

ConeQuest: A Benchmark for Cone Segmentation on Mars (Supplementary Material)

Mirali Purohit Jacob Adler Hannah Kerner
Arizona State University
{mpurohi3, jbadler2, hkerner}@asu.edu

A. Visualization of Predictions

This section discusses the visualization of predictions from models and compares inter-region and inter-size category results.

A.1. Single-region Evaluation

Figure 1 shows the visualization of prediction for the evaluation of in-distribution data for single-region training. From Figure 1, it can be observed that Isidis Planitia (IP) performs very well compared to Acidalia Planitia (AP) and Hypanis (HP). Predictions of HP are distorted and the model is not able to predict all the cones from the input sample. As discussed in Section 6.2.1 (main paper), this can be explained by the fact that IP has very dense cones per patch, i.e., an average of 3.52 cones per patch while AP and HP have 1.92 and 2.08 cones, respectively. Also, IP cones have a very different appearance than their surface, i.e., the distinct appearance of cones on the IP surface contributes to easier cone segmentation in this region. For AP and HP, it can be seen from Figure 1 that cones in these two regions are difficult to distinguish from their surface and hence it results in low performance. Also, as discussed in Section 4 and 6.2.1 (main paper) HP cones have more variability which is the cause for low performance.

A.2. Single-category Evaluation

Figure 2 shows the visualization of prediction for evaluating in-distribution data for single-size category training. From Figure 2, it is evident that all models are working better for Large size compared to Medium and Small. Moreover, predictions show improvement as we transition from small to large size. The reason behind this is that the model gets very little information while training only for the small-size category which leads to more FN pixels in prediction. Also, an increase in FN pixels does not only impact pixel-based metrics but highly affects instance-level metrics. This scenario can be observed from Table 3, the range of Object accuracy and Object IoU for small are 16-22 and 22-29, respectively; while for large, this range is 59-64 and 50-60. This indicates

although models are trained on an equal amount of data for all 3 size categories, models struggle on smaller size.

B. Additional Results and Analysis

In this section, we present all the additional results encompassing the performance of each model on in-distribution (\mathcal{D}_{id}), and the detailed analysis of results on out-of-distribution (\mathcal{D}_{ood}). All results are for \mathcal{T}_{pn} (when the model is trained on positive as well as negative samples), denoting the scenario where a model is trained on combined positive and negative samples.

B.1. Results on \mathcal{D}_{id} and \mathcal{D}_{ood}

For BM-1, Table 1 and 2 provide detailed results for each model for single-region and multi-region training, respectively. For BM-2, Table 3 and 4 provide detailed results for each model for single-category and multi-category training, respectively. In these tables, the *highlighted* rows indicate results on \mathcal{D}_{id} , while the *non-highlighted* rows indicate results on the \mathcal{D}_{ood} for both BMs. Detailed analysis on \mathcal{D}_{id} for both BMs can be found in Section 6 (main paper).

B.2. Analysis on \mathcal{D}_{ood}

Here, we provide a detailed analysis of the results on \mathcal{D}_{ood} for both BMs. As discussed in Section 6.2 (main paper), and as observed from Table 1, the models demonstrate lower performance on \mathcal{D}_{ood} compared to \mathcal{D}_{id} , with an average pixel IoU of 12.67 and 21.11 across all models for BM-1 using single-region and multi-region training, respectively. In contrast, the performance on \mathcal{D}_{id} yields an average IoU of 52.06 and 49.68 for the respective configurations. These results indicate that the models do not generalize well to \mathcal{D}_{ood} regions. A similar observation can be made for BM-2 on \mathcal{D}_{ood} , where the average pixel IoU across all models is 25.37 and 30.00 for single-category and multi-category training, respectively. Meanwhile, the performance on \mathcal{D}_{id} yields an average IoU of 42.88 and 41.08 for the corresponding configurations. This further demonstrates that the models struggle to generalize to \mathcal{D}_{ood} categories. Overall, this indicates that

the model’s generalization ability to both out-of-distribution regions and categories is limited, as shown by the lower IoU compared to performance on \mathcal{D}_{id} .

Interestingly, for BM-2, the medium category exhibits relatively better performance when the model is trained on either small or large categories, or trained on a combination of small and large categories (denoted as $S + L$ in tables). This can be attributed to the cone size in the medium category being closer in scale to both small and large categories. As a result, the model is better at handling the medium category when it is unseen during evaluation.

C. \mathcal{T}_{pn} vs. \mathcal{T}_p

In this section, we aim to provide a detailed discussion on having an advantage of incorporating negative samples during training. To accomplish this, we provide the performance of the model on \mathcal{T}_p (when the model is trained only on positive samples), denoting the scenario where the model is exclusively trained on positive samples.

C.1. Evaluation on BM-1

Table 5 and 6 show the results of \mathcal{T}_p for BM-1 on single-region and multi-region training, respectively. Similarly to \mathcal{T}_{pn} , the *highlighted* rows in these tables represent results on \mathcal{D}_{id} , while the *non-highlighted* rows represent results on \mathcal{D}_{ood} .

Performance on Single-region training: When evaluating cone-patches for \mathcal{T}_p on single-region training, we observe an average increase in pixel IoU of 4.18 on \mathcal{D}_{id} and 10.65 on \mathcal{D}_{ood} (across all models) compared to \mathcal{T}_{pn} . However, we do not observe significant improvements in other metrics. For instance, pixel precision shows a performance drop of 4.87 on \mathcal{D}_{id} and 23.43 on \mathcal{D}_{ood} for \mathcal{T}_p compared to \mathcal{T}_{pn} . Furthermore, we observe that the performance on non-cone patches for \mathcal{T}_p on \mathcal{D}_{id} shows the degradation of 1.85 in A_{FP} compared to \mathcal{T}_{pn} . Similarly, on \mathcal{D}_{ood} , the performance drop for A_{FP} is 8.26 for \mathcal{T}_p compared to \mathcal{T}_{pn} .

Performance on Multi-region training: Evaluation of model performance on cone-patches for single-region training, we can observe an average increase of 4.08 in pixel IoU on \mathcal{D}_{id} and 14.21 on \mathcal{D}_{ood} for \mathcal{T}_p compared to \mathcal{T}_{pn} . However, other metrics do not show significant improvements. Whereas, pixel precision decreases by 7.54 on \mathcal{D}_{id} and 18.07 on \mathcal{D}_{ood} for \mathcal{T}_p compared \mathcal{T}_{pn} . Additionally, the performance of non-cone patches deteriorates for \mathcal{T}_p , with A_{FP} showing a degradation of 2.35 on \mathcal{D}_{id} and 4.03 on \mathcal{D}_{ood} compared to \mathcal{T}_{pn} . From this, we can see that although cone-patches benefit from improved pixel IoU on both \mathcal{D}_{id} and \mathcal{D}_{ood} , there is a noticeable decline in pixel precision, and the performance of non-cone patches.

C.2. Evaluation on BM-2

Here, we can see similar observations as BM-1. Table 7 and 8 show results of \mathcal{T}_p for BM-2 on single-category and multi-category training, respectively.

Performance on Single-category training: Similar to the single-region training, evaluation of models on cone-patches for \mathcal{T}_p show an increase in an average pixel IoU by 4.86 (across all models) on \mathcal{D}_{id} and 6.66 (average across all models) on \mathcal{D}_{ood} compared to \mathcal{T}_{pn} for single-category training. However, there is no such significant improvement for other metrics. For instance, pixel precision shows performance drop by 7.08 on \mathcal{D}_{id} and 9.12 on \mathcal{D}_{ood} for \mathcal{T}_p compared to \mathcal{T}_{pn} . Moreover, evaluation of models on non-cone patches exhibits performance deterioration of A_{FP} by 1.44 and 1.39 on \mathcal{T}_p compared to \mathcal{T}_{pn} on \mathcal{D}_{id} and \mathcal{D}_{ood} , respectively.

Performance on Multi-category training: Upon evaluating models on cone-patches for single-category training for \mathcal{T}_p , we observe an increase in pixel IoU of 6.62 (average across all models) on \mathcal{D}_{id} and 8.11 on \mathcal{D}_{ood} when compared to \mathcal{T}_{pn} . However, we do not observe significant improvements in other metrics. Even, pixel precision demonstrates a performance drop of 7.15 on \mathcal{D}_{id} and 9.93 on \mathcal{D}_{ood} for \mathcal{T}_p compared to \mathcal{T}_{pn} . Furthermore, considering the performance on non-cone patches, we observe a degradation of 2.06 in A_{FP} compared to \mathcal{T}_{pn} on \mathcal{D}_{id} . Similarly, on \mathcal{D}_{ood} , the deterioration in performance for A_{FP} is 1.3 for \mathcal{T}_p compared to \mathcal{T}_{pn} .

Based on the aforementioned analysis in C.1 and C.1, it can be concluded that excluding negative samples has a positive impact on the performance of positive samples (cone-patches), but it adversely affects the performance of non-cone patches. This is evident from the degraded pixel precision, indicating that the model generates more false positives due to insufficient information about “*what is not a cone*”. Furthermore, the performance drop on non-cone patches is not desirable as it increases the workload for planetary scientists, since they would need to filter out the increased number of falsely annotated cones generated by the model.

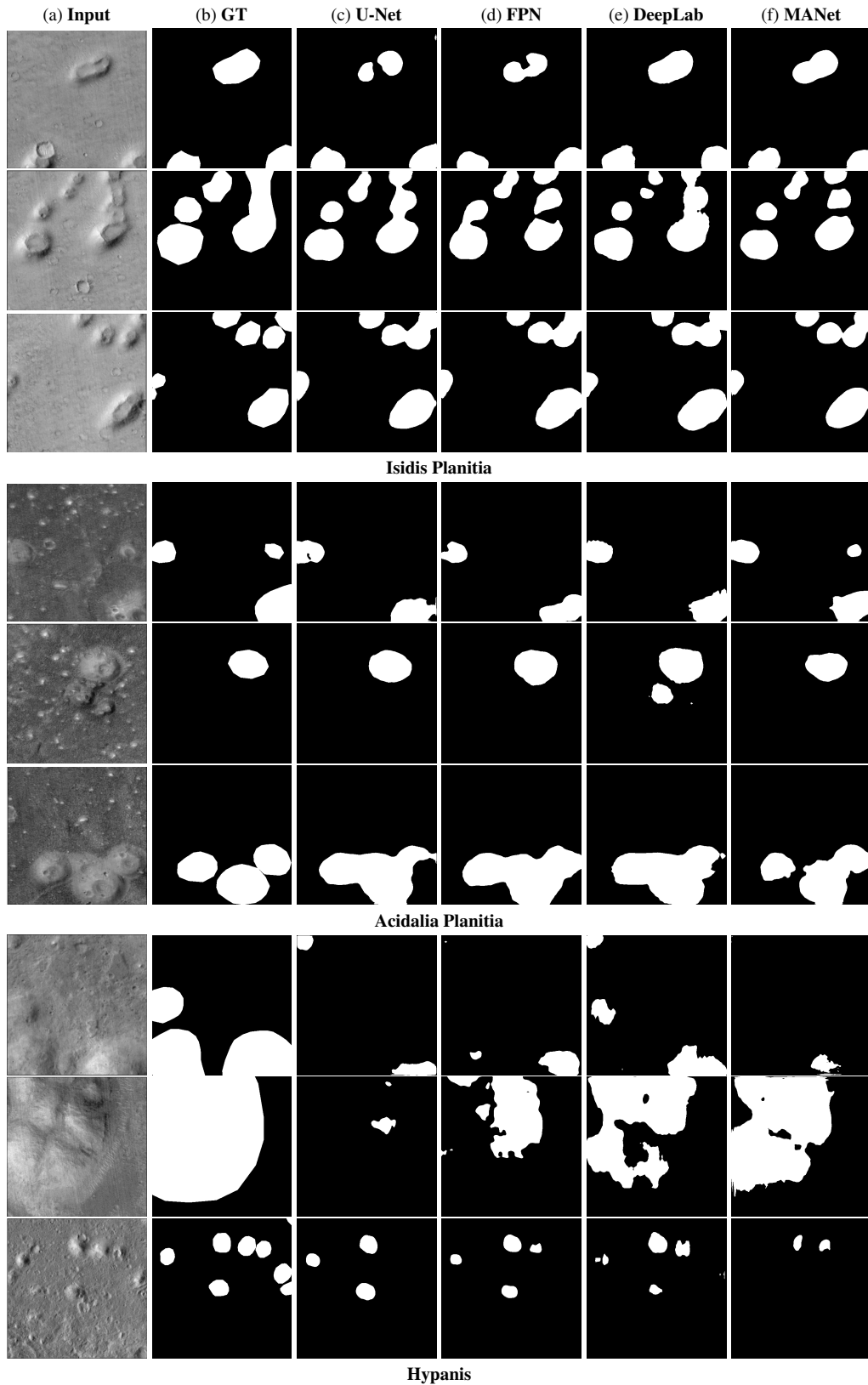


Figure 1. Illustration of predictions of in-distribution evaluation for single-region training (BM-1). Columns c, d, e, and f show prediction from all models with their corresponding Input (column a) and Ground Truth (GT) (column b).

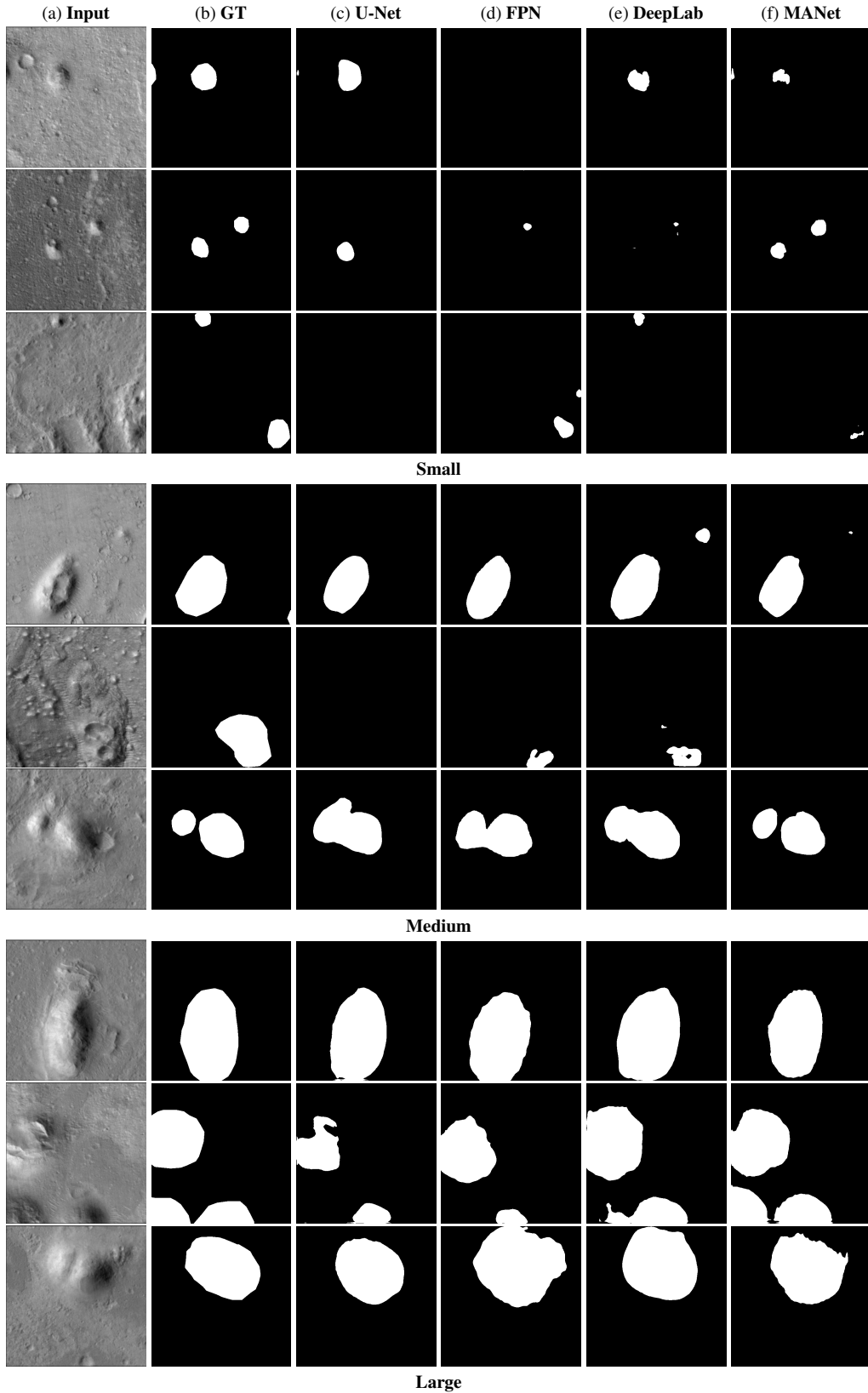


Figure 2. Illustration of predictions of in-distribution evaluation for single-size category training (BM-2). Columns c, d, e, and f show prediction from all models with their corresponding Input (column a) and Ground Truth (GT) (column b).

Training Region	Testing Region	Training Model	Cone											Non-cone
			Mask IoU	Pixel IoU	Pixel Accuracy	Pixel Precision	Pixel Recall	Panoptic Quality	mAP	Object IoU	Object Accuracy	Object Precision	Object Recall	A_{FP}
IP	IP	U-Net	63.90	67.99	96.66	79.56	83.69	54.38	36.67	74.17	59.83	69.13	76.07	0.00
		FPN	69.06	69.49	96.41	83.35	81.44	58.47	31.85	77.04	65.34	76.97	78.76	0.00
		DeepLab	62.11	67.28	96.80	85.41	76.98	53.50	32.29	73.79	58.58	70.08	72.61	0.00
		MA-Net	64.17	66.90	96.77	85.82	74.24	53.12	31.80	72.17	59.08	70.91	74.59	0.00
	AP	U-Net	6.72	3.99	89.58	28.60	9.04	1.75	0.67	2.19	2.47	3.70	2.47	5.70
		FPN	15.03	11.60	79.84	16.97	34.43	1.80	0.96	4.39	1.98	2.16	5.56	19.60
		DeepLab	7.82	4.87	92.75	74.96	5.20	1.46	0.54	2.93	1.23	3.70	1.23	7.45
		MA-Net	16.24	12.69	89.86	35.54	24.30	1.97	0.77	8.93	1.97	3.06	8.02	10.33
	HP	U-Net	25.02	23.29	88.14	69.08	29.01	11.99	3.83	23.37	13.56	16.18	22.41	1.09
		FPN	34.35	29.32	87.03	53.51	48.34	13.76	3.75	34.96	13.4	15.41	34.22	5.13
		DeepLab	24.01	21.57	87.36	70.05	28.67	11.47	3.34	24.51	12.79	15.94	23.72	1.11
		MA-Net	28.40	27.95	89.07	72.11	35.54	14.46	4.45	30.82	15.95	18.64	30.86	1.55
AP	IP	U-Net	8.37	11.21	89.53	46.7	19.71	4.75	2.16	11.28	5.02	7.65	8.80	0.70
		FPN	11.89	10.83	91.32	70.53	12.1	5.42	3.21	13.73	5.93	11.67	8.06	0.00
		DeepLab	0.69	1.53	90.36	95.16	1.86	0.87	0.22	2.18	0.93	1.23	1.85	0.00
		MA-Net	0.00	0.00	90.65	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	AP	U-Net	50.28	49.77	96.32	85.50	57.78	44.83	19.82	58.10	51.45	62.16	57.35	0.08
		FPN	47.96	46.85	96.22	80.55	56.05	38.62	20.43	50.68	45.49	54.94	50.56	0.38
		DeepLab	47.85	49.9	96.58	79.57	58.10	40.45	16.68	51.67	42.59	51.17	52.10	0.38
		MA-Net	34.42	36.81	95.72	86.24	39.15	31.74	20.84	36.73	38.27	40.74	43.46	0.06
	HP	U-Net	2.34	1.63	86.52	75.76	1.93	0.67	1.28	1.69	0.73	1.79	0.98	0.35
		FPN	4.27	3.39	86.79	86.01	3.81	0.86	1.61	1.19	0.93	1.61	1.11	0.13
		DeepLab	1.63	1.62	86.47	91.31	2.73	0.49	0.70	0.73	0.60	0.78	0.90	0.07
		MA-Net	1.50	1.67	86.42	95.25	2.90	0.56	1.73	1.30	0.50	0.55	1.33	0.05
HP	IP	U-Net	25.77	18.64	92.21	97.30	19.69	15.67	9.72	25.68	17.64	28.46	19.81	0.00
		FPN	38.32	33.07	93.60	97.69	34.49	26.41	15.83	39.63	28.72	41.05	31.31	0.00
		DeepLab	27.18	20.23	92.26	92.17	21.28	14.65	9.74	25.64	15.40	28.70	17.35	0.00
		MA-Net	5.00	1.80	88.07	93.15	6.34	2.45	1.63	6.6	2.44	3.00	6.79	0.00
	AP	U-Net	24.79	21.68	93.18	79.08	24.15	16.27	2.17	22.93	20.80	25.31	24.07	11.36
		FPN	11.29	11.08	92.80	91.06	11.42	5.44	1.29	9.14	7.22	8.64	9.88	7.56
		DeepLab	20.26	22.00	94.27	81.52	24.92	13.23	3.32	21.51	16.17	19.14	23.46	7.06
		MA-Net	8.00	8.44	89.93	70.04	10.36	2.06	0.68	4.08	2.59	3.09	4.94	7.21
	HP	U-Net	39.66	38.70	91.22	83.16	45.07	29.48	12.49	39.69	34.82	42.93	39.57	0.43
		FPN	43.08	43.87	92.31	84.70	50.32	33.50	17.35	45.47	39.92	47.88	46.11	0.55
		DeepLab	43.86	45.94	92.63	79.34	56.01	33.11	14.07	47.30	38.82	44.29	48.95	0.65
		MA-Net	38.94	41.20	92.18	83.65	48.54	29.56	13.66	41.19	34.30	39.79	42.60	0.35

Table 1. Results for BM-1 on \mathcal{D}_{id} (highlighted) and \mathcal{D}_{ood} for single-region \mathcal{T}_{pn} .

Training Region	Testing Region	Training Model	Cone											Non-cone
			Mask IoU	Pixel IoU	Pixel Accuracy	Pixel Precision	Pixel Recall	Panoptic Quality	mAP	Object IoU	Object Accuracy	Object Precision	Object Recall	A_{FP}
IP + AP	IP	U-Net	63.30	67.90	96.81	91.35	72.81	52.40	32.59	74.20	57.41	68.76	70.24	0.00
		FPN	66.92	69.44	96.70	82.85	80.91	59.10	36.76	76.35	67.03	76.81	81.1	0.00
		DeepLab	59.87	63.70	96.29	83.85	72.98	48.93	28.91	72.69	53.60	65.31	61.99	0.00
		MA-Net	66.66	69.29	96.91	82.92	79.46	59.41	38.32	76.01	65.50	77.34	77.40	0.00
	AP	U-Net	48.10	47.60	96.26	90.48	52.34	38.05	20.69	54.90	42.07	51.05	55.00	0.09
		FPN	36.08	35.38	95.46	94.01	38.52	28.75	19.82	38.11	33.58	39.20	40.37	0.23
		DeepLab	52.99	55.44	97.25	89.79	61.68	43.66	22.56	63.99	45.62	53.33	60.25	0.88
		MA-Net	44.18	44.61	96.45	86.35	48.84	33.92	20.61	45.09	37.41	45.00	46.05	0.03
	HP	U-Net	8.75	6.92	87.11	84.86	7.28	4.04	2.54	7.03	5.35	8.02	6.56	0.19
		FPN	14.42	12.35	87.46	84.76	13.28	7.15	3.66	10.83	9.71	12.02	11.48	0.49
		DeepLab	18.25	19.55	88.28	84.25	20.93	11.43	4.62	19.16	13.76	16.49	18.69	0.35
		MA-Net	4.26	3.97	86.82	97.59	4.24	2.13	6.29	3.15	2.72	3.22	3.33	0.14
HP + IP	IP	U-Net	58.90	63.58	96.16	78.02	79.80	47.24	25.91	70.93	50.80	65.79	58.76	0.00
		FPN	68.33	67.24	96.56	83.23	78.00	57.53	34.56	75.30	63.97	78.40	71.6	0.00
		DeepLab	59.32	61.55	96.09	83.42	73.72	47.08	22.96	71.52	50.38	65.95	61.28	0.00
		MA-Net	58.44	58.90	96.15	88.61	66.80	48.71	31.32	68.97	53.64	69.11	60.55	0.00
	AP	U-Net	24.45	21.61	94.28	87.44	22.02	12.93	4.08	22.16	16.89	20.49	21.11	6.41
		FPN	12.97	11.01	93.31	89.34	12.31	10.50	1.82	15.01	12.65	16.67	15.43	4.30
		DeepLab	20.19	20.21	93.28	79.97	22.91	15.90	2.29	25.80	19.60	27.38	24.69	10.32
		MA-Net	9.58	9.63	93.18	83.88	10.36	1.86	1.52	4.65	1.85	2.78	4.94	8.10
	HP	U-Net	46.02	45.84	92.5	78.06	55.89	33.59	15.40	46.38	39.58	48.59	46.29	1.19
		FPN	44.16	43.84	92.28	80.92	52.14	33.77	16.23	46.02	38.89	46.49	46.27	0.41
		DeepLab	41.75	45.38	92.34	77.02	56.27	32.49	11.93	46.51	37.34	42.37	47.45	0.77
		MA-Net	39.35	40.72	91.99	81.88	48.35	28.94	13.46	41.63	33.56	39.69	42.51	0.28
AP + HP	IP	U-Net	47.66	46.23	94.76	89.93	51.45	32.25	19.72	52.97	35.36	56.42	41.07	0.00
		FPN	58.25	57.70	95.14	81.41	71.36	44.79	24.25	69.93	48.20	70.31	53.92	0.00
		DeepLab	8.63	7.02	91.40	96.10	7.04	3.58	5.73	6.82	4.32	8.89	4.88	0.00
		MA-Net	36.53	37.11	93.56	90.40	41.35	24.22	15.05	45.95	22.82	36.62	30.63	0.00
	AP	U-Net	35.61	33.17	95.29	89.55	35.46	27.31	8.50	39.26	32.97	41.67	37.90	2.65
		FPN	45.38	44.81	96.22	90.60	49.74	38.21	17.40	47.96	44.81	56.30	46.67	0.70
		DeepLab	16.92	15.93	93.82	94.27	16.66	12.66	2.44	17.09	17.20	21.11	18.52	1.20
		MA-Net	48.19	50.78	96.75	83.64	58.69	35.11	17.87	46.51	36.55	40.12	49.01	1.29
	HP	U-Net	44.36	45.08	92.58	81.46	53.53	33.73	13.71	47.84	38.91	48.31	46.44	1.59
		FPN	44.71	45.79	92.44	81.59	55.21	34.30	16.53	48.14	39.63	47.80	48.00	1.10
		DeepLab	35.33	37.64	91.48	84.88	42.92	26.73	13.08	39.63	31.39	37.54	38.22	0.52
		MA-Net	46.22	47.96	92.48	77.66	60.00	35.27	12.72	52.77	39.46	45.47	53.26	1.11
IP + AP + HP	IP	U-Net	63.46	62.33	96.27	81.92	76.44	53.00	31.88	71.54	59.01	76.71	64.98	0.00
		FPN	65.94	67.91	96.62	85.96	78.00	56.60	33.54	76.45	62.42	78.40	66.03	0.00
		DeepLab	34.08	31.70	93.72	97.37	32.93	23.20	12.38	38.05	27.17	39.51	31.80	0.00
		MA-Net	59.90	63.50	96.44	91.52	67.92	49.38	29.83	69.09	56.98	69.14	65.00	0.00
	AP	U-Net	47.82	42.56	95.99	82.38	47.09	42.49	12.88	57.74	48.89	62.04	56.23	2.27
		FPN	48.26	45.91	96.51	87.12	51.16	40.66	17.63	52.17	45.56	54.57	50.37	0.76
		DeepLab	37.78	40.50	95.91	86.91	45.30	33.36	11.56	45.32	38.27	43.74	45.93	4.55
		MA-Net	49.94	54.05	97.20	83.96	59.85	45.62	23.38	59.47	49.20	55.00	57.47	0.45
	HP	U-Net	43.28	45.23	92.14	79.73	55.20	33.74	11.50	48.29	40.17	46.34	49.30	0.93
		FPN	48.39	49.13	92.98	82.66	58.38	37.96	17.77	50.42	44.69	53.45	52.18	1.36
		DeepLab	37.93	39.83	91.63	81.11	49.48	27.14	11.14	39.53	30.97	36.07	40.91	0.50
		MA-Net	42.44	44.13	92.46	82.88	51.18	32.57	13.42	45.70	38.17	43.90	48.02	0.35

Table 2. Results for BM-1 on \mathcal{D}_{id} (highlighted) and \mathcal{D}_{ood} for multi-region \mathcal{T}_{pn} .

Training Category	Testing Category	Training Model	Cone											Non-cone
			Mask IoU	Pixel IoU	Pixel Accuracy	Pixel Precision	Pixel Recall	Panoptic Quality	mAP	Object IoU	Object Accuracy	Object Precision	Object Recall	A_{FP}
S	S	U-Net	31.28	28.64	97.59	74.69	37.39	19.43	11.08	29.46	22.53	31.36	27.35	0.29
		FPN	25.06	25.06	97.51	76.26	31.68	15.64	8.27	23.31	17.89	21.62	22.65	0.05
		DeepLab	22.86	22.70	97.64	78.85	26.56	13.93	9.58	22.13	16.39	22.06	21.55	0.12
		MA-Net	27.41	26.99	97.70	76.21	33.96	17.56	12.09	25.93	20.55	25.74	27.38	0.03
	M	U-Net	37.26	33.50	92.08	77.37	40.56	25.43	12.36	39.01	31.46	40.73	35.74	0.15
		FPN	27.98	24.97	91.47	88.86	26.86	17.51	12.50	25.61	21.62	27.61	25.46	0.03
		DeepLab	28.97	27.74	91.90	91.70	29.37	20.02	10.37	27.71	25.98	31.30	29.45	0.05
		MA-Net	27.14	27.18	91.76	93.03	29.01	20.33	10.09	30.69	23.22	26.82	29.64	0.02
	L	U-Net	17.72	15.21	79.38	84.34	15.92	9.93	7.22	15.97	11.72	14.02	14.30	0.11
		FPN	15.51	14.86	79.70	91.64	15.33	9.43	5.93	13.24	12.78	15.02	14.91	0.01
		DeepLab	14.93	13.61	78.72	89.78	13.99	8.92	7.08	13.22	11.02	13.23	14.14	0.10
		MA-Net	9.81	9.46	78.01	93.44	9.97	6.96	4.66	9.46	9.08	9.54	12.39	0.05
M	S	U-Net	18.73	17.50	96.96	75.53	22.02	9.95	7.84	15.16	12.15	16.67	14.69	0.57
		FPN	26.27	25.42	97.30	69.99	33.05	16.84	10.43	24.24	20.04	28.44	21.97	0.22
		DeepLab	25.9	24.83	97.01	66.84	32.77	16.56	11.03	28.60	18.60	28.65	23.45	0.48
		MA-Net	29.04	28.68	97.35	70.84	37.84	17.77	11.37	28.71	19.98	28.59	25.22	0.31
	M	U-Net	36.26	33.19	92.43	81.92	37.17	20.92	12.54	36.28	24.23	36.05	28.58	0.35
		FPN	37.53	36.01	92.45	81.41	41.07	25.14	13.28	37.48	30.31	39.71	34.58	0.23
		DeepLab	38.29	40.80	92.82	80.68	48.21	29.42	10.96	44.83	35.23	43.75	43.03	0.31
		MA-Net	47.28	48.65	93.89	82.04	55.15	35.21	13.83	54.01	40.83	50.15	49.81	0.49
	L	U-Net	34.11	33.40	83.53	90.37	34.82	21.58	14.30	29.02	28.02	32.43	32.11	0.35
		FPN	36.97	34.83	83.11	90.43	36.62	26.31	15.41	33.31	32.99	38.30	36.62	0.31
		DeepLab	37.42	37.38	83.88	92.04	39.86	27.29	14.41	39.94	32.95	36.53	43.27	0.36
		MA-Net	43.82	44.02	85.44	91.21	47.17	34.27	18.82	48.67	39.69	44.15	53.21	0.30
L	S	U-Net	11.56	10.63	96.44	74.30	15.40	6.57	1.97	11.81	7.94	11.86	10.27	1.50
		FPN	16.76	15.81	95.30	49.56	27.48	7.47	4.84	13.84	8.58	13.21	11.44	1.39
		DeepLab	8.84	9.01	95.54	65.53	14.45	3.99	1.92	6.75	4.72	6.57	5.86	1.35
		MA-Net	12.73	12.13	93.70	58.47	25.67	6.16	1.81	12.07	7.57	10.08	11.09	0.88
	M	U-Net	36.38	36.29	91.79	77.73	45.80	21.86	7.67	39.88	24.59	34.69	31.55	0.78
		FPN	42.22	41.93	91.84	63.61	59.94	27.04	13.08	43.91	33.15	48.81	39.00	1.20
		DeepLab	31.74	34.10	91.04	69.21	46.22	19.36	6.51	36.39	20.26	30.12	26.26	1.22
		MA-Net	34.41	36.35	90.84	71.19	54.10	22.09	6.97	40.42	25.27	34.41	32.21	0.44
	L	U-Net	56.87	60.42	90.61	89.93	66.75	44.57	21.02	62.86	50.09	57.21	67.81	1.08
		FPN	59.49	63.69	90.40	81.33	75.75	47.49	21.45	59.68	56.28	63.49	64.45	1.00
		DeepLab	61.76	64.81	90.80	82.05	74.87	51.98	24.86	64.94	60.06	68.78	68.43	1.05
		MA-Net	59.47	63.61	90.89	84.67	73.21	47.50	22.85	61.79	53.68	59.97	66.51	0.38

Table 3. Results for BM-2 on \mathcal{D}_{id} (highlighted) and \mathcal{D}_{ood} for single-size category \mathcal{T}_{pn} .

Training Category	Testing Category	Training Model	Cone											Non-cone
			Mask IoU	Pixel IoU	Pixel Accuracy	Pixel Precision	Pixel Recall	Panoptic Quality	mAP	Object IoU	Object Accuracy	Object Precision	Object Recall	A_{FP}
S + M	S	U-Net	32.50	30.85	97.84	79.57	37.36	21.67	14.14	30.52	25.34	34.08	28.33	0.25
		FPN	24.86	23.97	97.53	82.85	28.39	16.36	7.42	24.44	19.53	27.48	22.46	0.13
		DeepLab	25.06	25.68	97.77	85.34	30.06	17.65	11.67	24.85	20.63	27.25	24.85	0.03
		MA-Net	39.17	37.39	97.52	62.84	54.41	26.54	14.21	41.84	30.58	38.42	41.57	0.42
	M	U-Net	45.19	41.88	93.27	85.45	47.18	34.49	19.15	48.00	41.69	54.05	46.08	0.50
		FPN	36.24	33.26	92.22	84.74	37.58	26.20	13.15	39.57	32.25	42.61	37.31	0.26
		DeepLab	29.96	28.14	91.80	91.75	30.26	20.13	11.96	29.31	24.94	31.90	28.41	0.07
		MA-Net	49.86	50.88	94.06	81.20	61.02	38.05	18.00	55.60	43.67	50.20	56.56	0.55
	L	U-Net	32.82	29.90	82.65	92.01	31.07	24.13	14.53	33.21	29.35	34.88	34.56	0.05
		FPN	24.17	22.83	81.06	93.26	23.51	16.70	10.72	22.56	22.63	25.61	25.61	0.09
		DeepLab	18.60	16.48	79.34	93.39	16.83	13.61	8.42	19.94	16.45	20.44	19.95	0.05
		MA-Net	50.00	51.11	86.35	90.39	56.31	39.35	18.89	53.78	44.56	48.19	63.07	0.33
M + L	S	U-Net	24.94	23.63	96.83	69.32	30.29	14.14	7.37	21.23	16.79	24.06	18.68	1.00
		FPN	19.34	19.01	96.60	73.26	25.05	8.21	7.23	15.24	8.94	14.33	11.87	0.65
		DeepLab	16.52	17.24	96.95	80.52	21.44	9.82	8.26	15.60	11.16	16.74	12.87	1.06
		MA-Net	24.41	22.42	97.02	76.47	28.04	13.51	6.36	21.96	15.71	23.10	17.55	0.81
	M	U-Net	47.06	46.28	92.98	75.76	55.25	32.41	14.89	48.40	37.85	50.29	43.68	0.70
		FPN	40.06	40.52	92.56	76.16	48.22	26.19	15.70	41.77	31.19	42.90	36.84	0.48
		DeepLab	40.07	40.49	92.49	81.37	47.06	27.67	10.12	44.56	31.98	43.67	37.41	0.51
		MA-Net	43.92	41.96	92.71	82.85	48.76	31.25	14.89	45.75	36.98	50.01	41.83	0.33
	L	U-Net	49.92	52.92	88.97	88.46	58.01	39.83	17.20	53.63	47.05	53.07	58.56	2.57
		FPN	51.83	53.70	88.67	88.55	59.06	41.37	17.94	54.81	48.26	56.02	57.57	2.88
		DeepLab	60.37	64.01	91.55	91.58	69.28	50.19	23.65	65.89	56.61	63.98	68.73	2.31
		MA-Net	57.52	60.06	90.42	92.17	64.18	48.82	23.54	61.74	57.25	63.24	67.13	1.76
L + S	S	U-Net	15.61	14.35	97.02	84.55	18.09	7.80	5.10	12.39	9.14	14.34	10.32	0.93
		FPN	23.82	24.01	97.28	81.13	32.41	16.12	6.93	23.82	18.47	24.34	23.08	0.48
		DeepLab	18.62	18.95	97.22	82.84	22.90	11.15	5.53	17.85	13.49	19.31	16.04	1.19
		MA-Net	32.27	31.44	97.16	65.87	45.44	19.56	7.33	31.66	22.91	27.81	33.09	0.95
	M	U-Net	34.43	35.82	92.78	86.03	40.87	22.62	9.83	35.77	26.82	37.47	29.91	0.38
		FPN	37.92	37.05	92.63	85.55	42.08	28.02	12.42	40.55	34.11	44.38	39.48	0.20
		DeepLab	31.96	34.30	92.10	81.33	40.50	20.40	8.11	36.56	23.43	30.83	29.79	0.31
		MA-Net	49.29	50.29	92.94	75.99	63.65	34.61	16.41	55.66	37.56	45.75	53.47	0.23
	L	U-Net	50.51	52.01	89.13	91.55	56.26	42.93	21.49	52.84	49.93	58.67	56.80	2.22
		FPN	49.39	53.22	88.22	87.40	60.37	40.01	20.24	51.47	45.58	50.81	55.73	1.40
		DeepLab	49.69	57.35	89.76	86.88	64.73	39.20	14.86	52.72	44.78	49.30	57.57	2.36
		MA-Net	58.73	62.65	90.80	84.17	71.61	43.32	14.74	59.74	49.18	52.64	67.89	2.50

Table 4. Results for BM-2 on \mathcal{D}_{id} (highlighted) and \mathcal{D}_{ood} for multi-size category \mathcal{T}_{pn} .

Training Region	Testing Region	Training Model	Cone											Non-cone
			Mask IoU	Pixel IoU	Pixel Accuracy	Pixel Precision	Pixel Recall	Panoptic Quality	mAP	Object IoU	Object Accuracy	Object Precision	Object Recall	A_{FP}
IP	IP	U-Net	62.62	65.60	96.75	89.51	71.99	52.27	31.84	73.62	58.50	68.63	72.12	0.00
		FPN	57.42	59.32	96.33	88.83	65.23	44.91	25.66	63.75	51.05	64.82	57.57	0.00
		DeepLab	57.41	64.37	96.61	85.27	72.34	45.88	29.38	69.45	49.98	57.94	68.32	0.00
		MA-Net	59.32	64.96	96.45	84.37	71.98	47.67	29.37	72.73	53.55	62.79	67.8	0.00
	AP	U-Net	17.36	14.36	91.47	62.60	17.80	6.71	5.60	17.73	6.04	10.54	12.20	9.96
		FPN	19.00	14.60	90.13	43.44	21.25	8.89	4.98	14.74	10.06	12.74	16.07	8.52
		DeepLab	21.22	14.94	90.07	37.72	26.57	4.63	3.05	17.99	4.42	4.90	19.17	7.46
		MA-Net	30.22	24.97	88.52	35.44	47.87	6.75	5.89	28.39	5.66	6.36	25.60	19.08
	HP	U-Net	30.05	27.92	89.74	72.56	35.22	16.87	7.11	28.85	19.65	23.59	30.59	1.31
		FPN	22.14	20.64	89.14	74.23	25.80	11.06	5.81	18.29	13.49	15.81	19.34	0.94
		DeepLab	24.54	24.03	89.55	68.04	29.27	11.81	4.29	25.86	12.85	14.26	27.89	0.95
		MA-Net	34.04	30.93	85.25	50.10	54.18	12.36	4.71	33.21	12.02	14.36	35.86	4.73
AP	IP	U-Net	7.00	11.82	32.22	12.95	86.48	0.79	1.81	4.15	0.78	1.15	1.85	85.26
		FPN	13.89	11.20	82.17	17.89	27.63	3.04	3.02	11.50	2.84	6.36	5.56	10.84
		DeepLab	8.24	6.53	78.24	10.92	17.76	0.22	1.11	1.89	0.23	0.25	1.85	10.71
		MA-Net	9.89	8.24	62.47	12.10	42.58	0.61	2.16	4.21	0.59	0.73	2.16	26.54
	AP	U-Net	61.30	60.29	96.48	70.42	74.78	50.23	27.90	68.59	52.69	60.54	73.21	3.30
		FPN	52.90	58.17	96.95	77.54	65.10	48.34	26.53	63.48	54.21	60.83	65.00	2.33
		DeepLab	53.78	59.87	96.82	76.65	69.44	48.19	25.25	62.99	54.44	60.71	66.19	1.34
		MA-Net	48.63	51.89	96.19	65.00	65.06	41.65	18.18	61.44	43.57	48.22	59.94	1.87
	HP	U-Net	26.04	22.67	71.75	29.07	66.11	7.27	2.69	20.72	7.09	8.72	22.32	30.40
		FPN	23.26	21.69	87.68	55.24	30.88	9.68	5.29	19.44	10.95	13.80	19.97	4.38
		DeepLab	19.10	17.96	86.75	50.37	26.18	5.97	3.90	14.42	6.69	8.42	15.02	5.20
		MA-Net	13.49	14.38	86.87	51.64	19.95	4.30	2.02	11.61	4.75	5.59	12.06	6.34
HP	IP	U-Net	42.62	44.23	93.79	74.00	59.09	24.89	9.99	54.69	27.18	45.99	33.49	0.00
		FPN	40.02	33.78	85.67	66.90	44.59	24.52	10.73	47.33	25.82	47.28	30.31	0.00
		DeepLab	40.55	38.44	86.34	72.64	46.94	27.37	11.16	52.15	29.55	50.43	33.92	0.00
		MA-Net	39.50	43.19	94.25	73.50	56.95	24.60	10.68	48.44	27.67	47.22	33.03	0.00
	AP	U-Net	30.31	33.29	92.29	67.55	44.88	17.64	8.49	26.81	20.77	25.89	27.68	9.29
		FPN	26.80	27.06	87.92	52.61	44.57	17.44	7.43	25.67	23.15	26.19	28.87	20.00
		DeepLab	28.48	27.79	89.65	67.22	39.68	18.86	8.56	31.81	22.47	27.56	31.85	13.29
		MA-Net	26.24	25.18	89.08	62.47	39.08	13.45	5.01	20.77	16.82	21.90	23.21	9.44
	HP	U-Net	48.45	48.51	93.41	77.77	58.63	36.34	18.13	50.38	42.28	51.18	49.52	2.66
		FPN	47.44	48.59	90.74	71.01	64.21	35.64	16.47	49.67	42.14	50.85	50.47	5.95
		DeepLab	47.14	47.59	91.34	76.21	56.85	34.61	15.07	49.80	39.01	46.65	50.54	4.57
		MA-Net	44.24	45.65	93.35	75.80	57.09	31.12	14.38	44.66	36.60	45.22	44.90	3.06

Table 5. Results for BM-1 on \mathcal{D}_{id} (highlighted) and \mathcal{D}_{ood} for single-region \mathcal{T}_p .

Training Region	Testing Region	Training Model	Cone											Non-cone
			Mask IoU	Pixel IoU	Pixel Accuracy	Pixel Precision	Pixel Recall	Panoptic Quality	mAP	Object IoU	Object Accuracy	Object Precision	Object Recall	A_{FP}
IP + AP	IP	U-Net	61.50	66.48	97.01	87.87	74.32	52.03	32.51	69.87	59.15	66.95	71.23	0.00
		FPN	56.59	58.70	96.07	87.53	64.58	42.66	26.58	66.17	47.27	62.88	54.58	0.00
		DeepLab	60.90	64.36	96.64	84.67	70.95	51.17	31.75	71.11	55.63	69.00	66.96	0.00
		MA-Net	58.47	63.56	96.74	86.47	69.49	49.84	28.39	68.69	56.31	67.38	64.22	0.00
	AP	U-Net	58.92	61.85	96.87	81.31	75.19	52.98	27.07	69.80	62.13	67.54	78.81	1.72
		FPN	59.29	60.35	96.82	77.09	72.28	51.50	27.62	62.39	58.49	65.48	65.60	1.74
		DeepLab	53.73	54.45	96.33	74.81	67.67	45.62	25.61	61.44	49.31	56.33	66.07	1.67
		MA-Net	51.51	53.54	96.17	60.95	72.16	39.28	17.02	59.43	42.01	48.10	63.69	1.98
	HP	U-Net	29.49	26.65	89.32	72.37	32.54	16.56	7.81	27.71	20.17	23.87	29.09	0.96
		FPN	33.10	30.94	89.74	70.08	39.60	18.31	8.77	31.06	20.91	26.46	30.47	1.54
		DeepLab	29.82	29.64	89.67	70.29	35.08	16.69	6.19	28.98	20.16	23.59	29.69	1.60
		MA-Net	27.39	27.68	89.68	71.80	32.66	14.43	6.94	25.10	16.61	18.69	27.65	1.10
HP + IP	IP	U-Net	58.85	58.41	96.07	84.29	65.90	44.36	26.65	67.89	46.35	66.54	52.91	0.00
		FPN	57.66	52.00	88.61	74.34	63.95	48.66	26.92	71.01	53.49	67.79	61.74	0.00
		DeepLab	43.94	46.36	94.62	75.83	49.39	32.80	15.98	49.12	37.55	48.60	41.23	0.00
		MA-Net	60.31	66.09	96.84	79.02	78.66	49.03	28.70	65.66	54.86	67.78	62.98	0.00
	AP	U-Net	28.83	25.69	86.98	62.01	44.35	16.60	5.66	30.20	20.29	24.90	29.46	15.92
		FPN	23.31	18.45	75.74	40.23	49.16	9.05	5.24	18.89	9.97	12.50	20.24	28.93
		DeepLab	26.28	26.44	91.28	63.61	36.22	17.15	6.33	34.27	18.98	23.54	35.42	11.26
		MA-Net	36.88	38.31	88.87	56.01	65.51	22.76	6.88	43.29	24.17	26.01	50.30	17.39
	HP	U-Net	49.65	50.26	93.21	75.10	62.85	38.16	18.04	53.79	43.24	51.10	53.90	3.63
		FPN	48.47	48.75	90.99	70.00	61.13	35.81	16.07	51.31	41.23	47.65	53.13	4.73
		DeepLab	46.02	47.94	93.65	77.01	57.39	32.74	13.25	48.65	37.10	43.59	49.82	3.44
		MA-Net	53.55	54.97	93.53	71.13	71.95	38.54	15.68	54.97	43.01	48.87	58.62	3.32
AP + HP	IP	U-Net	50.31	54.92	95.76	79.95	65.46	37.95	21.71	60.84	42.13	56.79	47.64	0.00
		FPN	55.44	55.73	96.10	83.86	63.87	41.62	21.88	63.45	46.56	66.36	50.1	0.00
		DeepLab	42.84	39.54	93.91	83.26	48.83	29.48	14.19	54.75	32.28	50.93	36.79	0.00
		MA-Net	50.74	49.80	95.24	79.57	60.62	34.37	16.71	55.95	40.93	56.61	44.42	0.00
	AP	U-Net	35.33	40.46	94.43	76.45	49.29	31.37	17.61	43.56	35.72	39.11	42.68	10.03
		FPN	45.35	46.30	94.31	79.74	54.31	35.64	21.74	50.92	39.70	49.70	49.82	8.70
		DeepLab	48.83	48.55	93.57	72.22	64.70	44.66	18.93	60.77	49.14	57.38	61.01	8.42
		MA-Net	50.52	54.94	94.55	71.33	72.74	45.68	20.74	61.35	50.54	54.11	65.77	7.58
	HP	U-Net	48.60	49.96	93.50	72.92	62.18	36.59	18.55	52.11	41.96	48.33	52.05	3.20
		FPN	49.84	49.54	93.68	79.43	58.12	38.33	22.65	52.06	45.35	53.36	54.14	1.58
		DeepLab	50.68	51.57	93.77	74.41	64.10	37.71	19.74	53.36	42.76	49.97	54.71	2.23
		MA-Net	50.92	51.80	93.86	72.27	64.46	38.08	17.77	53.54	42.93	49.99	55.24	2.36
IP + AP + HP	IP	U-Net	57.63	55.90	96.04	88.12	60.07	43.47	26.69	67.67	46.89	64.75	52.66	0.00
		FPN	57.13	59.66	96.01	83.72	67.59	43.78	24.63	66.97	48.26	65.74	55.51	0.00
		DeepLab	54.36	53.52	95.65	86.86	58.72	37.29	18.87	58.08	40.87	58.44	45.41	0.00
		MA-Net	55.02	55.88	95.96	86.90	61.55	42.32	22.09	64.19	45.25	60.29	53.33	0.00
	AP	U-Net	54.48	50.87	94.01	73.59	67.10	44.62	27.66	54.02	53.69	59.17	61.61	9.45
		FPN	47.84	48.94	93.87	74.32	60.52	39.56	21.61	52.21	44.11	50.89	54.46	10.58
		DeepLab	48.26	49.39	94.08	71.10	64.71	40.16	22.92	53.11	46.67	50.00	60.12	9.73
		MA-Net	53.83	57.92	96.43	76.17	69.55	47.61	23.61	66.01	49.88	56.55	65.18	1.12
	HP	U-Net	49.89	49.55	93.70	77.38	60.54	36.62	19.47	52.64	42.47	50.65	53.07	2.14
		FPN	50.17	51.84	93.60	70.81	67.14	35.99	18.53	53.19	39.18	46.19	53.39	3.20
		DeepLab	45.72	47.13	92.72	76.59	56.76	33.12	15.92	48.62	38.06	43.59	50.52	3.96
		MA-Net	52.37	53.25	94.11	76.63	65.24	40.47	18.87	55.09	46.27	53.57	56.69	3.44

Table 6. Results for BM-1 on \mathcal{D}_{id} (highlighted) and \mathcal{D}_{ood} for multi-region \mathcal{T}_p .

Training Category	Testing Category	Training Model	Cone											Non-cone
			Mask IoU	Pixel IoU	Pixel Accuracy	Pixel Precision	Pixel Recall	Panoptic Quality	mAP	Object IoU	Object Accuracy	Object Precision	Object Recall	
S	S	U-Net	26.56	24.68	96.88	70.36	29.01	17.04	8.59	24.99	21.74	28.00	23.68	0.30
		FPN	30.00	29.05	97.08	62.67	39.30	19.10	12.20	34.10	21.76	26.54	32.39	0.48
		DeepLab	29.54	30.56	97.21	66.39	37.39	21.15	9.14	34.46	24.95	29.57	32.72	0.26
		MA-Net	29.93	30.58	97.24	66.91	38.51	20.39	10.33	32.80	23.82	28.24	30.75	0.32
	M	U-Net	36.16	34.46	95.60	86.72	37.74	21.62	11.24	39.11	26.27	33.24	35.99	0.20
		FPN	53.46	48.02	96.43	86.15	52.11	30.98	13.58	49.78	34.51	47.06	50.56	0.39
		DeepLab	40.88	41.49	96.44	76.02	44.54	28.66	14.95	44.01	31.75	35.29	47.48	0.17
		MA-Net	41.93	39.83	96.04	82.33	43.67	28.77	10.26	42.26	33.04	38.45	42.86	0.33
	L	U-Net	10.63	10.91	81.04	70.44	11.15	3.98	5.08	5.42	6.06	6.82	6.82	0.19
		FPN	22.80	22.28	82.91	80.77	23.51	13.28	7.15	17.47	18.94	21.97	22.73	0.46
		DeepLab	18.65	19.25	81.99	80.70	19.68	13.79	7.91	18.38	17.95	18.94	21.97	0.28
		MA-Net	18.07	18.89	82.06	93.43	19.15	13.75	9.17	20.62	17.80	20.83	27.27	0.45
M	S	U-Net	21.64	22.35	95.91	49.27	32.77	7.66	11.26	17.53	7.89	14.21	12.02	1.70
		FPN	30.18	28.35	96.24	56.61	41.94	18.65	10.20	30.62	21.93	30.98	26.73	1.17
		DeepLab	27.76	31.39	96.54	61.04	45.49	20.27	11.12	33.31	22.67	27.88	27.21	1.17
		MA-Net	32.44	30.46	95.61	50.99	47.27	14.36	11.53	24.42	14.74	23.08	22.88	1.73
	M	U-Net	47.18	47.69	93.93	67.30	63.57	32.28	14.04	52.66	37.24	45.25	53.31	1.19
		FPN	51.33	51.73	94.97	77.50	61.96	40.38	20.35	55.67	48.29	56.54	57.76	0.76
		DeepLab	51.13	52.67	95.14	77.23	63.83	38.43	17.29	58.28	43.19	49.70	60.41	0.76
		MA-Net	51.97	52.89	94.91	77.20	64.18	40.88	16.73	58.35	48.86	58.18	59.27	1.48
	L	U-Net	41.63	43.67	86.65	83.07	49.41	26.60	10.38	42.56	32.99	38.10	49.24	1.30
		FPN	38.11	37.96	86.00	83.91	40.91	29.19	15.58	38.53	37.01	43.33	42.42	0.90
		DeepLab	45.58	49.24	87.55	84.77	53.45	36.46	19.61	58.41	40.06	44.47	54.92	0.97
		MA-Net	53.00	57.38	89.35	90.03	61.66	44.45	27.76	57.72	51.94	60.38	60.61	1.59
L	S	U-Net	15.59	11.93	91.97	73.41	21.34	5.73	3.05	12.70	6.76	11.54	9.58	3.78
		FPN	23.91	24.14	91.23	44.07	43.74	12.12	6.57	22.01	14.06	21.70	17.79	5.11
		DeepLab	23.18	23.40	92.26	56.19	42.89	9.65	6.59	18.24	10.71	17.31	12.02	5.17
		MA-Net	23.72	24.60	91.88	41.44	44.99	13.39	8.02	30.55	14.56	22.12	21.76	4.09
	M	U-Net	34.11	28.90	90.61	70.71	39.67	22.42	8.65	33.52	26.88	34.56	32.91	2.91
		FPN	41.92	38.74	88.21	55.71	61.16	21.32	9.96	32.40	26.41	30.88	35.99	4.14
		DeepLab	40.98	38.53	91.62	61.08	62.69	24.85	10.65	35.97	30.69	36.27	37.82	3.99
		MA-Net	42.45	42.44	92.79	49.20	66.00	27.91	9.89	40.63	32.88	38.73	44.82	3.39
	L	U-Net	55.82	58.25	87.79	85.57	67.06	46.60	25.31	58.41	54.38	61.85	63.82	3.79
		FPN	61.67	61.98	88.45	77.24	78.36	50.62	22.26	66.47	57.63	64.3	75.84	5.09
		DeepLab	62.52	63.98	88.98	76.67	82.22	49.71	23.42	62.77	57.62	64.56	70.18	4.28
		MA-Net	64.51	68.85	91.13	80.00	84.16	50.94	21.08	68.08	55.75	62.40	72.97	3.91

Table 7. Results for BM-2 on \mathcal{D}_{id} (highlighted) and \mathcal{D}_{ood} for single-size category \mathcal{T}_p .

Training Category	Testing Category	Training Model	Cone											Non-cone
			Mask IoU	Pixel IoU	Pixel Accuracy	Pixel Precision	Pixel Recall	Panoptic Quality	mAP	Object IoU	Object Accuracy	Object Precision	Object Recall	A_{FP}
S + M	S	U-Net	32.65	32.54	96.84	66.34	45.60	23.83	13.45	37.43	28.21	36.64	34.33	1.41
		FPN	21.64	19.95	96.75	77.80	23.45	11.08	10.53	21.85	13.13	19.79	16.83	0.46
		DeepLab	30.66	31.66	97.29	71.83	41.11	21.60	11.69	34.84	24.49	30.59	31.29	1.14
		MA-Net	34.60	35.01	97.25	63.54	47.05	22.53	15.60	36.01	25.65	30.99	35.54	1.49
	M	U-Net	53.70	54.59	94.62	70.96	73.28	38.61	16.33	59.33	44.70	51.06	60.91	4.28
		FPN	42.45	39.61	94.54	90.53	42.62	27.66	14.84	43.06	33.99	44.23	40.76	1.01
		DeepLab	51.10	51.70	95.08	77.11	63.31	37.44	18.97	53.76	42.44	47.81	56.31	2.67
		MA-Net	53.55	55.13	95.15	77.39	66.78	37.15	16.66	56.83	40.86	45.69	59.90	2.74
	L	U-Net	46.93	45.05	87.50	82.83	48.95	37.00	17.43	51.50	41.86	52.65	51.89	0.77
		FPN	33.00	31.85	84.94	89.20	32.55	25.21	11.35	34.09	32.58	39.02	37.88	0.40
		DeepLab	35.72	38.63	85.91	83.11	41.14	27.82	9.82	42.09	31.69	33.00	43.56	0.75
		MA-Net	51.80	54.29	88.42	89.20	59.73	38.23	19.24	56.47	41.01	44.67	58.33	0.97
M + L	S	U-Net	29.02	26.06	95.54	71.13	35.95	18.29	8.87	30.77	20.82	28.55	26.25	2.32
		FPN	21.94	19.10	95.48	62.98	28.58	11.27	7.25	21.26	12.54	18.91	17.44	1.65
		DeepLab	35.51	29.91	95.20	65.54	50.61	23.22	11.11	41.13	25.57	33.43	38.11	2.45
		MA-Net	27.86	28.72	94.69	46.06	46.68	16.77	7.75	31.23	19.79	23.85	28.37	2.25
	M	U-Net	52.22	52.67	94.76	79.65	64.14	38.98	18.20	56.44	46.22	56.26	54.61	1.99
		FPN	45.55	43.95	94.32	81.91	53.34	31.62	14.57	47.73	38.69	47.52	47.60	1.15
		DeepLab	55.40	55.62	94.60	71.99	70.30	43.07	17.08	60.28	50.14	56.14	63.17	2.09
		MA-Net	49.59	50.54	94.31	76.50	64.11	36.45	18.05	54.67	42.68	49.39	55.70	1.87
	L	U-Net	61.63	62.74	90.89	86.69	70.60	50.47	27.10	61.61	57.98	64.60	67.00	3.61
		FPN	51.84	54.54	89.00	88.06	59.12	44.69	22.92	58.17	53.63	59.45	66.02	2.80
		DeepLab	64.36	67.11	90.97	83.98	77.64	51.65	19.56	67.81	55.80	60.63	72.80	4.69
		MA-Net	60.98	63.59	90.44	83.96	73.04	49.26	23.85	65.65	54.97	60.66	74.34	4.14
L + S	S	U-Net	31.94	30.27	95.95	59.95	44.87	20.85	12.19	36.11	23.10	29.22	33.25	2.82
		FPN	26.01	24.83	94.38	61.55	40.27	14.62	10.81	25.11	17.67	25.29	21.11	4.06
		DeepLab	26.52	26.52	96.41	68.21	36.65	15.41	9.81	28.44	18.14	26.22	22.74	2.54
		MA-Net	33.19	32.33	96.45	57.88	47.59	21.23	12.37	34.69	23.37	29.59	33.28	2.37
	M	U-Net	52.29	49.67	96.17	70.60	63.37	31.90	15.00	50.57	34.40	39.89	58.12	1.90
		FPN	40.17	39.37	93.38	65.25	63.57	23.86	12.16	34.95	30.92	37.25	36.97	3.49
		DeepLab	39.68	45.69	96.20	79.97	56.90	30.96	12.51	49.82	35.98	41.76	52.66	1.63
		MA-Net	49.84	49.07	95.71	72.53	60.84	31.44	12.59	52.68	35.88	42.35	58.54	2.23
	L	U-Net	64.87	66.61	91.06	82.20	78.37	50.93	23.53	69.71	54.54	58.90	74.59	5.95
		FPN	61.10	63.40	89.52	78.31	78.76	47.69	25.40	58.50	55.44	62.77	66.70	8.78
		DeepLab	61.30	64.56	91.32	85.28	72.91	49.41	23.20	65.23	55.25	60.36	72.23	4.65
		MA-Net	63.26	65.60	90.90	81.93	78.18	47.95	19.35	66.31	51.76	55.57	75.93	6.45

Table 8. Results for BM-2 on \mathcal{D}_{id} (highlighted) and \mathcal{D}_{ood} for multi-size category \mathcal{T}_p .