IDDAW: A Benchmark for Safe and Robust Segmentation of Drive Scenes in Unstructured Traffic and Adverse Weather

iddaw.github.io

Furqan Ahmed Shaik,
Abhishek Reddy Malreddy, Nikhil Reddy Billa,
Kunal Chaudhary, Sunny Manchanda, Girish Varma

Machine Learning Lab
Center for Security, Theory & Algorithmic Research
IIIT Hyderabad

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

Safety and Robustness in autonomous navigation

Report: Nearly 400 crashes by 'self-driving' cars in the US

Data collected by a US regulatory agency will allow for greater transparency on safety of semiautonomous vehicles

A Waymo minivan drives passengers during an autonomous vehicle ride in Chandler, Arizona [File: Ross D Franklin/AP]

https://www.aljazeera.com/economy/2022/6/15/report-nearly-400-crashes-by-self-driving-cars-in-the-us

Tesla's self-driving technology fails to detect children in the road, group claims

Safe technology campaigners release 'disturbing' video advert showing car in Full Self-Driving mode hitting child-sized mannequin

https://www.theguardian.com/technology/2022/aug/09/tesla-self-driving-technology-safety-children

California hits pause on GM Cruise self-driving cars due to safety concerns

Decision is latest instance of regulatory agencies expressing concern over safety of autonomous vehicles.

https://www.aljazeera.com/news/2023/10/24/california-hits-pause-on-gm-cruise-self-driving-cars-due-to-safety-concerns

Cruise recalls all self-driving cars after grisly accident and California ban

All 950 of the General Motors subsidiary's autonomous cars will be taken off roads for a software update

Cruise self-driving cars outside the company's headquarters in San Francisco. Photograph: Heather Somerville/Reuters

https://www.theguardian.com/technology/2023/nov/08/cruise-recall-self-driving-cars-gm

Problem: We need to create benchmarks for robust datasets as well as design appropriate metrics for calculating safety.

Adverse weather driving challenges in India:

Amid heavy rain in Delhi, auto driver, woman killed, 5 others injured

Two people were killed in rain-related incidents in Delhi on Sunday after record-breaking rain in the national capital. Five other people were injured in separate incidents.

https://www.indiatoday.in/cities/delhi/story/delhi-heavy-rain-death-injured-accident-road-weather-flood-2404180-2023-07-10

News

Fog envelops north India, low visibility leads to accident in UP

Delhi saw cold wave conditions for the fifth consecutive day and fog reduced visibility to just 25 metres.

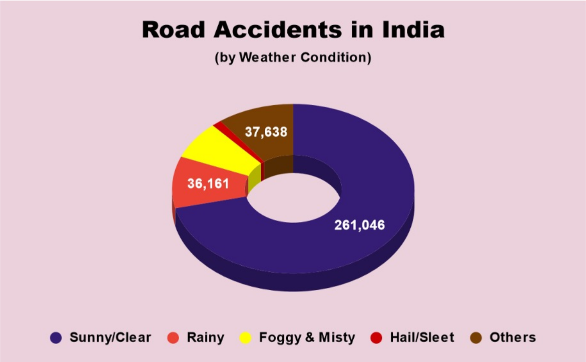
News / India News / Cold wave grips north India; 3 killed, 40 injured in fog-related ...

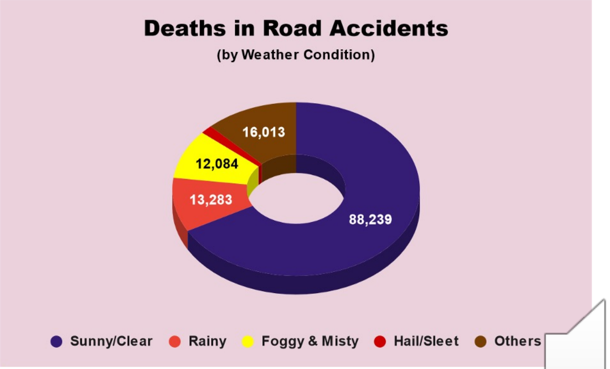
Cold wave grips north India; 3 killed, 40 injured in fog-related accidents

PTI |

Dec 20, 2022 10:51 PM IST Join Us

https://www.hindustantimes.com/india-news/cold-wave-grips-north-india-3-killed-40-injured-in-fog-related-accidents-101671556602987.html

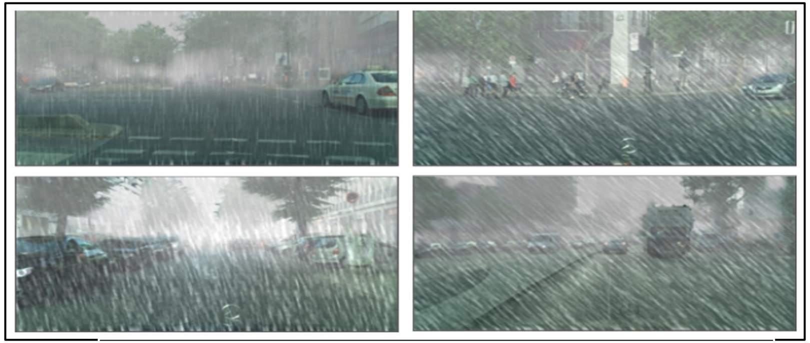




https://www.news18.com/news/india/crash-course-most-accident-deaths-on-straight-roads-under-clear-weather-in-2020-shows-govt-data-5562337.html

Challenges with Existing Adverse Weather Datasets

- Synthetic Images because real time data is harder to capture.
- No unstructured traffic.
- Skewed datasets.
- Insufficient sensors and data to accurately capture the effects of adverse weather conditions.



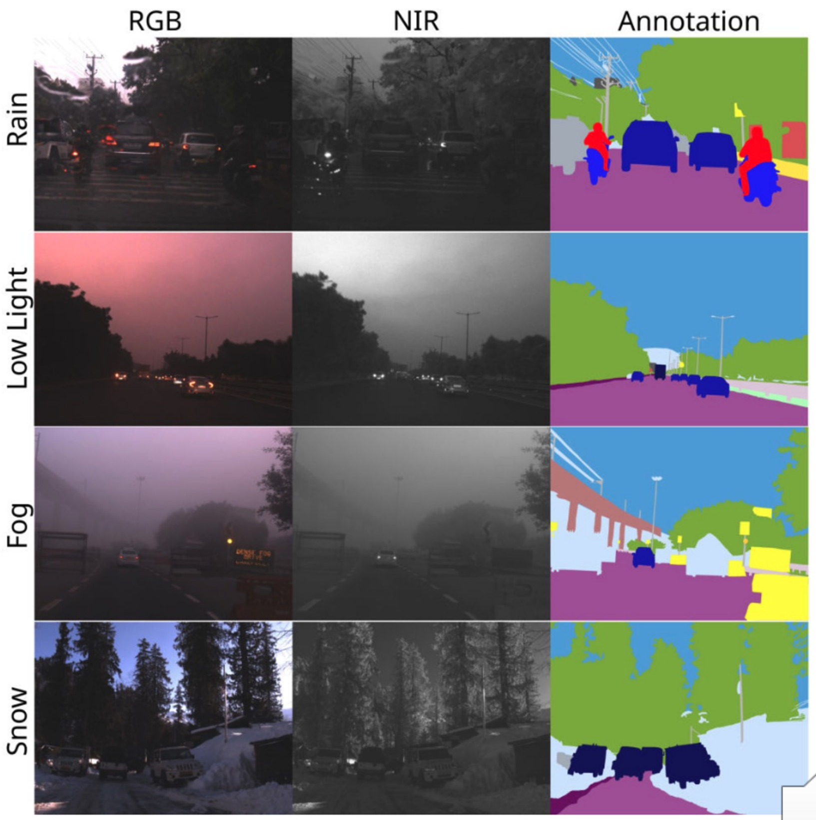
Synthetic Rain Sample Images

ACDC Sample Images

IDD-AW DATASET:

- Benchmark dataset for driving scenes In adverse weather and unstructured traffic.
- Rain, Fog, Lowlight and Snow.

- Collected across various states of India, from highways of Hyderabad and Delhi, to foggy hills of Ooty and Snowy mountains and roads of Manali.



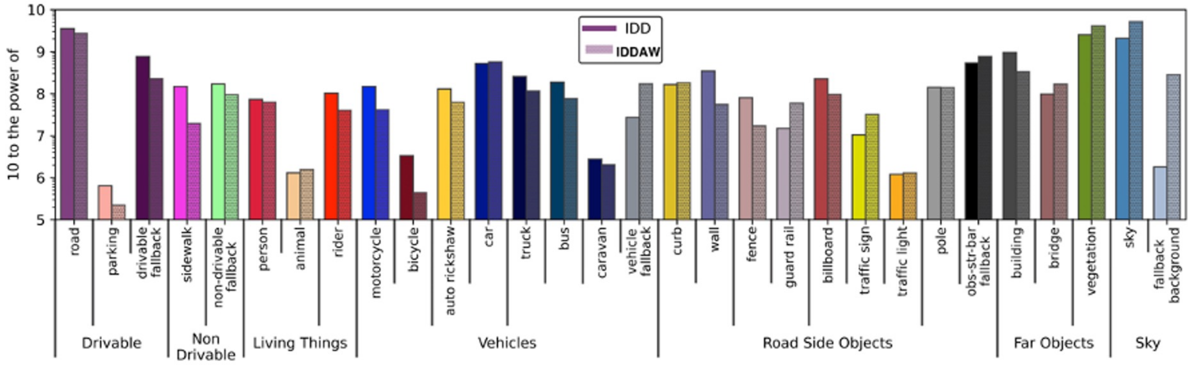
Comparison with other SOTA Datasets

- 202 drive sequences across various Weather conditions.
- 5000 RGB-NIR Image pairs
- Higher label set indicating more Diversity and types of instances When compared to other datasets
- Semantic/Pixel wise annotations.

Dataset	Labeled images	Rain	Fog	Snow	Lowlight	Labels	NIR
Foggy Driving	101	0	101	0	0	19	
Foggy Zurich	40	0	40	0	0	19	
Nighttime Driving	50	0	0	0	50	19	
Dark Zurich	201	0	0	0	201	19	
Raincouver	326	326	0	0	95	19	
WildDash	226	13	10	26	13	19	
BDD100K	1346	213	23	345	765	19	
ACDC	4006	1000	1000	1000	1006	19	
IDD-AW	**5000**	**1500**	**1500**	**1000**	**1000**	**30**	✓

IDD-AW Statistics

- It has almost identical pixelwise Comparison for each class even Though collected in adverse weather

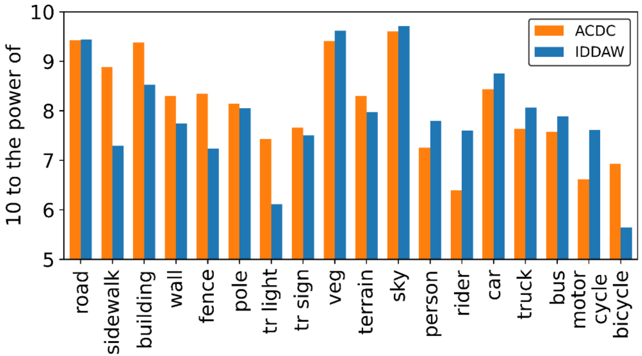


Uses same label hierarchy as of IDD Dataset

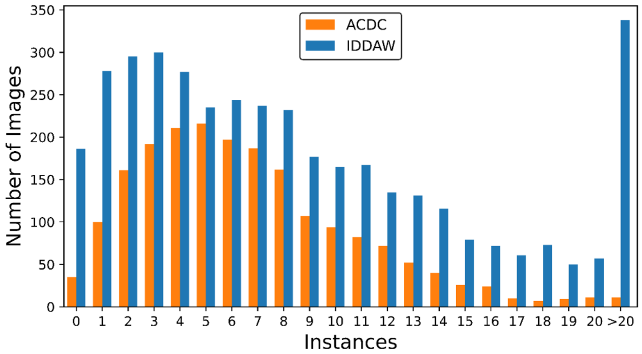
Girish Varma, Anbumani Subramanian, Anoop Namboodiri, Manmohan Chandraker, and CV Jawahar. IDD: A dataset for exploring problems of autonomous navigation in uncon strained environments. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1743–1751. IEEE, 2019.

Comparison with ACDC

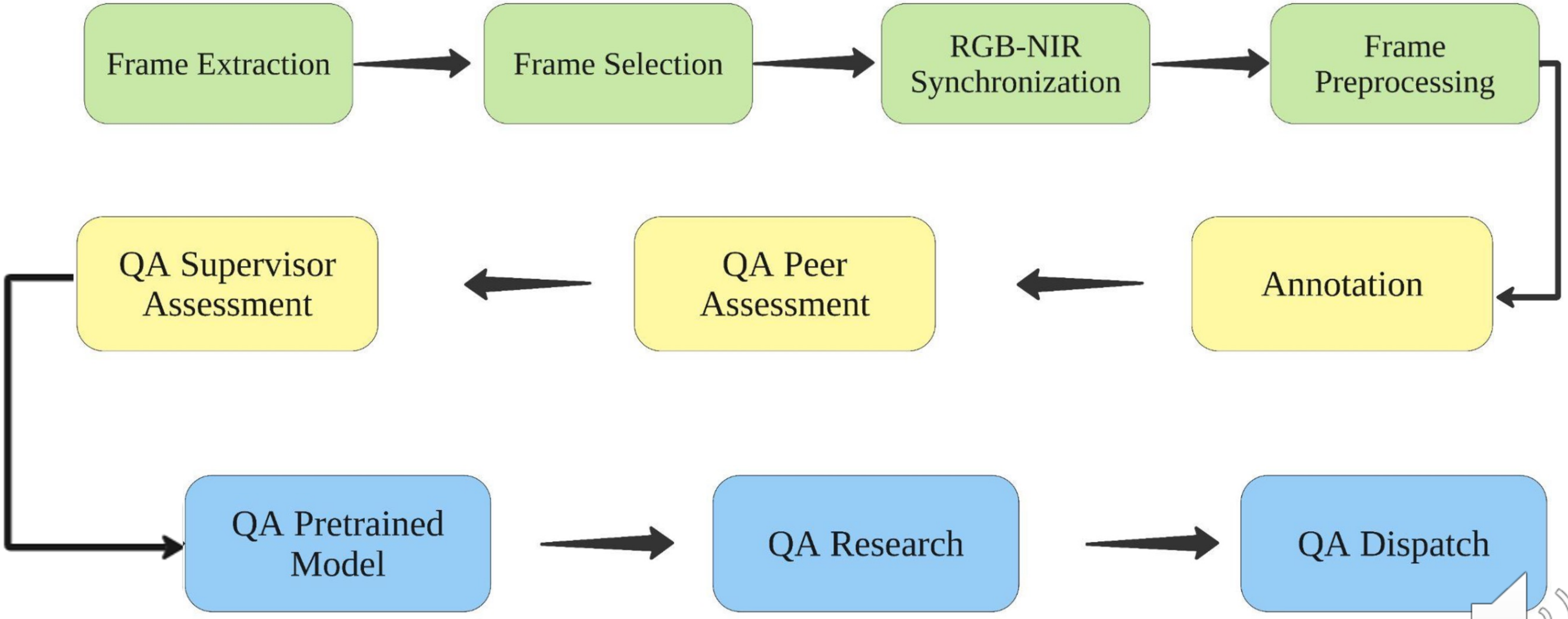
- Higher number of pixel counts per class when compared with ACDC.



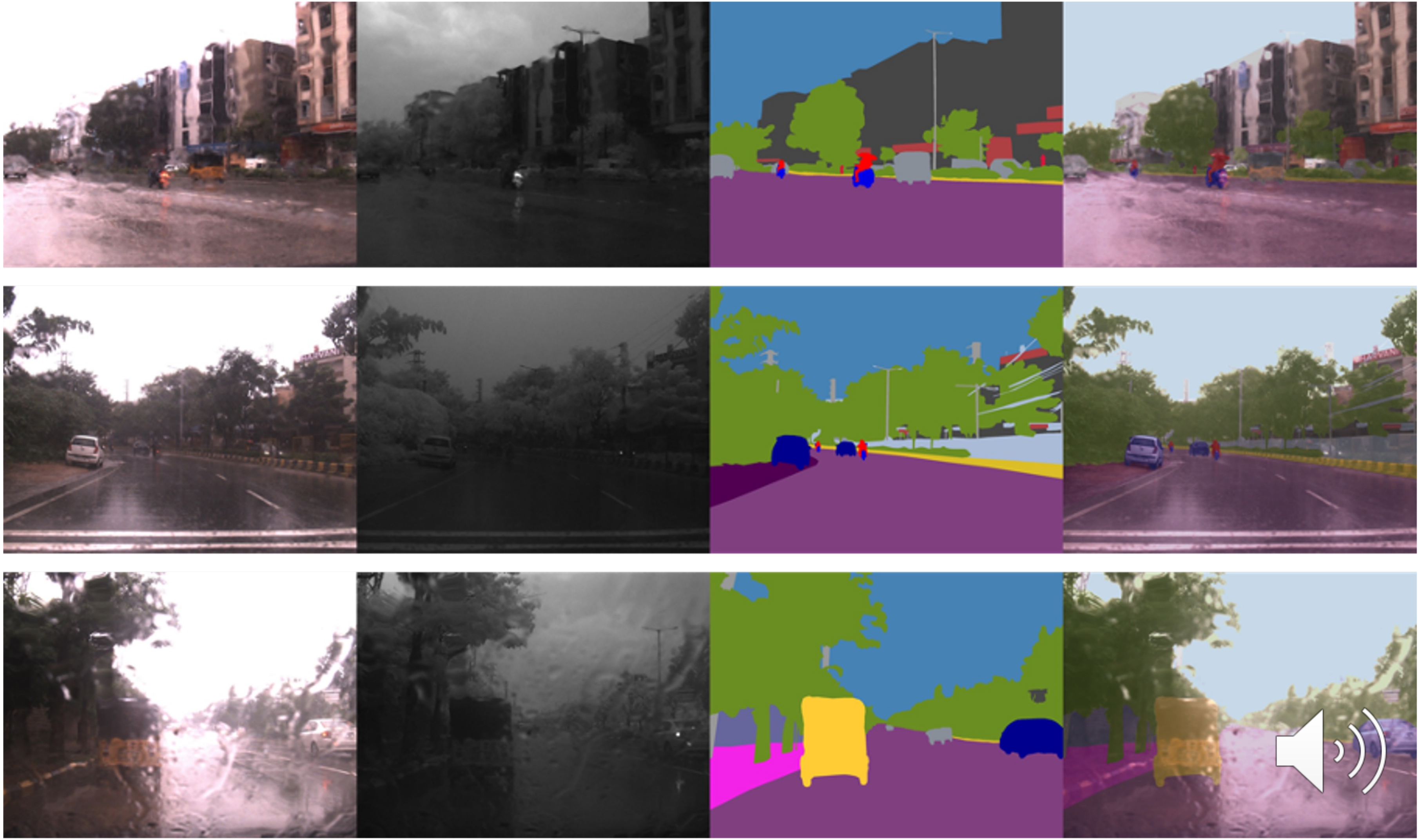
- Number of instances and traffic participants per image is significantly higher in comparison with ACDC.



Annotation Pipeline



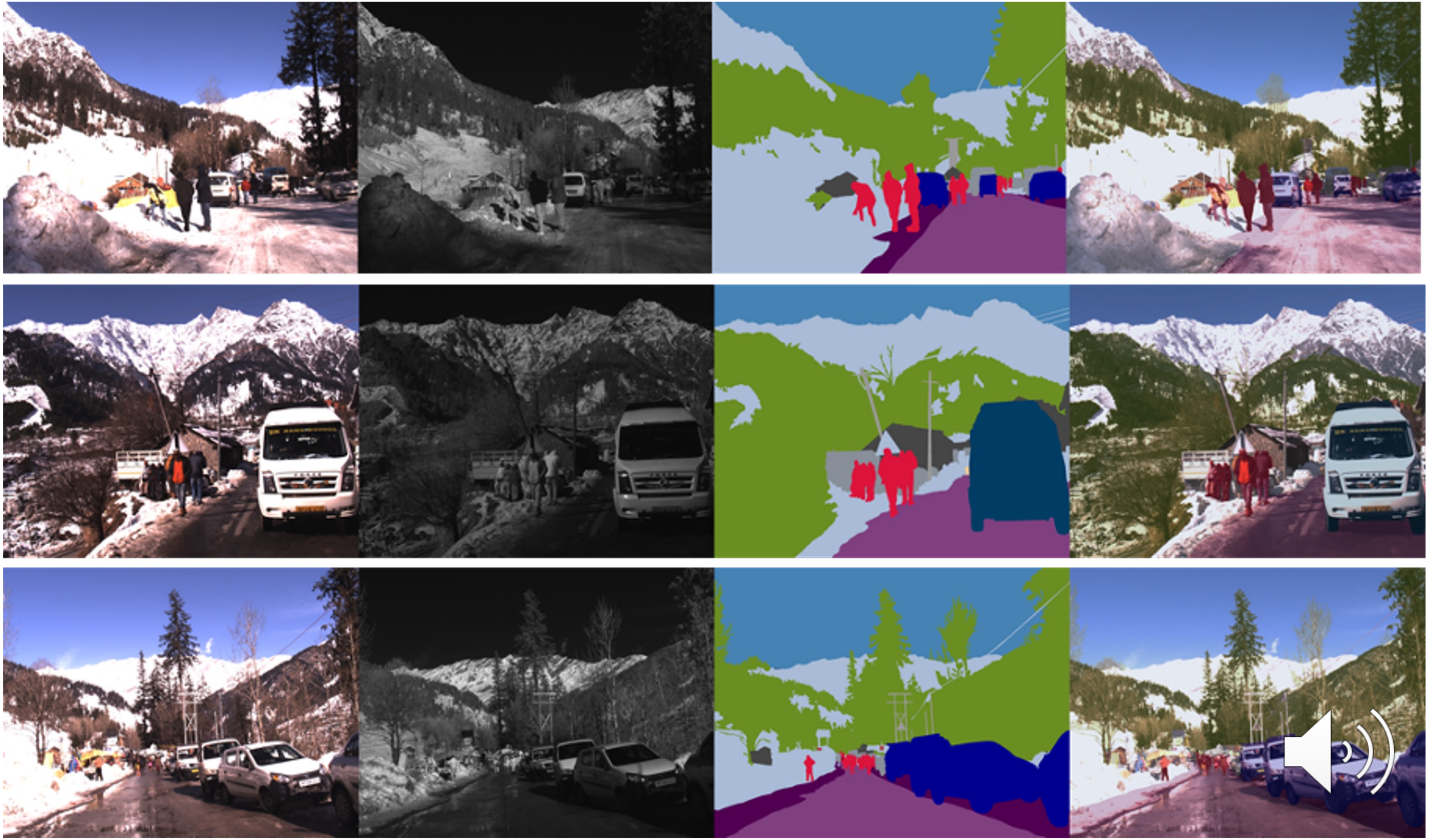
Rain Images



FOG

Lowlight

Snow



Sample Data Collection videos:

Rain data collection

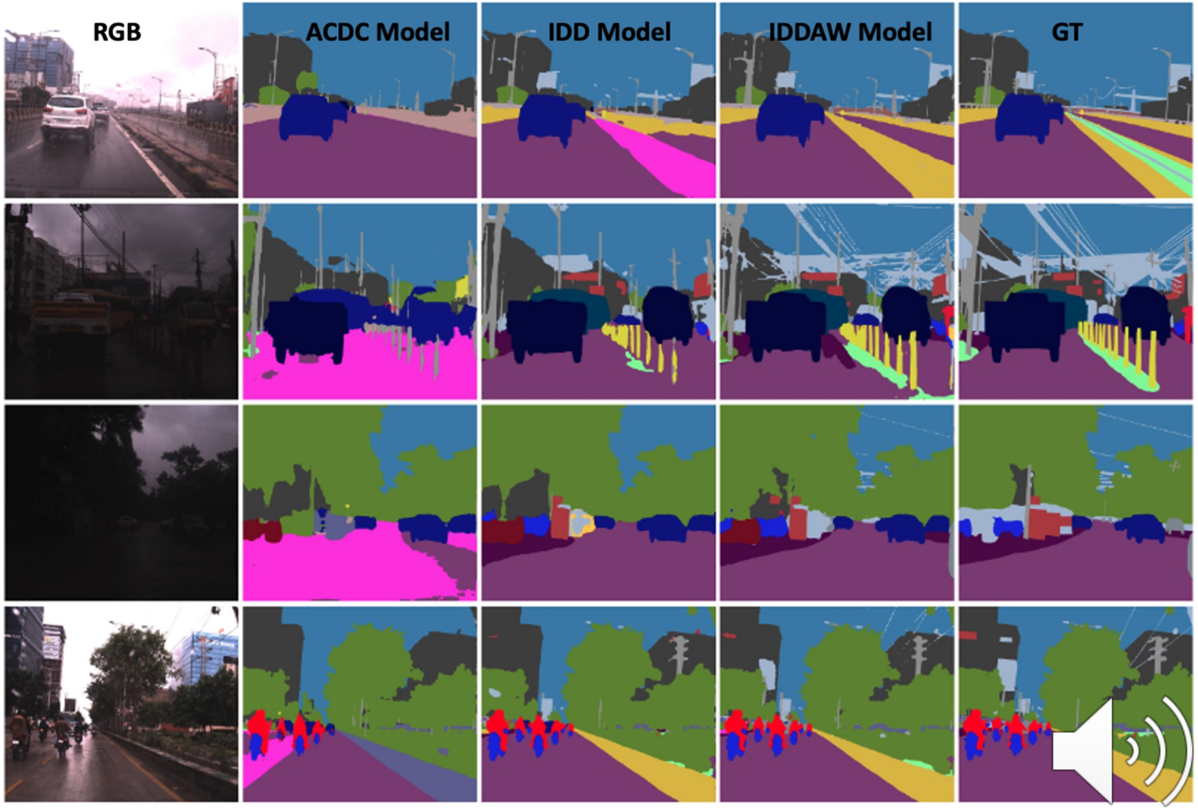
Fog / Lowlight

Snow

Qualitative Results

- Comparisons across Cityscapes, ACDC, IDD and IDD-AW

- We use InternImage-b framework
  As the standard model for all the Datasets.



Results: Semantic Segmentation (mIoU)

- IDD-AW pretrained models outperform others with a notable 10-15% difference in mIoU compared to CS, ACDC, and IDD pretrained models.
- IDD-AW pretrained models demonstrate superior performance on datasets beyond IDD-AW, surpassing the performance of other pretrained models on the IDD-AW test set.

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**

Result: This shows better inclusive data in terms of both adverse weather as well as unstructured traffic scenes.

Metrics

Semantic Segmentation Metrics

- **Traditional Approach:** mIoU (mean of Intersection over Union)
    - Commonly used for evaluating segmentation quality

- **Challenge:** Limitations in assessing safety in driving scenes
    - ***Limitations :***
        - Equal treatment of all classes regardless of safety significance
        - Inability to capture severity in misclassifications, especially in critical classes like pedestrians, vehicles, and traffic signs
        - Failure to distinguish between tolerable and dangerous misclassifications in real-world driving scenarios

Why Safe mIoU

- **Objective**:
  - Introduce a refined metric - Safe mIoU (SmIoU)
  - Overcome the shortcomings of mIoU in the context of driving scene safety
- **Significance:**
  - Safety in driving scenes crucial for real-world applications
  - Need for a metric that considers the severity of misclassifications

Hierarchical Labelling and Tree distance

Hierarchical labeling :

- refers to the structured organization of classes into a tree-like hierarchy, capturing semantic relationships and dependencies between them in the context of semantic segmentation

- **Advantages:**
  - Inherent semantic relationships between classes.
  - Provides a structured hierarchy for a comprehensive understanding of class dependencies.

Hierarchical Labelling and Tree distance

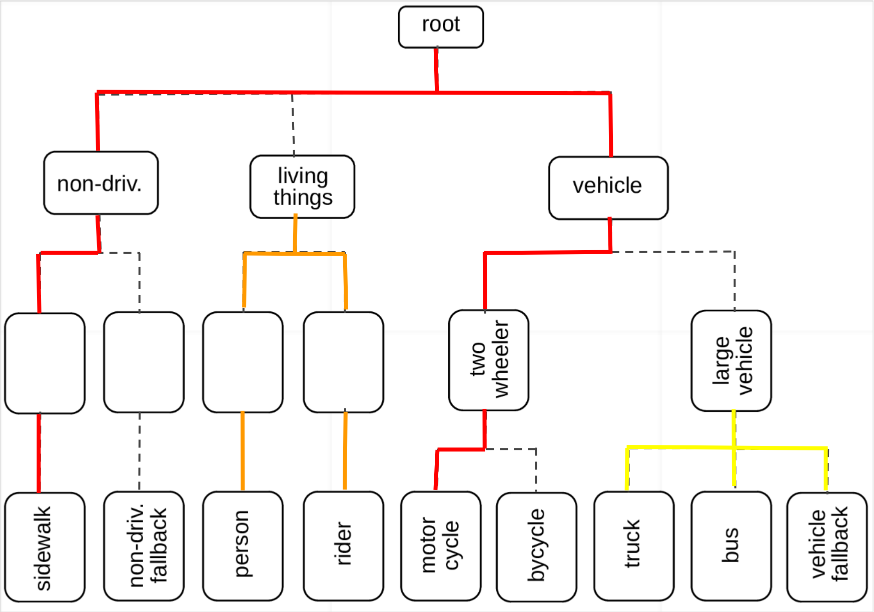
Tree distance:

- Length of the shortest path in the class hierarchy tree.
- To ensure its appropriate scaling, it is divided by 2.

- Purpose of Tree Distance:
  - Quantifies the semantic relationship between classes.
  - Guides the severity of penalties for misclassifications.

Tree distance: Danger levels of mispredictions

- Yellow indicates immediate siblings and hence tolerable misclassification
- Orange suggests same L2 category but has some implications.
- Red suggests highest distance and hence highest penalty.



Safe mIoU: Calculation

- **I safe_c :** Calculation of Safe IoUs for each critical class (c) incorporating penalties based on tree distance.(eq1)

- **Final SmIoU Score :** Obtained by taking the mean of I safe_c.(eq2)

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

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

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

Results

- Cross - Evaluation results presented for different weather condition experts, comparing mIoU and SmIoU metrics.

- InternImage-b framework experiences a consistent drop of over 15% in SmIoU for each individual condition when considering important classes.

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

Comparison of mIoU (%) with SmIoU (%) metric at different levels and label sets for various adverse weather conditions

Results

Implication:

- Table highlights the discrepancy and dangers of relying solely on traditional mIoU in hierarchical autonomous driving datasets.

- Results demonstrates the critical role of SmIoU in revealing safety concerns and better quantifying the performance in safety-critical evaluations.

SmIoU vs mIoU Comparison

Safety Concerns:

- Stressing the potential dangers of misclassifying these crucial classes, especially in driving scenarios.

Metric	Condition	road	drivable fallback	sidewalk	person	rider	bike	bicycle	rickshaw	car	truck	bus	vehicle fallback	curb	wall	guard rail	billboard	traffic sign	traffic light
mIoU %	All	95	51	48	76	72	68	5	83	85	74	76	45	78	52	74	62	58	52
	Rain	96	49	49	48	73	68	2	86	87	79	52	47	81	47	46	64	54	52
	Fog	97	64	24	62	75	73	8	67	91	82	80	49	79	57	78	61	69	47
	LL	95	60	54	73	73	68	38	75	80	69	86	29	65	50	72	58	35	53
	Snow	85	42	-	80	62	40	0	-	82	58	70	48	23	56	64	60	37	-
SmIoU %	All	92	32	16	64	58	52	- 22	77	81	68	70	21	69	32	61	42	40	27
	Rain	94	28	25	15	59	54	-7	81	84	74	46	23	74	22	20	46	32	26
	Fog	95	51	-48	41	62	58	-63	43	87	78	77	28	71	52	69	45	53	12
	LL	92	46	22	60	59	52	14	70	76	61	82	2	48	26	52	35	15	32
	Snow	79	20	-	70	44	-2	-99	-	76	48	56	23	-46	23	42	38	5	-

Table 4. Comparison of class-wise labels for important classes between mIoU vs SmIoU for InternImage-b model on segmentation.

- **Significant Disparities:**
  - Highlighted in classes like bicycles, traffic signals, and sidewalks.

SmIoU vs mIoU Comparison

Negative SmIoU:

Bicycle, Sidewalk, Curb:

- Indicate dangerous misclassifications overlooked by traditional mIoU.
- **Negative values highlight the effect of SmIoU and dangerous mispredictions.**

Metric	Condition	road	drivable fallback	sidewalk	person	rider	bike	bicycle	rickshaw	car	truck	bus	vehicle fallback	curb	wall	guard rail	billboard	traffic sign	traffic light
mIoU %	All	95	51	48	76	72	68	5	83	85	74	76	45	78	52	74	62	58	52
	Rain	96	49	49	48	73	68	2	86	87	79	52	47	81	47	46	64	54	52
	Fog	97	64	24	62	75	73	8	67	91	82	80	49	79	57	78	61	69	47
	LL	95	60	54	73	73	68	38	75	80	69	86	29	65	50	72	58	35	53
	Snow	85	42	-	80	62	40	0	-	82	58	70	48	23	56	64	60	37	-
SmIoU %	All	92	32	16	64	58	52	- 22	77	81	68	70	21	69	32	61	42	40	27
	Rain	94	28	25	15	59	54	-7	81	84	74	46	23	74	22	20	46	32	26
	Fog	95	51	-48	41	62	58	-63	43	87	78	77	28	71	52	69	45	53	12
	LL	92	46	22	60	59	52	14	70	76	61	82	2	48	26	52	35	15	32
	Snow	79	20	-	70	44	-2	-99	-	76	48	56	23	-46	23	42	38	5	-

Table 4. Comparison of class-wise labels for important classes between mIoU vs SmIoU for InternImage-b model on segmentation.

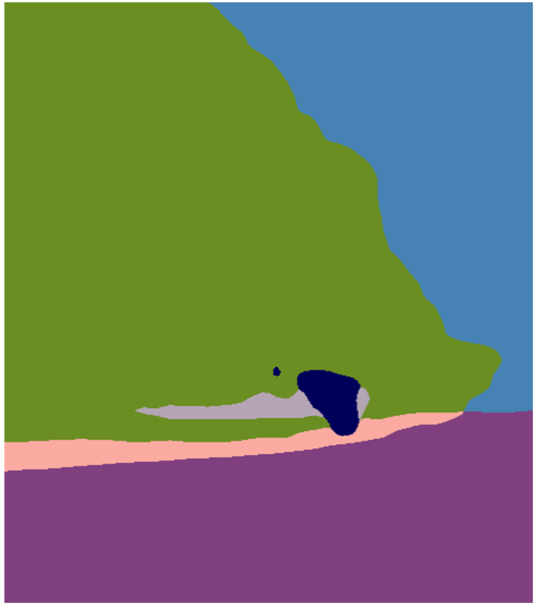
Qualitative examples with severity map

Ground Truth   Prediction   Severity

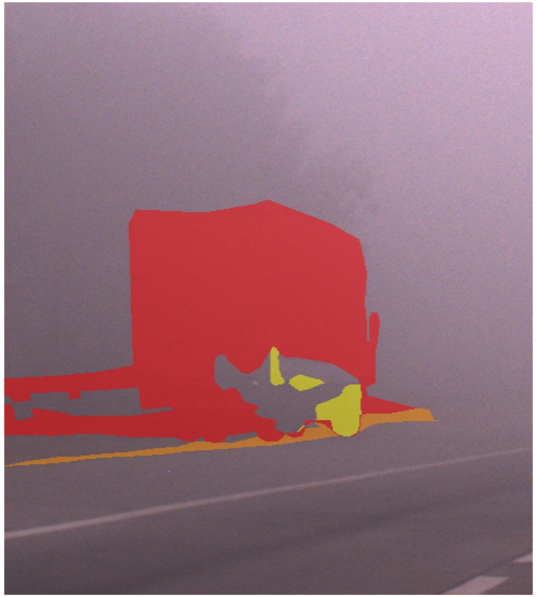
A pedestrian is misclassified in the above example

Qualitative examples with severity map

Ground Truth



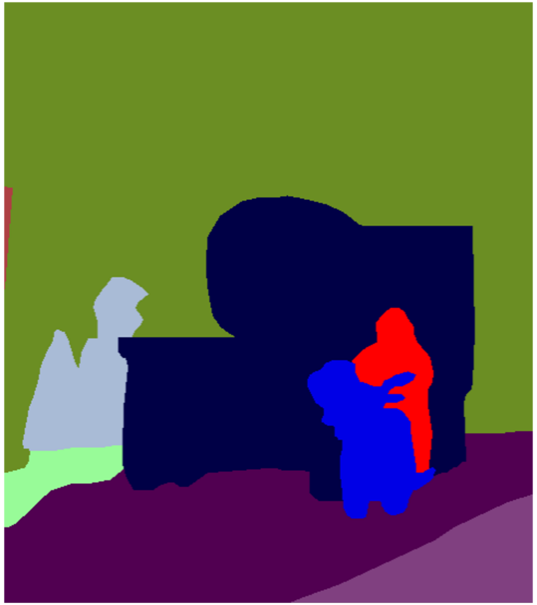
Prediction



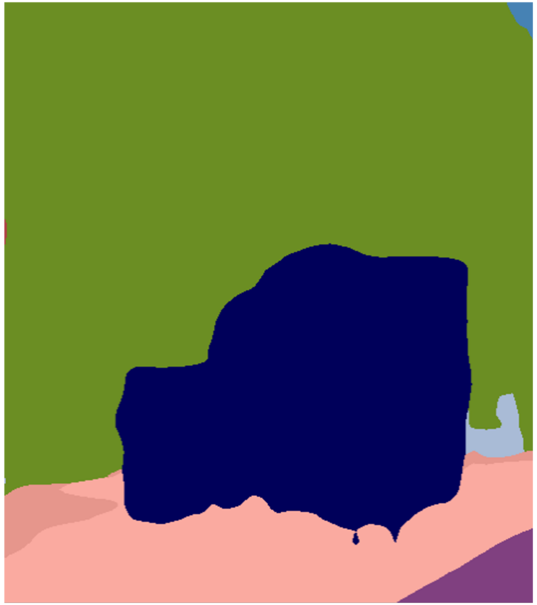
Severity

A truck is misclassified in the above example

Qualitative examples with severity map



Ground Truth

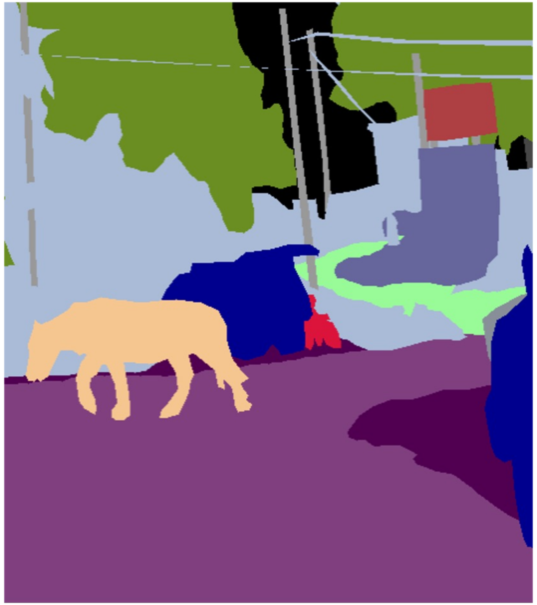


Prediction

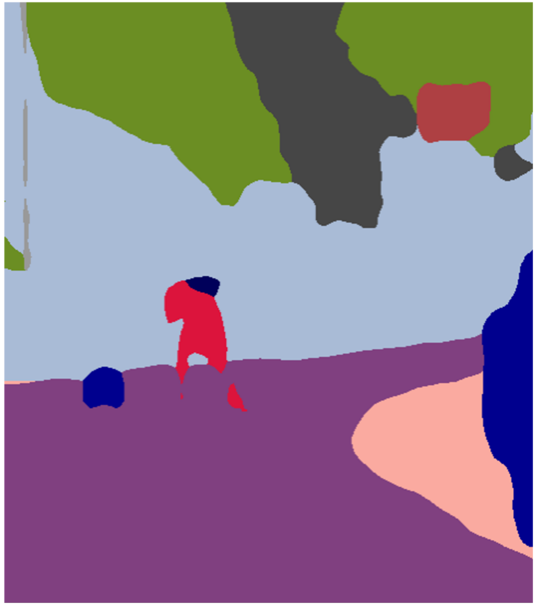
Severity

A rider is misclassified in the above example

Qualitative examples with severity map



Ground Truth



Prediction

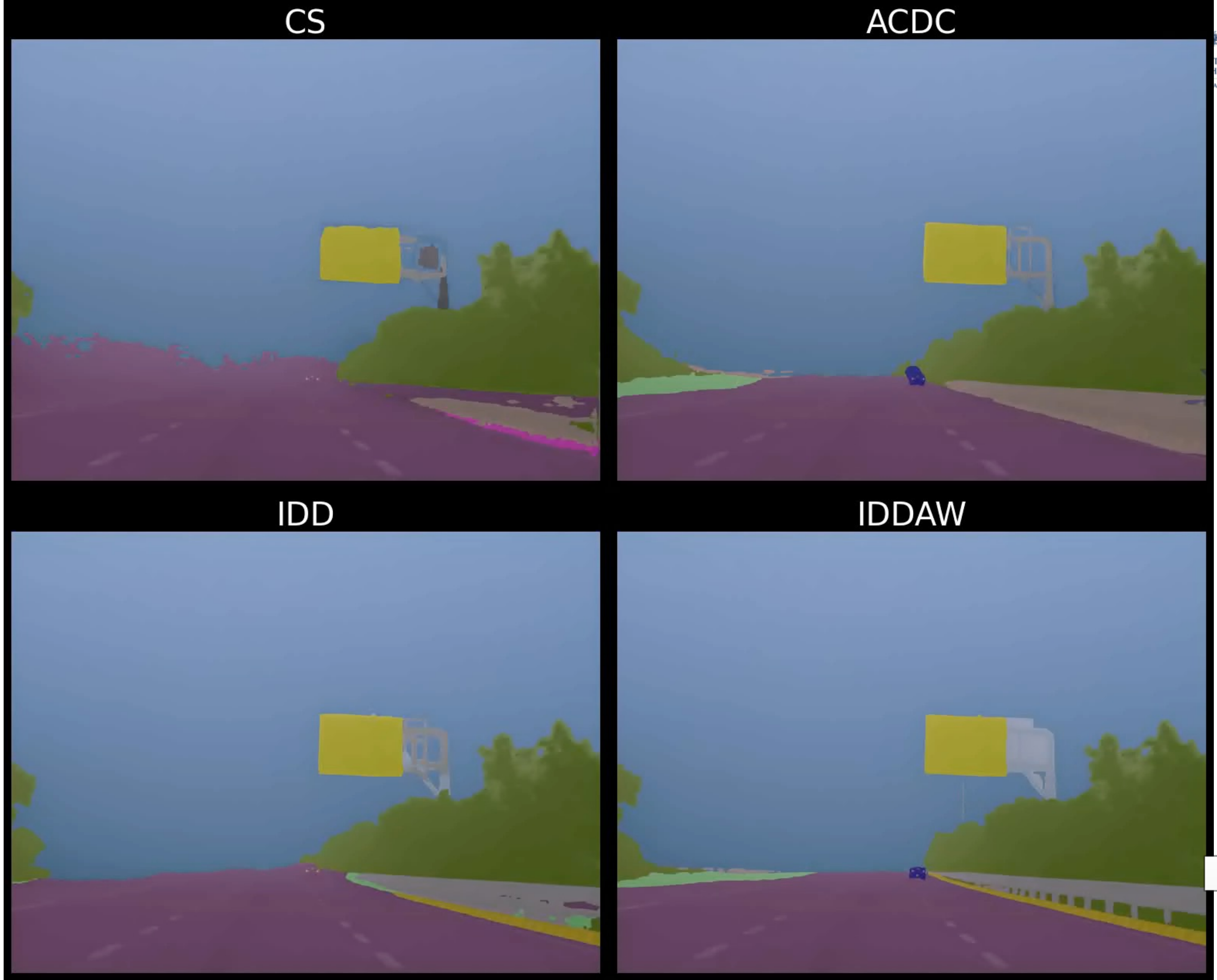
Severity

An animal is misclassified in the above example

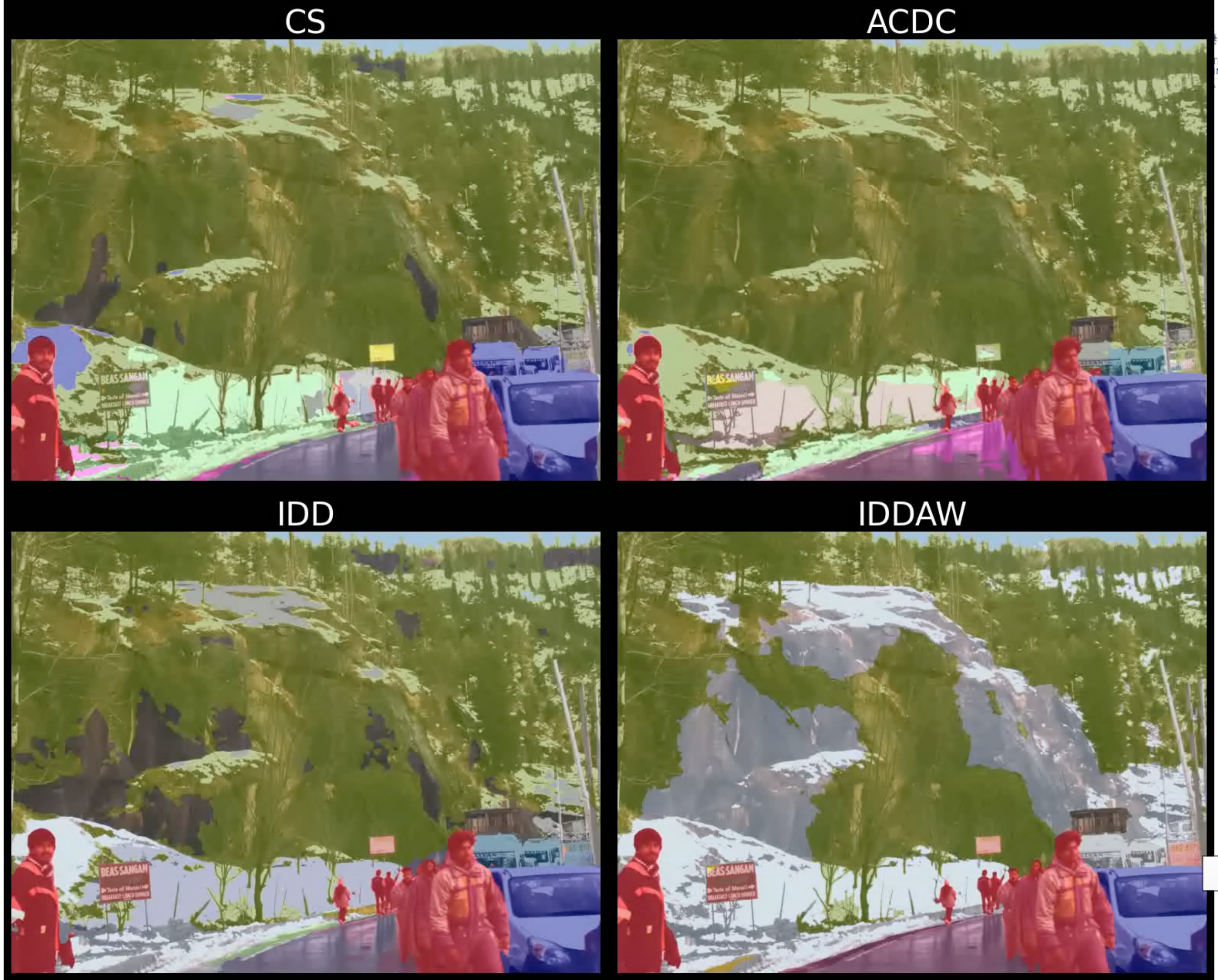
Comparison of models on rain:

Comparison of models on fog:

Lowlight:



Comparison of models on Snow:



Conclusion:

- We have presented IDD-AW, a large-scale dataset and a benchmark suite for semantic driving scene understanding in adverse weather and unstructured driving conditions.
- We also present a new metric called Safe mIoU which incorporates safety concerns in the definition of mIoU.
- We benchmark state-of-the-art models for semantic segmentation in IDD-AW and also show the differences between traditional mIoU and safe mIoU while considering important classes.
- Finding appropriate loss functions, which can better optimize safe mIoU more efficiently is an interesting direction for future work.

Thanks!

Questions?

Project: http://iddaw.github.io

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

IIIT Hyderabad

CSTAR