55.85	71.88	**82.78**	63.51	**91.43**	73.85	69.25
Res-U-Net50	56.93	72.74	81.87	65.44	90.60	75.11	69.89
Res-U-Net101	56.23	71.88	81.48	64.31	90.08	74.80	69.62
Dense-U-Net-121	**57.12**	**73.03**	81.64	66.07	90.50	75.59	70.06
Dense-U-Net-169	54.82	70.84	81.49	62.65	89.70	72.46	67.42
Dense-U-Net-201	57.04	72.95	81.36	**66.11**	90.24	75.94	**70.13**
Dense-U-Net-121*	55.49	71.44	81.45	63.62	90.59	73.79	68.69

TABLE II
BASELINE RESULTS FOR VARIOUS FCNN ARCHITECTURES ON THE CUSTOM VALIDATION SPLIT OF DG (* NOT PRE-TRAINED)

Model	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
DeepLabv3+*	59.65	75.19	77.36	73.13	78.91	77.80	64.80
DenseASPP	61.46	76.78	77.20	76.36	81.94	80.94	69.14
D-LinkNet34	59.00	74.78	79.17	70.85	81.28	76.34	65.02
D-LinkNet50	58.12	74.04	80.53	68.51	83.13	74.35	64.75
D-LinkNet101	60.24	75.67	79.49	72.20	82.28	77.10	66.37
U-Net*	61.97	76.82	80.84	73.18	83.33	79.26	68.65
Res-U-Net18	63.45	77.89	80.96	75.05	84.70	79.83	70.22
Res-U-Net34	64.03	78.35	81.33	75.58	86.05	79.75	71.10
Res-U-Net50	64.55	78.62	80.80	**76.55**	84.79	81.40	71.59
Res-U-Net101	61.63	76.47	76.42	76.52	78.32	80.48	65.84
Dense-U-Net-121	**65.13**	**79.19**	82.80	75.89	**87.16**	80.50	72.43
Dense-U-Net-169	65.12	78.98	83.90	74.60	85.96	**81.70**	**72.72**
Dense-U-Net-201	65.10	79.01	**84.00**	74.58	86.48	81.18	72.60
Dense-U-Net-121*	62.12	76.99	83.53	71.40	85.34	78.71	69.75

TABLE III
LOSSES COMPARISON FOR DENSE-U-NET-121 ON THE CUSTOM TEST SPLIT OF MA

Loss	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
BCE	57.12	73.03	**81.64**	66.07	90.50	**75.59**	**70.06**
DICE	**57.50**	**73.53**	79.77	68.19	**91.11**	74.88	69.76
IoU	57.19	73.38	78.79	**68.66**	91.02	74.68	69.40
BCE + DICE	57.65	73.43	**79.08**	68.54	**90.16**	75.74	70.08
BCE + IoU	**58.12**	**73.84**	77.37	**70.62**	89.01	**77.13**	**70.48**
BCE + DICE + IoU	57.99	73.71	77.80	70.02	89.44	76.44	70.20
BCE + Sigmoid	**57.88**	**73.51**	**80.05**	67.95	**89.74**	76.50	**70.60**
BCE + Unhinged	57.72	73.47	77.81	**69.58**	89.02	76.16	69.82
BCE + Savage	57.56	73.15	79.36	67.84	88.41	**76.73**	70.06
BCE HB $\beta = 0.7$	57.54	73.08	76.62	69.85	87.90	76.77	69.90
BCE SB $\beta = 0.8$	57.93	73.54	**78.64**	69.06	88.72	**76.88**	70.29
DICE SB $\beta = 0.6$	**58.26**	**73.91**	77.27	**70.83**	**89.82**	76.61	**70.74**

TABLE IV
LOSSES COMPARISON FOR DENSE-U-NET-121 ON THE CUSTOM VALIDATION SPLIT OF DG

Loss	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
BCE	65.13	**79.19**	**82.80**	75.89	**87.16**	80.50	**72.43**
DICE	**65.18**	79.02	78.74	79.30	84.15	82.18	72.08
IoU	63.19	77.58	74.52	**80.90**	78.08	**83.19**	68.01
BCE + DICE	65.43	79.37	78.25	**80.52**	83.49	**83.37**	72.08
BCE + IoU	63.83	78.35	79.63	77.11	**85.28**	79.82	70.73
BCE + DICE + IoU	**65.57**	**79.42**	**80.14**	78.71	85.13	81.90	**72.29**
BCE + Sigmoid	65.76	79.63	**82.08**	77.33	**87.36**	81.14	73.02
BCE + Unhinged	**65.99**	79.63	78.15	**81.16**	84.56	**83.38**	**73.15**
BCE + Savage	65.89	**79.65**	80.24	79.07	85.59	82.49	73.14
BCE HB $\beta = 0.7$	65.80	**79.58**	78.12	81.10	85.29	82.73	73.12
BCE SB $\beta = 0.7$	**65.87**	**79.58**	**78.65**	80.54	**86.08**	82.18	**73.28**
DICE SB $\beta = 0.4$	65.29	79.08	76.28	**82.10**	82.53	**83.63**	71.74

TABLE V
RESULTS COMPARISON FOR DENSE-U-NET-121 TRAINED WITH SYNTHETIC LABEL NOISE ON THE CUSTOM TEST SPLIT OF MA, TRAINED WITH BCE, BCE SB $\beta = 0.7$ AND DICE SB $\beta = 0.7$ LOSSES.

Noise type and level	Loss	Noise Correction	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
Shift1	BCE	×	57.19	73.10	**80.36**	67.05	**89.55**	76.54	**70.27**
	BCE	✓	**57.54**	**73.30**	76.54	70.32	87.62	77.40	69.90
	DICE	✓	56.85	72.85	71.22	**74.56**	85.89	**77.94**	69.24
Shift2	BCE	×	57.01	72.92	**81.33**	66.08	**89.98**	76.44	**70.45**
	BCE	✓	**57.16**	72.99	77.38	69.07	87.92	76.64	69.51
	DICE	✓	56.79	**73.02**	71.68	**74.41**	85.95	**77.86**	68.97
Shift3	BCE	×	47.26	64.47	**85.79**	51.64	**92.87**	66.90	63.42
	BCE	✓	**53.80**	**70.84**	72.34	69.41	90.38	70.22	**65.23**
	DICE	✓	41.63	60.56	49.95	**76.88**	87.79	**71.08**	64.72
Shift4	BCE	×	04.18	08.25	**32.69**	04.72	25.76	09.06	05.95
	BCE	✓	12.16	22.58	29.33	18.35	23.56	23.47	12.43
	DICE	✓	**23.24**	**39.39**	28.75	**62.53**	**30.53**	**40.09**	**20.83**
Ablation1	BCE	×	56.09	71.91	**82.68**	63.63	74.81	69.82	69.82
	BCE	✓	57.15	72.89	79.32	67.43	**89.45**	75.85	**69.91**
	DICE	✓	**57.67**	**73.53**	76.25	**70.99**	88.40	**76.78**	69.85
Ablation2	BCE	×	00.00	00.00	**100.00**	00.00	**100.00**	00.02	00.01
	BCE	✓	38.34	55.35	92.52	39.49	94.72	65.80	63.47
	DICE	✓	**57.11**	**72.92**	77.75	**68.65**	88.05	**77.31**	**70.11**
Shift2+Ablation1	BCE	✓	56.93	72.80	**77.41**	68.70	**87.04**	76.67	69.08
	DICE	✓	**57.42**	**73.22**	73.84	**72.61**	86.02	**77.82**	**69.31**
Shift2+Ablation2	BCE	✓	40.20	57.31	**92.27**	41.56	**94.31**	68.75	65.92
	DICE	✓	**57.06**	**72.60**	76.72	**68.90**	88.27	**76.06**	**69.64**

[12] S. Aich, W. van der Kamp, and I. Stavness, "Semantic binary segmentation using convolutional networks without decoders," in *CVPR Workshops*, 2018.

[13] A. Buslaev, S. Seferbekov, V. Iglovikov, and A. Shvets, "Fully convolutional network for automatic road extraction from satellite imagery," in *CVPR Workshops*, 2018.

[14] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction," in *CVPR Workshops*, 2018.

[15] O. Filin, A. Zapara, and S. Panchenko, "Road detection with eosresunet and post vectorizing algorithm," in *CVPR Workshops*, 2018.

TABLE VI

RESULTS COMPARISON FOR DENSE-U-NET-121 TRAINED WITH SYNTHETIC LABEL NOISE ON THE CUSTOM VALIDATION SPLIT OF DG, TRAINED WITH BCE, BCE SB $\beta = 0.7$ AND DICE SB $\beta = 0.7$ LOSSES.

Noise type and level	Loss	Noise Correction	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
Shift1	BCE	×	66.05	79.49	**82.77**	76.46	85.18	83.13	73.36
	BCE	✓	66.36	79.85	79.96	79.75	83.78	84.06	72.94
	DICE	✓	**68.03**	**81.13**	81.91	**80.37**	**85.40**	**85.15**	**74.77**
Shift2	BCE	×	66.00	79.65	**84.12**	75.64	86.14	82.85	73.61
	BCE	✓	66.03	79.61	79.73	79.49	82.74	**84.77**	72.66
	DICE	✓	**67.72**	**80.91**	82.10	**79.76**	**87.11**	83.52	**74.89**
Shift3	BCE	×	61.14	76.27	**83.64**	70.10	**86.99**	72.26	65.87
	BCE	✓	**63.01**	**77.65**	75.09	80.38	84.29	77.36	68.55
	DICE	✓	59.34	75.21	67.47	**84.95**	85.38	**77.58**	**69.71**
Shift4	BCE	×	34.28	51.10	**74.02**	39.02	**56.48**	33.74	27.47
	BCE	✓	41.80	58.45	56.54	60.50	49.14	47.42	34.48
	DICE	✓	**42.58**	**59.71**	48.47	**77.73**	54.64	**60.21**	**42.76**
Ablation1	BCE	×	65.47	79.32	**84.00**	75.13	86.01	82.62	73.37
	BCE	✓	65.88	79.69	81.86	77.63	86.05	82.35	73.08
	DICE	✓	**67.04**	**80.54**	82.06	**79.08**	**86.34**	**83.80**	**74.46**
Ablation2	BCE	×	00.03	00.06	**99.84**	00.03	86.97	01.10	00.41
	BCE	✓	45.90	63.04	95.26	47.11	**91.99**	64.35	60.87
	DICE	✓	**64.41**	**78.73**	81.93	**75.77**	85.02	**80.78**	**71.31**
Shift2+Ablation1	BCE	✓	65.00	78.92	**83.18**	75.08	**86.23**	80.92	72.42
	DICE	✓	**67.30**	**80.78**	81.14	**80.42**	85.06	**85.19**	**74.61**
Shift2+Ablation2	BCE	✓	47.87	64.86	**93.71**	49.59	**89.79**	69.25	64.29
	DICE	✓	**64.83**	**79.05**	81.33	**76.89**	85.53	**81.01**	**71.69**

TABLE VII

COMPARISON OF DENSE-U-NET-121 WITH THE STATE-OF-THE-ART ON MA'S OFFICIAL TEST SET; * WITHOUT POST-PROCESSING; BEST AND SECOND BEST; TTA - TEST-TIME AUGMENTATION.

Models	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
RSRCNN [4]	49.46	66.20	60.60	72.90	-	-	-
Modified U-Net [5]	59.76	74.54	74.15	75.48	-	-	-
JointNet [6]	64.00	78.05	71.90	**85.36**	-	-	-
WRAU-Net [7]	64.58	78.48	74.50	**82.90**	-	-	-
MFPN [8]	65.70	79.30	**85.10**	74.20	-	-	-
RDRCNN* [9]	66.28	79.72	84.64	75.33	-	-	-
RDRCNN [9]	**67.10**	**80.31**	**85.35**	75.75	-	-	-
Dense-U-Net-121 + BCE SB $\beta = 0.8$ + IoU*	65.16	78.89	79.55	78.25	90.15	86.18	78.85
Dense-U-Net-121 + BCE SB $\beta = 0.8$ + IoU w/ TTA	**66.61**	**79.98**	81.67	78.35	92.93	86.07	80.80

TABLE VIII

COMPARISON OF DENSE-U-NET-121 WITH THE OFFICIAL LEADERBOARD ON THE VALIDATION SET OF DG; OUR REPLICATION OF D-LINKNET RESULTS, ALSO MODIFIED TO USE THE BCE SB LOSS.

Models	IoU
Stacked U-Nets with Multi-Output [10]	60.00
Ensemble U-Nets + CNN + Post-Processing [11]	60.58
ResNet50-D2S [12]	60.60
U-Net-Like-ResNet34 [13]	64.00
D-LinkNet [14]	64.12
D-LinkNet [14]	63.29
D-LinkNet [14] + BCE SB	64.36
EOSResUNet [15]*	65.60
Dense-U-Net-121 + BCE SB $\beta = 0.7$ + Ramp	61.93
Dense-U-Net-121 + BCE SB $\beta = 0.7$ + Ramp w/ TTA	63.52

TABLE IX

RESULTS COMPARISON OF DENSE-U-NET-121 TRAINED WITH DIFFERENT LOSSES ON OUR OWN TEST SPLIT OF CT, COMPARED TO A NOT PRE-TRAINED U-NET. (* NOT DIRECTLY COMPARABLE AS USES A DIFFERENT SPLIT)

Loss	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
Topology loss [3]*	-	-	-	-	88.44	95.13	84.61
BCE [U-Net]	81.79	89.83	**92.72**	87.11	**93.71**	91.37	86.48
BCE	81.96	89.96	91.71	88.28	92.22	92.94	86.54
BCE + IoU	**82.90**	**90.53**	90.78	90.28	92.64	93.41	87.31
BCE + Sigmoid	82.41	90.20	90.56	89.85	91.29	93.34	86.15
BCE SB $\beta = 0.5$	82.32	90.18	91.03	89.34	93.43	93.24	**87.82**
DICE SB $\beta = 0.4$	82.53	90.31	89.98	**90.65**	91.00	**93.79**	86.17

TABLE X

RESULTS COMPARISON OF DENSE-U-NET-121 TRAINED WITH DIFFERENT LOSSES ON OUR OWN TEST SPLIT OF EM, COMPARED TO D-LINKNET50. (* NOT DIRECTLY COMPARABLE AS USES A DIFFERENT SPLIT)

Loss	IoU	F1	Prec.	Rec.	Corr.	Comp.	Qual.
Topology loss [3]*	-	-	-	-	72.27	73.58	57.22
BCE [D-LinkNet50]	64.64	78.53	79.93	77.17	67.62	72.32	53.69
BCE	65.41	79.08	78.95	79.22	66.18	**73.33**	53.34
IoU	66.55	79.93	**80.46**	79.41	**71.53**	71.75	**55.79**
BCE + Unhinged	**67.55**	**80.64**	77.79	**83.70**	68.27	72.94	54.47
BCE SB $\beta = 0.5$	66.41	79.83	80.24	79.42	70.22	72.29	55.31
DICE SB $\beta = 0.5$	66.80	80.09	80.00	80.19	68.18	72.31	54.08

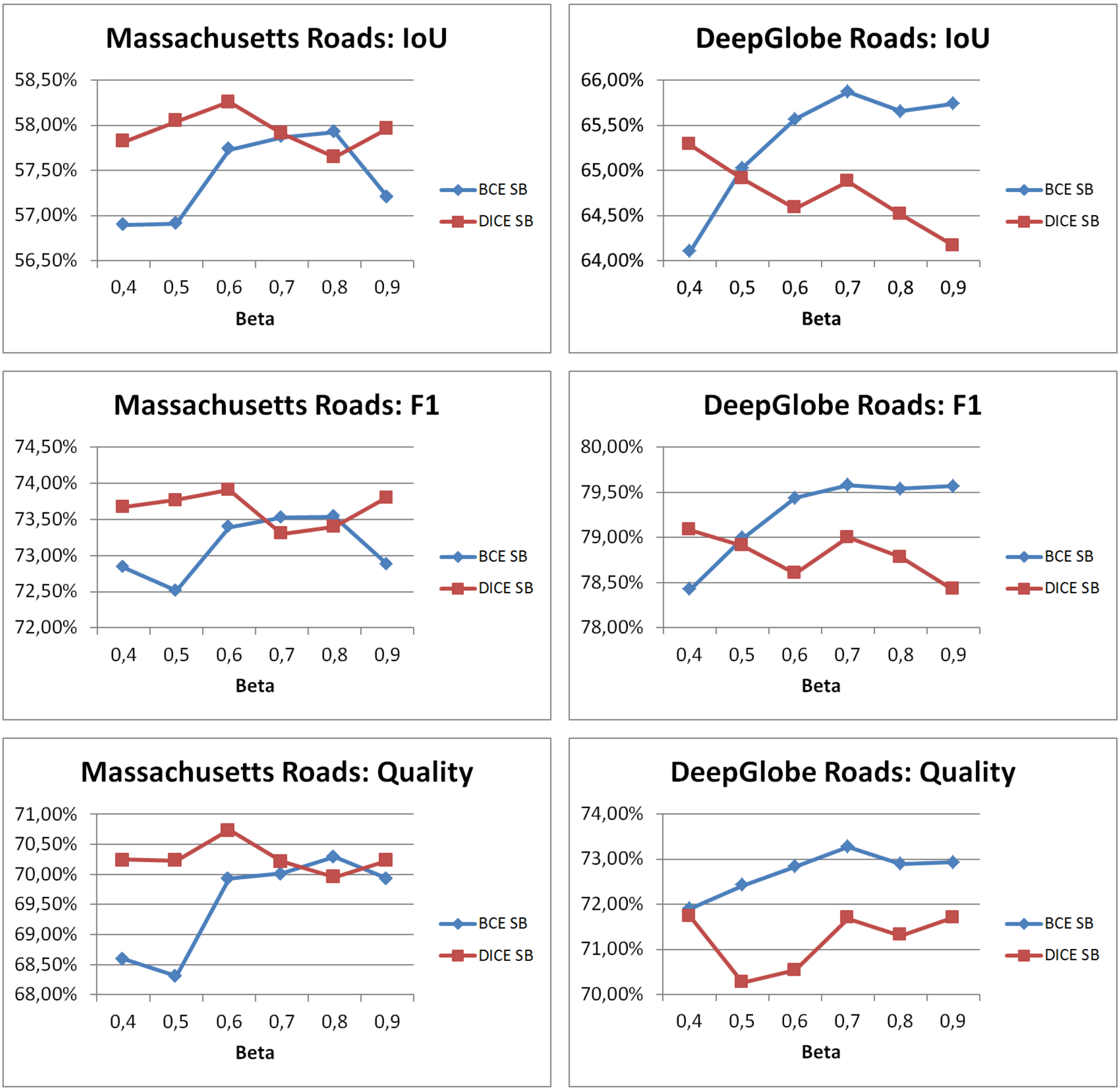

Fig. 1. Cross-validation of the Beta parameters for the Soft-Bootstrapped Cross-Entropy (BCE SB) and Soft-Bootstrapped DICE (DICE SB) on the custom validation sets on Massachusetts Roads dataset and the DeepGlobe Roads dataset.