IDD-AW: A Benchmark for Safe & Robust Segmentation of Drive Scenes in Unstructured Traffic & Adverse Weather

Furqan Ahmed Shaik[1], Abhishek Malreddy[1], Nikhil Reddy Billa[1], Kunal Chaudhary[2], Sunny Manchanda[2], Girish Varma[1]
1. Machine Learning Lab, IIIT Hyderabad 2. DYSL-AI, DRDO | http://iddaw.github.io

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD
IIIT Hyderabad
JAN 4-8 WACV 2024 WAIKOLOA HAWAII

Safety and Robustness in Segmentation

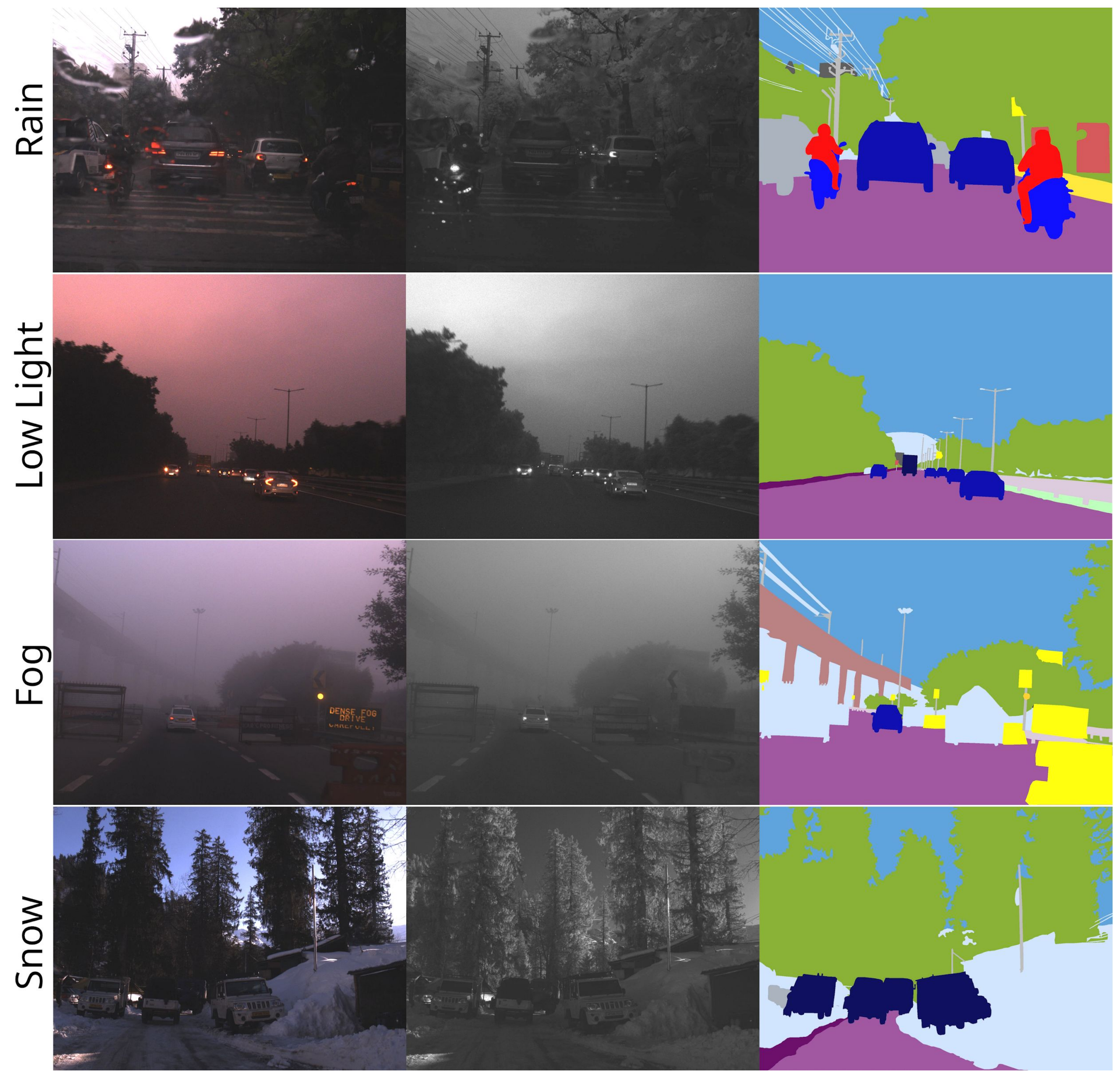
- Safety and Robustness of a segmentation model possess two main challenges:
- The robustness of the model to high amounts of traffic participants and adverse weather conditions.
- Handling corner conditions and image distortions due to high traffic areas and weather changes is necessary.
- Secondly, the metrics used for segmentation should accurately depict the safety of the model.

India Driving Dataset in Adverse Weather

- Hierarchical Adverse weather dataset collected in unstructured traffic scenes all across India
- 5000 RGB-NIR Image pairs
- Semantic / Instance segmentation annotations.
- Pixel level annotations with 30 class labels.
- Rain, Fog, Snow and Lowlight weather conditions.

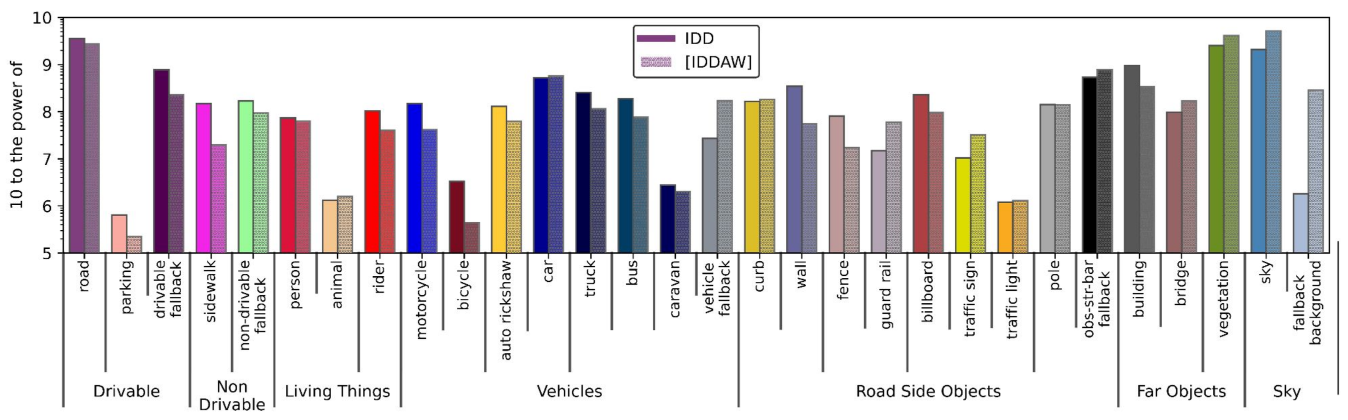
Dataset	Labeled Images	Rain	Fog	LowLight	Snow	Label	NIR
BDD100K	1346	213	23	345	765	19	-
ACDC	4006	1000	1000	1006	1000	19	-
IDD	10000	-	-	-	-	30	-
IDDAW	5000	1500	1500	1000	1000	30	✓

RGB/NIR/Annotation under Adverse Conditions

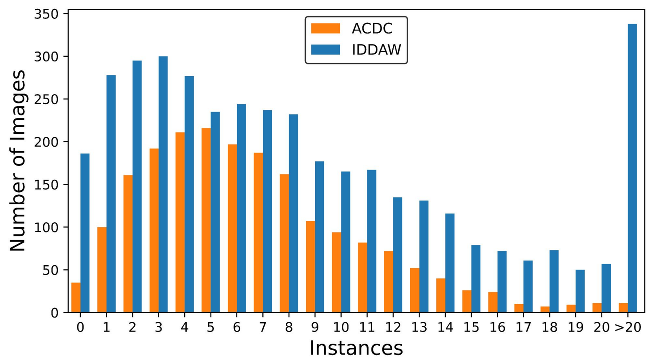


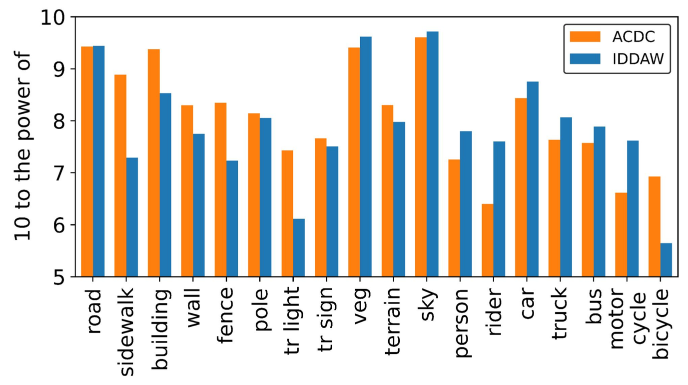
Label Hierarchy and Statistics

Pixel wise comparison for each class between IDD and IDD-AW: Almost identical in spite of our dataset being collected in adverse weather conditions.



Higher pixel wise count per class and instances per image when compared with ACDC.



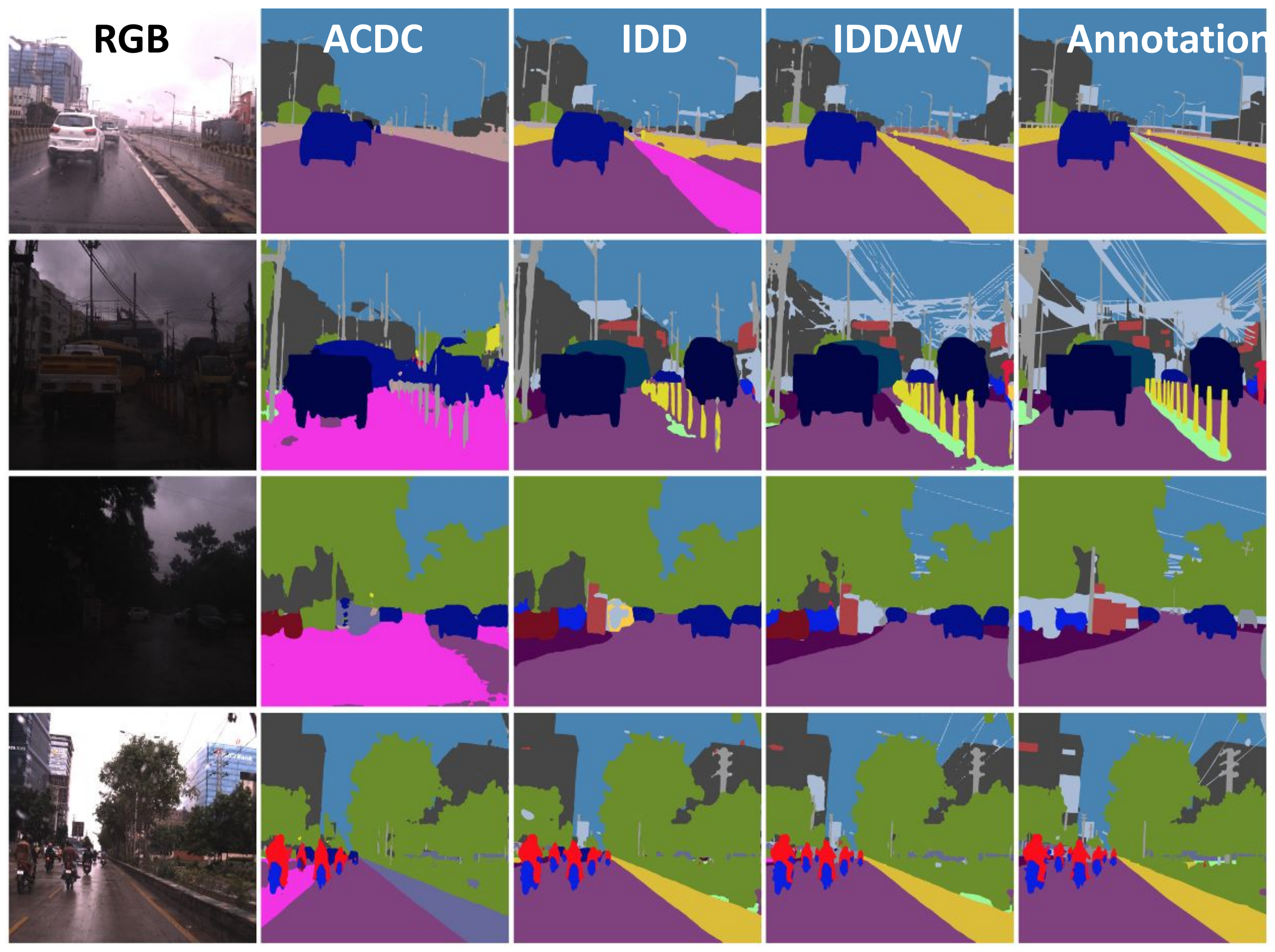


Benchmarking

- Semantic Segmentation using InternImage-b framework.
- Comparison with Cityscapes, IDD and ACDC

Dataset / Test → / Train ↓	CS	ACDC	IDD	Rain	Fog	LL	Snow	IDD-AW
CS RGB	83	-	-	46	45	42	43	46
ACDC RGB	-	75	-	47	51	42	38	48
IDD RGB	-	-	73	52	55	50	33	54
IDD-AW RGB	49	51	51	62	64	62	53	64
IDD-AW NIR	-	-	-	61	58	57	51	61
IDD-AW NIR+RGB	-	-	-	**66**	**65**	**63**	**53**	**67**

Qualitative Comparisons

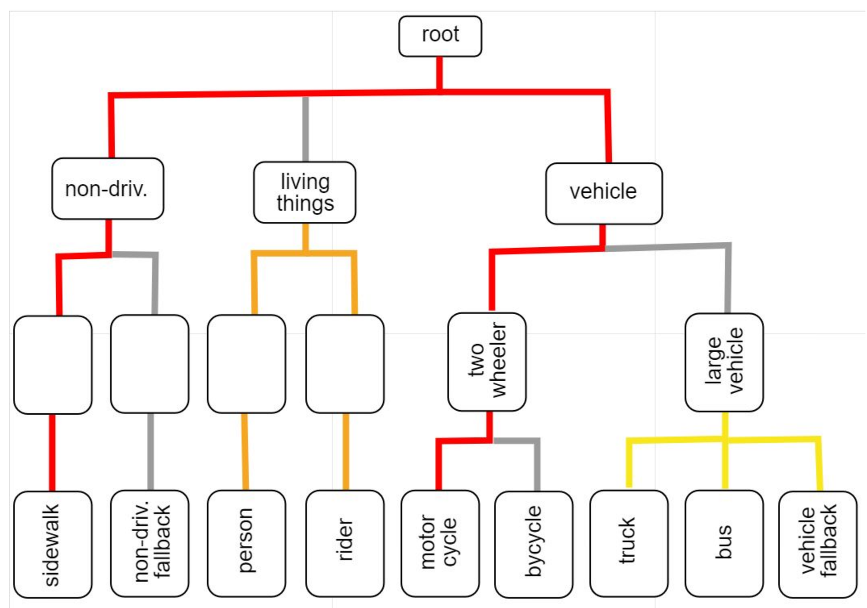


Safety challenges in mIoU metric

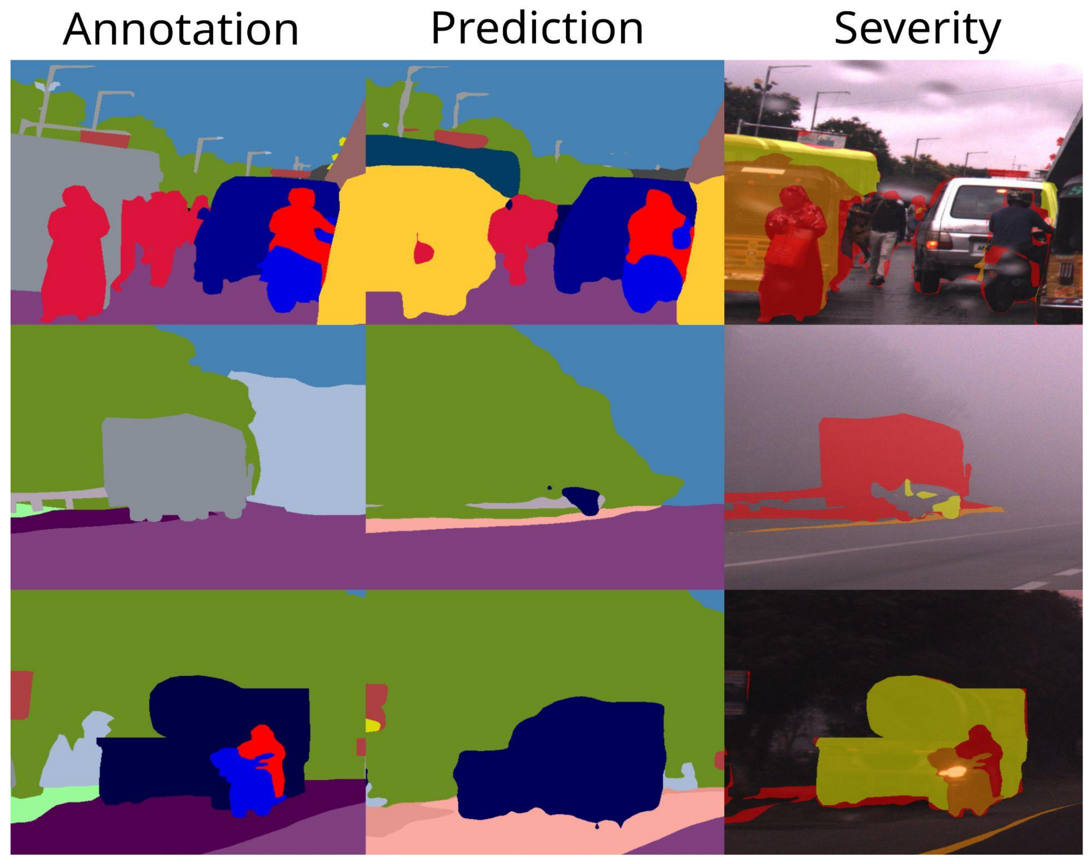
- **Traditional Approach:** mIoU
  - Commonly used for evaluating segmentation quality
  - ***Limitations :***
    - Equal treatment of all classes and predictions regardless of safety significance
    - Inability to capture severity in misclassifications, especially in critical classes like pedestrians, vehicles, and traffic signs

We use tree distance is used to determine the severity of misprediction.

d(sidewalk, motorcycle) = 3 > d(person, rider) = 2 > d(truck, bus) = 1



Severity map based on tree distance:



Severity: Red, orange, yellow indicate danger level. In the third row, bike as a truck (orange) is less dangerous than rider as a vehicle (red). mIoU treats all mispredictions equally, while proposed safe mIoU penalizes orange and red more. Proposed safe mIoU is over 20% lower than mIoU in these cases.

Safe mIoU (SmIoU):

- Proposed metric SmIoU takes tree distance to overcome the limitations of traditional mIoU.
- The penalty for misclassification is based on the tree distance between expected and predicted classes and the severity levels is expressed as yellow, orange and red colors based on the distance

$$I^{\text{safe}}_{c,s} = \frac{|\text{gt}_c \cap \text{pred}_s|}{|\text{gt}_c \cup \text{pred}_c|} \qquad I_{c,c} = \frac{|\text{gt}_c \cap \text{pred}_c|}{|\text{gt}_c \cup \text{pred}_c|}$$

$$I^{\text{safe}}_{c} = \begin{cases} I_{c,c} - \sum\limits_{s \in C, s \neq c} \frac{d(c,s)}{n} I^{\text{safe}}_{c,s} & \text{if } c \in C_{\text{imp}} \\ I_{c,c} - \sum\limits_{s \in C_{\text{imp}}} \frac{d(c,s)}{n} I^{\text{safe}}_{c,s} & \text{else.} \end{cases}$$

$$\text{SmIoU} = \frac{\sum_{c \in C} I^{\text{safe}}_c}{|C|}$$

SmIoU vs mIoU

InternImage-b experts evaluated on IDD-AW: first four columns for weather conditions, last three columns compare mIoU and SmIoU across diverse adverse conditions.

eval →	cross				same		
Test → / Train ↓	Rain	Fog	LL	Snow	mIoU	SmIoU (tp)	SmIoU
IDD	52	55	50	33	-	-	-
IDD-AW	-	-	-	-	64	60	51
Rain	-	55	40	29	64	58	48
Fog	51	-	53	29	64	58	47
LL	52	57	-	30	62	58	48
Snow	35	38	33	-	53	43	28

Histogram shows Safe mIoU (%) vs mIoU (%) distribution for IDD-AW test set. Shift in SmIoU to lower values indicates detection of significant, unaccounted mispredictions compared to mIoU.