TESTING ROBUSTNESS OF COMPUTER VISION SYSTEMS

MARKUS MURSCHITZ

Autonomous Systems

Center for Vision, Automation & Control

AIT Austrian Institute of Technology GmbH

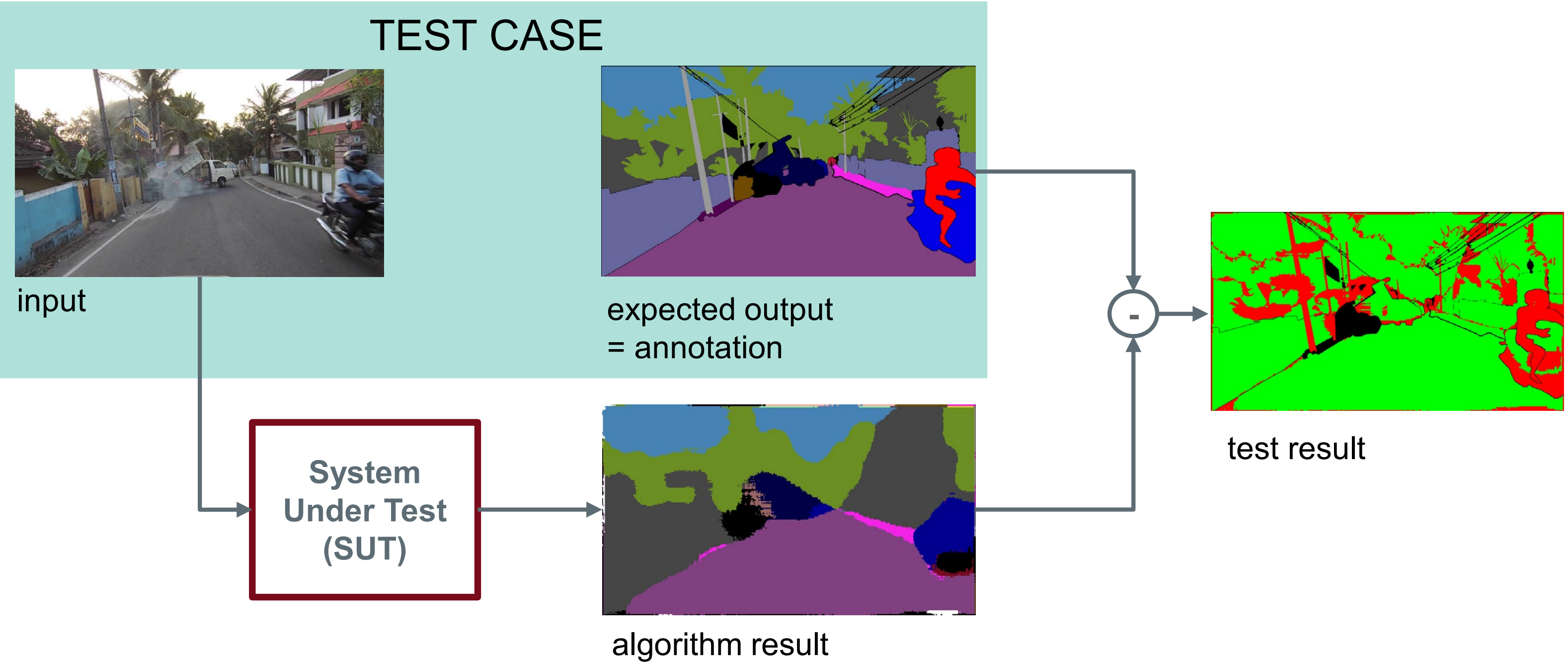
www.vitro-testing.com, wilddash.cc

INTRODUCTION

- Many different computer vision applications

- Many are machine learned and sometimes hard to reason about
- Common question: How good do they work?

TESTING COMPUTER VISION – AN EXAMPLE



WHAT WE WANT FROM TESTING

- Multiple solutions for a problem, which is the best? => fairness
- Does the system have weaknesses? If so, which? => challenges/hazards

VISION TESTING FOR ROBUSTNESS

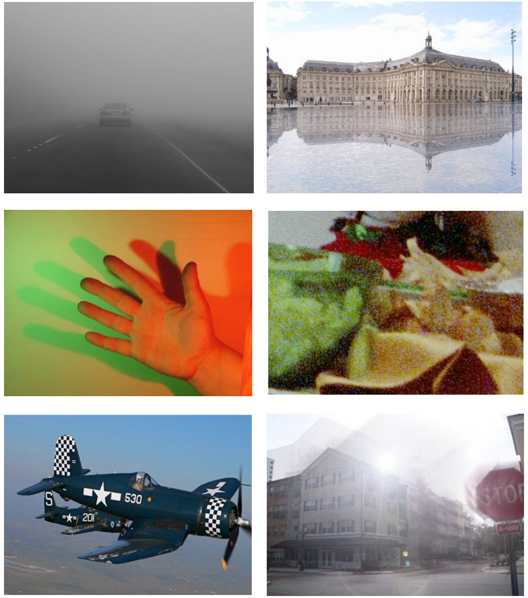
⇒ We need many test cases which are well selected and organized!

FINDING AND ORGANIZING CHALLENGES

- We performed a Hazard and Operability Study on Computer Vision (CV-HAZOP)
  - Established method to find vulnerabilities in the chemical industries
  - Yields a list of ~1500 potential weakness for CV algorithms
    - https://vitro-testing.com/cv-hazop/

- The Checklist can be used for:
    - Evaluating datasets
    - Combining datasets
    - Planning new datasets

CHALLENGES - EXAMLES



TIMELINE OF CV-HAZOP

2014 | 2015 | 2016 | 2017 | 2018

CV-HAZOP

Application to Stereo Vision

Application to Semantic Labeling

June18, Salt Lake City

We where co-hosting the CVPR Workshop

http://www.robustvision.net/

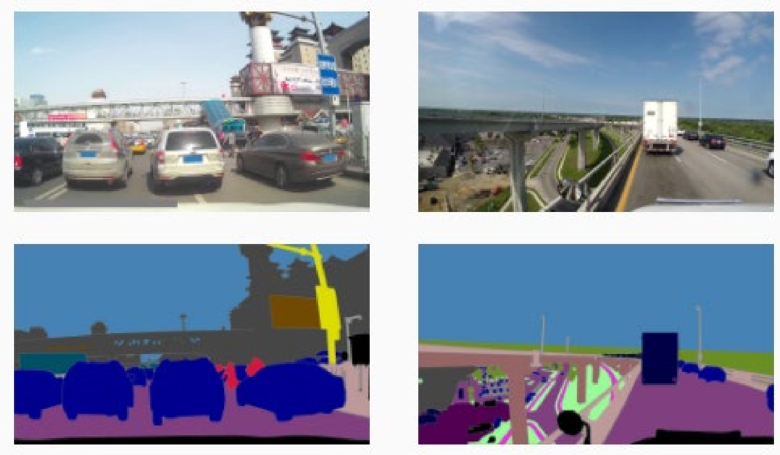
with our semantic segmentation dataset

http://www.wilddash.cc/

[2015] O.Zendel, M.Murschitz, M.Humenberger, and W.Herzner, CV-HAZOP: Introducing test data validation for computer vision, ICCV
[2016] O.Zendel, M.Murschitz, M.Humenberger, and W.Herzner, How Good Is My Test Data? Introducing Safety Analysis for Computer Vision, IJCV
[2017] O.Zendel, K.Honauer, M.Murschitz, M.Humenberger, and G.D. Fernandez, Analyzing Computer Vision Data - The Good, the Bad and the Ugly, CVPR
[2018] O.Zendel, K.Honauer, M.Murschitz, D.Steininger, G.D. Fernandez, WildDash - Creating Hazard-Aware Benchmarks ECCV

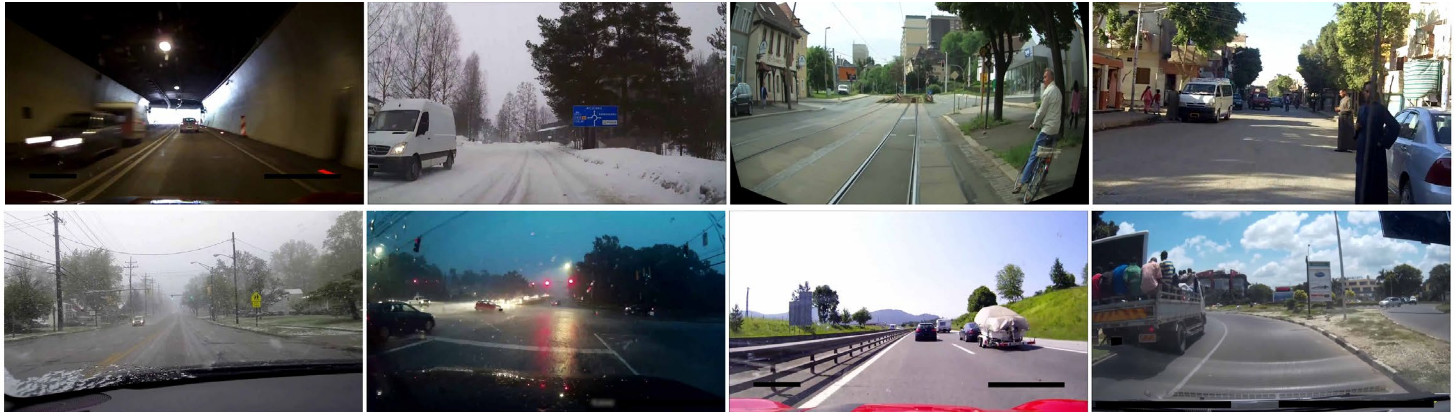
WILDDASH

- Risk-Aware Benchmarking for Semantic Segmentation & Instance Segmentation
- Diverse scenes from all over the world
- Includes challenging visual conditions (e.g. underexposure, overexposure, poor weather) and negative test cases
- http://www.wilddash.cc



WILDDASH SCENARIOS

- Driving Scenes from all over the world
- Mined from public internet sources
- Diverse mixture of countries, situations, weather conditions (fairness)
- Many different cameras / noise levels / compression qualities (challenges)



CV-HAZOP FOR SEMANTIC SEGMENTATION

- Group main hazards by their influence on output image
  - Blur (motion, focus, compression)
  - Road Coverage
  - Distortion
  - Occlusion
  - Overexposure
  - Particles (mist, fog, rain, snow, falling leaves)
  - Underexposure
  - Intra-Class Variations
  - Windscreen (interior refl., smudges, water)
  - Hood

SEVERITY OF VISUAL CHALLENGES

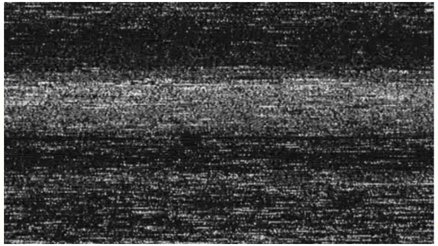
- For each image evaluate severity of each challenge/hazard
  - Three severity levels: none, low, high

- Identified hazards guide selection of images for dataset
  - > 15 frames per hazard and severity level
  - => now we can investigate the impact of each hazards

NEGATIVE TEST CASES

- Tests where we expect the algorithm to fail e.g.:
  - Mixed up color channels / transmission errors / lots of noise
  - Blocked sensor
  - Completely out-of-scope images






- Good algorithm should mark pixels as invalid (= „void“)
- Bad algorithm will likely „hallucinate“ events => creates false positives

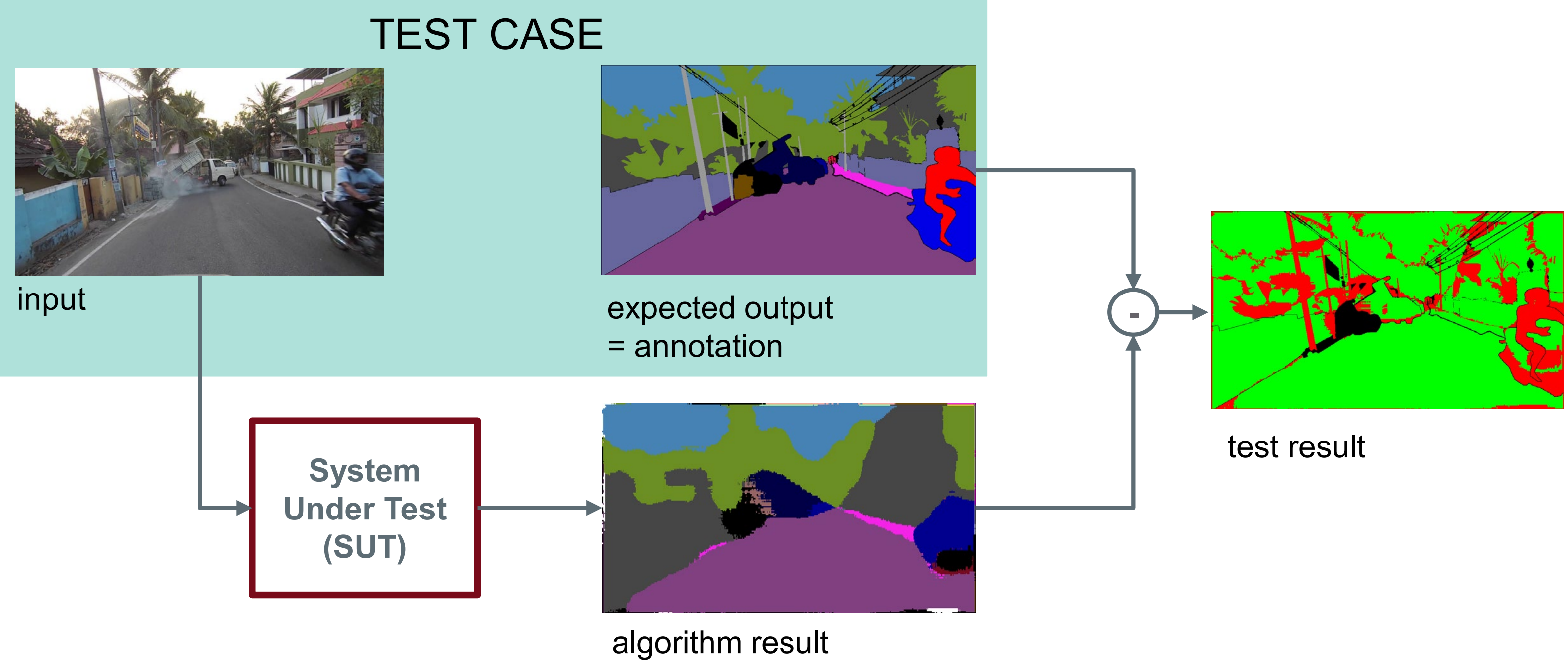
AUSTRIAN INSTITUTE OF TECHNOLOGY

THE CHALLANGE - SUBMISSIONS

Algorithm	Meta AVG	Classic				Negative	Impact (IoU class)									
	IoU Class	IoU Class	iIoU Class	IoU Cat.	iIoU Cat.	IoU Class	Blur	Coverage	Distortion	Hood	Occl.	Overexp.	Particle	Screen	Underexp.	Variation
AHiSS_ROB	39.0%	41.0%	32.2%	53.9%	39.3%	43.6%	-11%	-12%	-2%	-24%	0%	-27%	-13%	-13%	**-28%**	-16%
MapillaryAI_ROB	38.9%	41.3%	38.0%	60.5%	57.6%	25.0%	-15%	-5%	-4%	-23%	0%	-23%	-12%	-21%	**-25%**	-6%
PSP-IBN-SA_ROB	38.5%	39.4%	33.6%	60.6%	51.0%	65.3%	-18%	-3%	-5%	-18%	-3%	**-27%**	-17%	-13%	-27%	-12%
IBN-PSP-SA_ROB	33.6%	34.7%	30.8%	55.1%	38.9%	68.5%	-8%	0%	0%	-22%	0%	-27%	-23%	-23%	**-36%**	-8%
IBN-PSA-SA_ROB	32.5%	33.6%	30.1%	53.8%	39.3%	69.5%	-9%	-1%	0%	-25%	0%	-28%	-25%	-20%	**-32%**	-11%
LDN2_ROB	32.1%	34.4%	30.7%	56.6%	47.6%	29.9%	-7%	-0%	-11%	-36%	0%	-37%	-16%	-24%	**-42%**	-6%
BatMAN_ROB	31.7%	31.4%	17.4%	51.9%	37.3%	36.3%	-9%	-8%	-11%	-20%	-11%	-29%	-5%	-10%	**-37%**	-6%
Mapillary_ROB	31.6%	32.7%	27.5%	55.2%	51.1%	22.7%	-12%	-7%	-15%	-23%	-1%	-26%	-12%	-28%	**-31%**	-3%
ifly	31.4%	31.3%	25.3%	58.0%	51.1%	19.0%	-10%	-18%	-13%	-19%	-7%	-22%	0%	-8%	**-30%**	0%
HiSS_ROB	31.3%	31.0%	16.3%	50.3%	34.6%	44.1%	-11%	-10%	-11%	-25%	-10%	-32%	-2%	-10%	**-44%**	-0%
DeepLabv3+_CS	30.6%	34.2%	24.6%	49.0%	38.6%	15.7%	-13%	-15%	-15%	-34%	0%	**-55%**	-17%	-23%	-53%	-6%
AdapNetv2_ROB	29.5%	28.7%	16.5%	51.5%	38.0%	43.6%	-15%	-10%	-20%	-24%	-14%	-21%	-8%	-7%	**-37%**	-7%
VlocNet++_ROB	29.2%	28.4%	16.4%	51.3%	37.3%	39.4%	-19%	-8%	-17%	-23%	-14%	-23%	-4%	-9%	**-36%**	-11%
DRN_MPC	28.3%	29.1%	13.9%	49.2%	29.2%	15.9%	-17%	-8%	-15%	-32%	-5%	**-47%**	-3%	-12%	-34%	-9%
VENUS_ROB_update	28.2%	29.8%	22.7%	51.5%	35.0%	50.6%	-3%	-0%	0%	-32%	0%	-42%	-15%	-31%	**-43%**	-21%
DN_2_4_CITY_WD	27.2%	28.3%	18.2%	50.6%	38.6%	17.5%	-5%	-3%	-10%	-40%	0%	**-45%**	-15%	-23%	-44%	0%
DRN_MPS	26.3%	27.4%	11.9%	47.5%	27.1%	12.9%	-19%	-12%	-14%	-32%	-8%	**-51%**	-9%	-12%	-45%	-14%
VENUS_ROB	25.1%	26.4%	19.8%	46.9%	29.8%	54.4%	-2%	-0%	0%	-37%	0%	**-49%**	-17%	-30%	-48%	-16%
GoogLeNetV1_ROB	22.9%	22.4%	17.3%	36.7%	36.6%	50.7%	-21%	-21%	-43%	-26%	-9%	-29%	-21%	-28%	**-46%**	-2%
APMoE_seg_ROB	22.2%	22.5%	12.6%	48.1%	35.2%	22.8%	-11%	-2%	-23%	-23%	-4%	-44%	-12%	-11%	**-46%**	0%
PAG_ROB	22.1%	21.7%	12.5%	48.8%	35.6%	34.1%	-9%	-10%	-20%	-27%	-3%	-35%	-6%	-8%	**-41%**	-3%
DRN_CS	14.8%	15.4%	7.1%	28.9%	14.2%	7.2%	-43%	-9%	-29%	-29%	-15%	-27%	-18%	-24%	**-74%**	-35%
FCN101_ROB	12.2%	11.1%	2.1%	29.3%	8.3%	38.7%	0%	-7%	-26%	-27%	-11%	**-49%**	-17%	-4%	-32%	-10%
PSPNetv0	8.3%	8.5%	5.5%	17.7%	15.5%	10.1%	-17%	-33%	-10%	-20%	0%	-34%	-26%	**-52%**	-30%	-32%

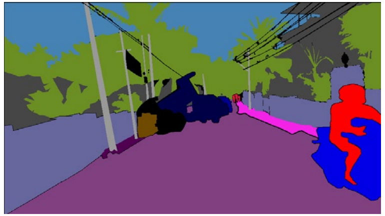
[Cached June 13, 2018, 7:42 p.m. UTC+0]

TESTING COMPUTER VISION - RECAP



TESTING COMPUTER VISION – THE ISSUE

input

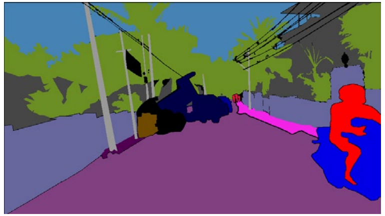


expected output
= annotation

<=>

TESTING COMPUTER VISION – THE ISSUE

input



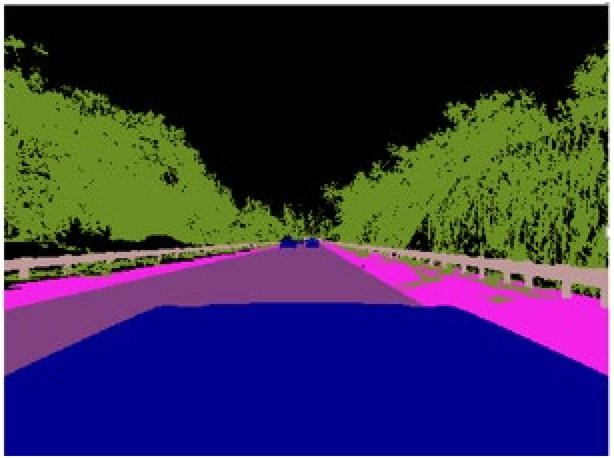
expected output
= annotation

<=>

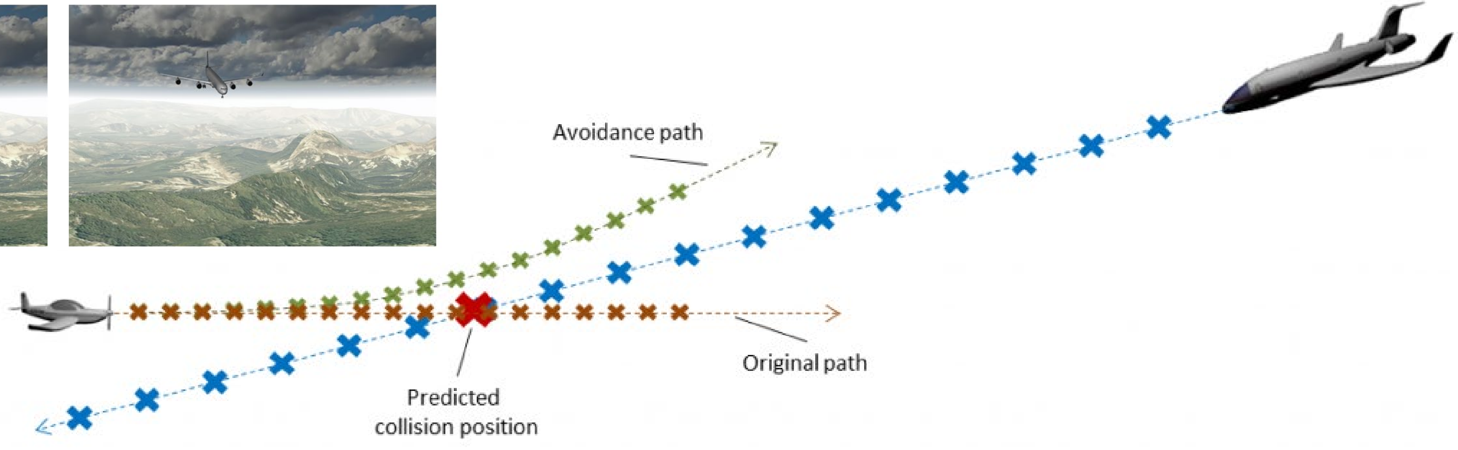
- => Synthetic test data by generating both input and expected output
- => Can also be used for training data

SYTHETIC TEST DATA - RESULTS

SYTHETIC TEST DATA – RESULTS - WEATHER



SYNTHETIC TEST DATA – AERIAL COLLISION AVOIDANCE

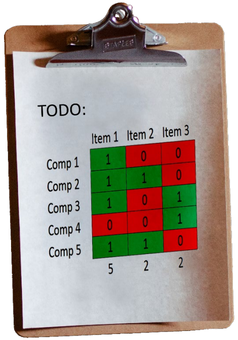


[Mur2016] Murschitz, M., Zendel, O., Humenberger, M., Sulzbachner, C., & Domınguez, G. F. An Experience Report on Requirements-Driven Model-Based Synthetic Vision Testing. Quality Assurance in Computer Vision 2016

CONCLUSION

- Use Checklists to increase the quality of datasets
- CV-HAZOP is a good starting point / framework
- WildDash allows the calculation of hazard impact factors

  ⇒ allows the backtracking of bad results to actual reasons



Better data ⇒ better systems

CVPR 2018 Robust Vision Challenge: http://www.robustvision.net

Access CV-HAZOP and datasets: www.vitro-testing.com

Contact: markus.Murschitz@ait.ac.at, oliver.zendel@ait.ac.at

REFERENCES

[Bow2001] K. Bowyer, C. Kranenburg, and Sean Dougherty. Edge Detector Evaluation Using Empirical ROC Curves. In Computer Vision and Image Understanding 84ff, 2001.

[Ble2011] M. Bleyer, C. Rhemann, and C. Rother. Patchmatch stereo-stereo matching with slanted support windows. In British Machine Vision Conference, 2011.

[Don2013] A. Donath, D. Kondermann. Is crowdsourcing for optical ow ground truth generation feasible? In Prroceeding to the International Conference on Vision Systems, 2013.

[But2012] D. J. Butler, J. Wulff, G. B. Stanly, and M. J. Black. A Naturalistic Open Source Movie for Optical Flow Evaluation. In European Conference on Computer Vision, 2012.

[Gai2016] A. Gaidon, Q. Wang, Y. Cabon and E. Vig. Virtual Worlds as Proxy for Multi-Object Tracking Analysis. In Computer Vision and Pattern Recognition, 2016.

[Gei2012] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? The kitti vision benchmark suite. In Computer Vision and Pattern Recognition, 2012.

[Hir2008] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30(2):328ff, 2008.

[Hum2010] M. Humenberger, C. Zinner, M.Weber,W. Kubinger, and M. Vincze. A fast stereo matching algorithm suitable for embedded real-time systems. Computer Vision and Image Understanding, 2010.

[Kon1998] K. Konolige. Small vision systems: Hardware and implementation. In Robotics Research. Springer, 1998.

[Kon2015] D. Kondermann, R. Nair, S. Meister, W. Mischler, B. Güssefeld, K. Honauer, S. Hofmann, C. Brenner, and B. Jähne. Stereo ground truth with error bars. In Asian Conference on Computer Vision, 2015.

[Mei2013] X. Mei, X. Sun, W. Dong, H. Wang, and X. Zhang. Segment-tree based cost aggregation for stereo matching. In Computer Vision and Pattern Recognition, 313ff, 2013.

[Men2015] M. Menze and A. Geiger. Object Scene Flow for Autonomous Vehicles. Conference on Computer Vision and Pattern Recognition, 2015.

[Pin2008] N. Pinto, D. D. Cox, and J. J. DiCarlo. Why is real-world visual object recognition hard? PLoS Comput Biol, 4(1), 2008.

[Pon2006] J. Ponce, T. L. Berg, M. Everingham, D. A. Forsyth, M. Hebert, S. Lazebnik, M. Marszalek, C. Schmid, B. C. Russell, A. Torralba, et al. Dataset issues in object recognition. In Toward category-level object recognition, pages 29–48. Springer, 2006.

[Rhe2011] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In Computer Vision and Pattern Recognition, pages 3017–3024, 2011.

[Ros2016] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. Lopez. The SYNTHIA Dataset. In Computer Vision and Pattern Recognition, 2016.

[Sch2002] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision, 47(1):7ff, 2002.

[Sch2011] M. Schulze. A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method, In Social Choice and Welfare, 2011

[Sch2014] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In Pattern Recognition, pages 31–42. Springer, 2014.

[Tor2011] A. Torralba and A. A. Efros. Unbiased look at dataset bias. In Computer Vision and Pattern Recognition, pages 1521–1528, 2011.

[Zen2015] O.Zendel, M.Murschitz, M.Humenberger, and W.Herzner, CV-HAZOP: Introducing test data validation for computer vision, ICCV 2015 pp. 2066-2074

[Zen2016] O.Zendel, M.Murschitz, M.Humenberger, and W.Herzner, How Good Is My Test Data? Introducing Safety Analysis for Computer Vision, IJCV Volume 125/1–3, pp 95

[Zen2017] O. Zendel, K. Honauer, M. Markus, M. Humenberger, and G.D. Fernandez, Analyzing Computer Vision Data - The Good, the Bad and the Ugly, CVPR 2017

CVPR 2018 Robust Vision Challenge: http://www.robustvision.net

Access CV-HAZOP and data sets: www.vitro-testing.com

Contact: markus.murschitz@ait.ac.at, oliver.zendel@ait.ac.at

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

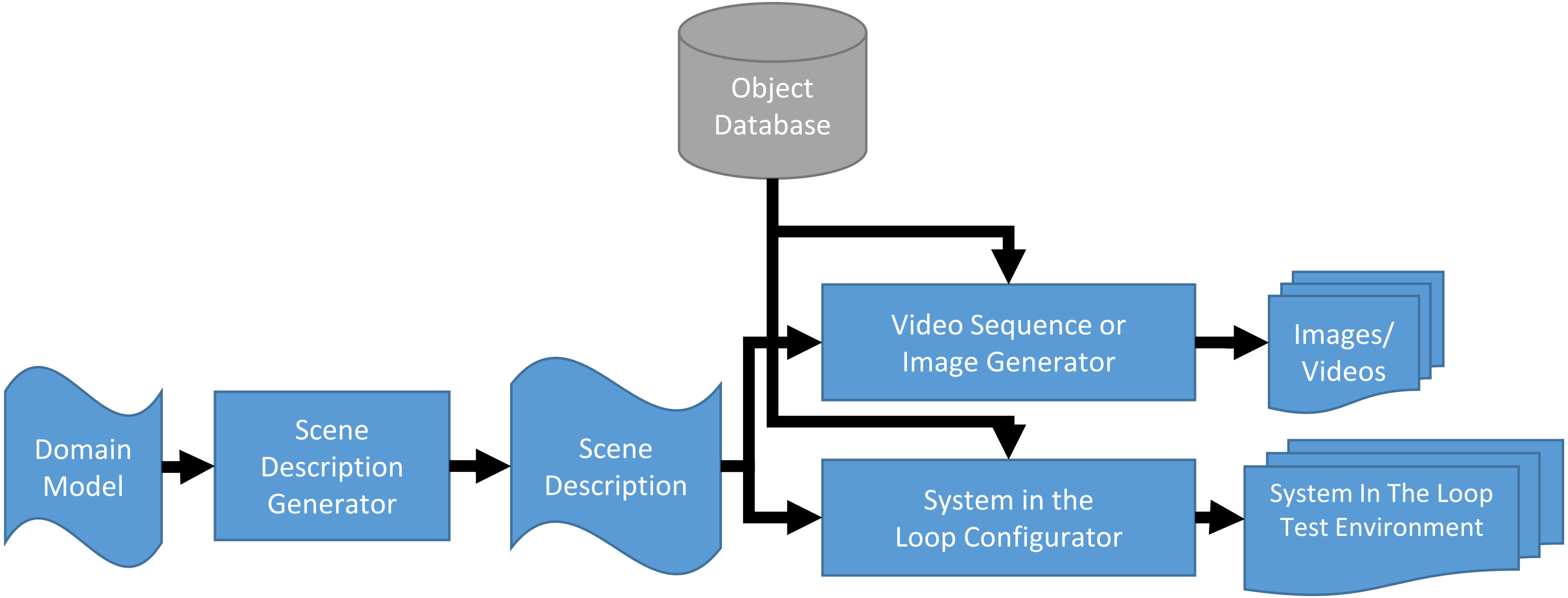
THANK YOU!

TESTING ROBUSTNESS OF COMPUTER VISION SYSTEMS

Additional slides

SYNTHETIC TEST DATA - TRAMWAYS

SYTHETIC TEST DATA



SYNTHETIC TEST DATA – COMPLEX EVALUATIONS – OBJECT DETECTION