MUAD: Multiple Uncertainties for Autonomous Driving, a benchmark for multiple uncertainty types and tasks

Gianni Franchi[1 *], Xuanlong Yu[1,2 *], Andrei Bursuc[3], Angel Tena[4], Rémi Kazmierczak[1], Severine Dubuisson[5], Emanuel Aldea[2], David Filliat[1]

[1]U2IS, ENSTA Paris, IPP [2]SATIE, Paris-Saclay University [3]valeo.ai [4]Anyverse [5]Aix Marseille University

TLDR: We provide a new synthetic dataset with multiple uncertainty sources for autonomous driving in order to better simulate and evaluate aleatoric uncertainty (domain-shift caused by unknow weather conditions) and epistemic uncertainty (Out-of-Distribution samples).

1/ Motivation and contributions

Motivation:

(1). Disentangling the different types and sources of uncertainty is non trivial for most datasets, especially since there is no ground truth for uncertainty

(2). Adverse weather conditions of varying intensities are usually under-represented in both training and test sets in public datasets.

Contributions:

(1). MUAD: a new automotive dataset with annotations for multiple tasks and multiple uncertainty sources.

(2). A wide range of benchmarks on MUAD dataset for multiple computer vision tasks and settings to further support research in this area.

(3). An extensive study on uncertainty quantification for 2D output tasks for recent Transformer-based architectures.

Download Dataset:

Scan the QR code on the right and fill the form, then download link will be permanently provided. (Test sets will be fully provided after the challenge.)

MUAD Challenge:

MUAD Uncertainty Estimation for Semantic Segmentation Challenge is held on CodaLab, feel free to participate.

2/ Dataset Overview:

10413 annotated images:
3420 images in the train set, 492 in the validation set and 6501 in the test set. 2/3 being day images and 1/3 night images

3 types of adversity conditions with 2 intensity levels:
Fog, Rain, Snow

1024×2048 high resolution images

21 classes:
19 ID classes (same as Cityscapes), 2 OOD classes (object anomalies and animals)

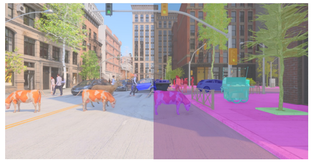
4 supported tasks:
Semantic segmentation
Depth estimation
Object detection 2D/3D
Instance segmentation

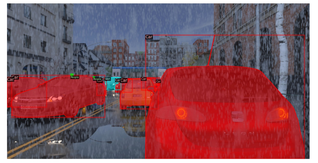
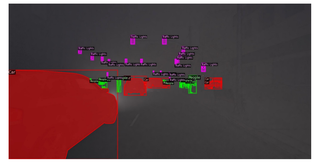
7 test sets:
Normal sets
Normal set overhead sun
OOD set
Low adv. Set
High adv. Set
Low adv. with OOD set
High adv. with OOD set

Photorealistic rendering
Sky: a physical model of the light coming from the, the factors includes ozone, humidity etc.

Rain and snow: the falling speed and size of raindrops are chosen according to observed real rain and snow.

Fog: a full volumetric approach for the simulation where scattering effects are considered.

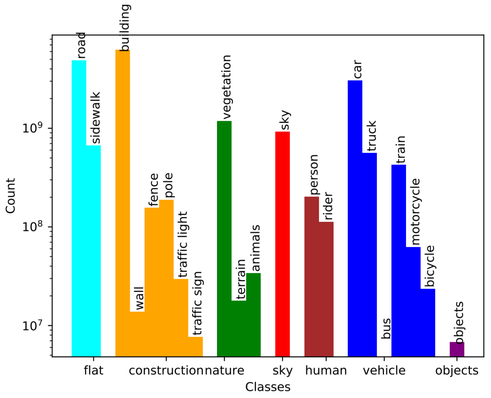


Overview of the different datasets for uncertainty on autonomous driving:

Dataset	Adversarial annotations	Fog	Night	Rain	Snow	Classes	Out of distribution	Depth	Object detection 2D/3D	Instance segmentation
Foggy Driving [🇨🇭]	101	✓	-	-	-	19	-	-	✓	-
Foggy Zurich [🇨🇭]	40	✓	-	-	-	19	-	-	-	-
Nighttime Driving [🇨🇭]	50	-	✓	-	-	19	-	-	-	-
Dark Zurich [🇨🇭]	201	-	✓	-	-	19	-	-	-	-
Raincouver [🇨🇭]	326	-	✓	✓	-	3	-	-	-	-
WildDash [🇨🇭]	226	✓	✓	✓	✓	19	-	-	-	-
BDD100K [🇨🇭]	1346	✓	✓	✓	✓	19	-	-	-	-
ACDC [🇨🇭]	4006	✓	✓	✓	✓	19	-	✓	✓	-
Virtual KITTI 2 [🇨🇭]	21260	✓	-	✓	-	14	-	✓	✓	✓
Fishyscapes [🇨🇭]	373	-	-	-	-	19+2	✓	-	-	-
LostAndFound [🇨🇭]	1203	-	-	-	-	19+9	✓	-	-	-
RoadObstacle21 [🇨🇭]	327	-	✓	-	✓	19+1	✓	-	-	-
RoadAnomaly21 [🇨🇭]	100	-	-	-	✓	19+1	✓	-	-	-
Streethazard [🇨🇭]	6625	-	-	-	-	13+250	✓	-	-	-
BDD anomaly [🇨🇭]	810	✓	✓	✓	✓	17+2	✓	-	-	-
MUAD	10413	✓	✓	✓	✓	16+9	✓	✓	✓	✓

Number of annotated pixels per class in MUAD:



3/ Experiment results:

Semantic segmentation:

Methods	Architectures	normal set mIoU ↑	normal set ECE ↓	low adv. without OOD set mIoU ↑	low adv. without OOD set ECE ↓	high adv. without OOD set mIoU ↑	high adv. without OOD set ECE ↓	OOD set mIoU ↑	OOD set ECE ↓	OOD set AUROC ↑	OOD set AUPR ↑	OOD set FPR ↓	low adv. with OOD set mIoU ↑	low adv. with OOD set ECE ↓	low adv. with OOD set AUROC ↑	low adv. with OOD set AUPR ↑	low adv. with OOD set FPR ↓	high adv. with OOD set mIoU ↑	high adv. with OOD set ECE ↓	high adv. with OOD set AUROC ↑	high adv. with OOD set AUPR ↑	high adv. with OOD set FPR ↓
Baseline (MCP) [🇨🇭]	DeepLab v3+ [🇨🇭]	68.90%	0.0138	38.77%	0.3238	22.51%	0.4567	57.32%	0.0607	0.8624	0.2604	0.3943	31.84%	0.3078	0.6349	0.1185	0.6746	18.94%	0.4356	0.6023	0.1073	0.7547
Baseline (MCP) [🇨🇭]	SegFormer-B0 [🇨🇭]	69.04%	0.0135	48.67%	0.1004	30.0%	0.2396	58.91 %	0.06465	0.8578	0.21576	0.4106	**40.22%**	0.1544	0.7448	0.1361	0.5876	**27.51 %**	0.6564	0.6564	0.1071	0.7011
MC-Dropout [🇨🇭]	DeepLab v3+ [🇨🇭]	65.33%	0.0172	42.08%	0.2587	27.68%	0.3846	55.62%	0.0645	0.8439	0.2225	0.4575	33.38%	0.1329	0.7506	0.1545	0.5807	20.77%	0.3809	0.6864	0.1185	0.6751
MC-Dropout [🇨🇭]	SegFormer-B0 [🇨🇭]	68.55%	0.0119	45.01%	**0.0758**	26.59%	**0.1594**	58.81%	**0.0574**	0.8811	0.2535	0.3435	39.64%	**0.1172**	**0.7698**	**0.1557**	**0.5498**	26.52%	**0.1771**	**0.6965**	**0.1237**	**0.6633**
Deep Ensembles [🇨🇭]	DeepLab v3+ [🇨🇭]	69.80%	0.0129	42.81%	0.2444	23.91%	0.4500	58.29%	0.0588	0.871	**0.2802**	0.3760	34.91%	0.2447	0.6543	0.1212	0.6425	20.19%	0.4227	0.6101	0.1162	0.7212
Deep Ensembles [🇨🇭]	SegFormer-B0 [🇨🇭]	**70.00%**	**0.0115**	**49.10%**	0.0837	**31.67%**	0.3167	**59.50%**	0.05928	**0.8843**	0.2611	**0.3342**	40.00 %	0.1400	0.6933	0.1198	0.6290	25.89 %	0.3305	0.5939	0.0959	0.7287

Monocular depth estimation:

Methods	normal set Depth results d1 ↑	normal set Depth results AbsRel ↓	normal set Depth results RMSE ↓	normal set Uncertainty results AUSE RMSE ↓	normal set Uncertainty results AUSE Absrel ↓	low adv. without OOD set Depth results d1 ↑	low adv. without OOD set Depth results AbsRel ↓	low adv. without OOD set Depth results RMSE ↓	low adv. without OOD set Uncertainty results AUSE RMSE ↓	low adv. without OOD set Uncertainty results AUSE Absrel ↓	high adv. without OOD set Depth results d1 ↑	high adv. without OOD set Depth results AbsRel ↓	high adv. without OOD set Depth results RMSE ↓	high adv. without OOD set Uncertainty results AUSE RMSE ↓	high adv. without OOD set Uncertainty results AUSE Absrel ↓	normal set overhead sun Depth results d1 ↑	normal set overhead sun Depth results AbsRel ↓	normal set overhead sun Depth results RMSE ↓	normal set overhead sun Uncertainty results AUSE RMSE ↓	normal set overhead sun Uncertainty results AUSE Absrel ↓	OOD set Depth results d1 ↑	OOD set Depth results AbsRel ↓	OOD set Depth results RMSE ↓	OOD set Uncertainty results AUSE RMSE ↓	OOD set Uncertainty results AUSE Absrel ↓	low adv. with OOD set Depth results d1 ↑	low adv. with OOD set Depth results AbsRel ↓	low adv. with OOD set Depth results RMSE ↓	low adv. with OOD set Uncertainty results AUSE RMSE ↓	low adv. with OOD set Uncertainty results AUSE Absrel ↓	high adv. with OOD set Depth results d1 ↑	high adv. with OOD set Depth results AbsRel ↓	high adv. with OOD set Depth results RMSE ↓	high adv. with OOD set Uncertainty results AUSE RMSE ↓	high adv. with OOD set Uncertainty results AUSE Absrel ↓
Baseline	0.922	0.114	3.357	-	-	0.786	**0.147**	5.005	-	-	0.632	**0.207**	6.989	-	-	0.951	0.090	3.646	-	-	0.896	0.125	3.616	-	-	0.713	2.637	4.764	-	-	0.555	0.459	6.916	-	-
Deep Ensembles [🇨🇭]	**0.929**	**0.111**	**3.199**	**0.291**	0.060	0.767	0.156	4.892	**0.740**	0.105	0.566	0.243	7.498	**1.182**	0.153	**0.955**	**0.083**	3.479	**0.336**	0.055	**0.903**	**0.114**	3.447	**0.427**	0.074	0.709	**1.810**	4.707	**0.692**	**0.129**	0.521	**0.331**	7.411	**1.072**	**0.151**
MC Dropout [🇨🇭]	0.919	0.119	3.209	0.634	0.061	**0.798**	0.151	**4.580**	1.063	0.098	**0.657**	**0.207**	**6.278**	1.382	**0.128**	0.948	0.092	**3.407**	0.786	0.058	0.893	0.145	**3.432**	0.724	0.080	**0.744**	3.925	**4.364**	0.927	0.206	**0.610**	0.545	**6.176**	1.245	0.314
Single-PU [🇨🇭]	0.905	0.132	3.230	0.313	0.081	0.773	0.159	4.865	0.789	0.112	0.571	0.248	7.680	1.740	0.171	0.946	0.105	3.546	0.358	0.079	0.888	0.132	3.463	0.447	0.095	0.714	4.349	4.716	0.744	0.482	0.529	0.351	7.627	1.347	0.156
SLURP [🇨🇭]	0.922	0.114	3.357	0.467	**0.048**	0.786	**0.147**	5.005	1.167	**0.090**	0.632	**0.207**	6.989	1.707	**0.128**	0.951	0.090	3.646	0.525	**0.033**	0.896	0.125	3.616	0.721	**0.068**	0.713	2.637	4.764	1.072	0.212	0.555	0.459	6.916	1.564	**0.151**

Object detection:

Evaluation data	normal set mAP ↑	normal set AP50 ↑	normal set PDQ ↑	low adv. without OOD set mAP ↑	low adv. without OOD set AP50 ↑	low adv. without OOD set PDQ ↑	high adv. without OOD set mAP ↑	high adv. without OOD set AP50 ↑	high adv. without OOD set PDQ ↑	OOD set mAP ↑	OOD set AP50 ↑	OOD set PDQ ↑	low adv. with OOD set mAP ↑	low adv. with OOD set AP50 ↑	low adv. with OOD set PDQ ↑	high adv. with OOD set mAP ↑	high adv. with OOD set AP50 ↑	high adv. with OOD set PDQ ↑
Faster R-CNN (ResNet101)	**39.91%**	**54.91%**	**16.88%**	25.00%	36.89%	**8.61**	**13.97%**	**22.01%**	**0.041**	**35.85%**	**48.9%**	**14.33%**	**24.73%**	**35.70%**	**8.49%**	**12.41%**	**19.66%**	**3.86%**
Faster R-CNN (ResNet50)	38.43%	53.13%	15.02%	**25.19%**	**37.38%**	8.18%	13.29%	21.53%	0.0389%	34.52%	47.63%	12.96%	23.93%	34.51%	7.95%	12.11%	19.46%	3.64%
Gaussian YOLOV3 [🇨🇭]	20.81%	32.84%	2.22%	8.79%	16.40%	0.57%	3.28%	6.30%	0.22%	17.44%	28.16%	1.52%	10.80%	18.71%	0.64%	3.21%	6.15%	0.26%

Semantic segmentation simple domain adaptation from MUAD to Cityscapes

Training set	mIoU ↑
Baseline trained on Cityscapes	76.84%
Baseline trained on MUAD	16.71%
Baseline trained on MUAD with histogram eq.	32.12%
Baseline trained on GTA [🇨🇭]	32.85%
Baseline trained on SYNTHIA [🇨🇭]	29.45%

Monocular depth estimation simple domain adaptation from MUAD to KITTI eigen-split

Training set	KITTI d1↑	KITTI d2↑	KITTI d3↑	KITTI Abs Rel↓	KITTI Sq Rel↓	KITTI RMSE↓	KITTI RMSE log↓	KITTI SILog↓
KITTI [🇨🇭]	0.975	0.997	0.999	0.052	0.148	2.072	0.078	6.9859
Virtual KITTI 2 [🇨🇭]	0.835	0.957	0.989	0.129	0.706	4.039	0.177	15.534
MUAD	0.731	0.927	0.983	0.187	1.059	4.754	0.227	18.581