VCL Challenges

1st Workshop on Visual Continual Learning, @ICCV 2023

Overviews

Introduction

The challenges are designed to encourage the development of novel methods for visual continuous learning (VCL) and to provide a benchmark for evaluating the performance of different methods. The challenges are open to all participants.

General Awards

We will award the top three teams of each challenge with **a certificate and a prize of 1000, 500, and 300 USD**, respectively. The winners of each challenge will be invited to give a **15min presentation** at the workshop. Teams will be selected based on the performance of their methods on the test set.

Innovation Awards

Besides the general awards, we will also award one team from each challenge with an innovation award. The innovation award is given to the team that proposes the most innovative method and/or insightful analysis. The winner will receive **a certificate and an additional prize of 300 USD**.

To be eligible for the innovation award, the team must submit a technical report to the workshop. The report should be 2-4 pages long and include the following information: the motivation of the method, the model description, implementation and training details, the analysis of the results, and the discussion of the limitations. The report should be in PDF format and submitted to the workshop email address vcl.iccvworkshop.2023@gmail.com with the subject "[Challenge X Technical Report Submission]".

Challenge Results

Last update: 2023/09/10. We're delighted to announce that a total of **47 teams** have participated in our three challenges. Additionally, **8 technical reports** have been submitted for the innovation award. A big thank you to all participants for their outstanding contributions!

Winners of the challenges

Congratulations to the following champions of each challenge!

- **Challenge A - Multitask Learning for Robustness**
  Winner: UniNet-USTC&PKU
  Members: Zehui Chen (USTC), Qiuchen Wang (USTC), Zhenyu Li (KAUST), Jiaming Liu (PKU), Shanghang Zhang (PKU), Feng Zhao (USTC)
- **Challenge B - Continual Test-time Adaptation for Object Detection**
  Winner: YSVnL
  Members: Dongjae Jeon, Taeheon Kim, Seongwon Jo, Minhyuk Seo, Jonghyun Choi (Yonsei University)
- **Challenge B - Continual Test-time Adaptation for Semantic Segmentation**
  Winner: PKU_BLV_LAB
  Members: Ran Xu (PKU), Peidong Jia (PKU), Senqiao Yang (PKU), Jiayi Ni (PKU), Jiaming Liu (PKU), Zehui Chen (USTC), Feng Zhao (USTC), Yandong Guo (AI2Robot), Shanghang Zhang (PKU)
- **Challenge C - Robust RAW Object Detection**
  Winner: IPIU-XDU
  Members: Xiaoqiang Lu, Licheng Jiao, Xu Liu, Yuting Yang, Zhongjian Huang, Jiaxuan Zhao, Fang Liu (Xidian University)

Winners of the Innovation Award

We will announce the winner team of the Innovation Award soon.

Scores

You can view the detailed scores in the tables below. The final rankings are determined by the "overall" score from the testing set of each challenge. Please note: only the teams that have successfully submitted their test results are displayed. If you have any questions regarding the scores, please contact us at vcl.iccvworkshop.2023@gmail.com.

Challenge A - Multitask Learning for Robustness

	Team	Overall	Insseg_mAP	Depth_SILog	Det3D_mAP	Det3D_mTPS
1 ★	UniNet-USTC&PKU	49.2	39.3	16.9	9.8	74.5
	(Baseline: Mask3D)	31.5	14.1	31.5	19.9	67.0

Challenge B - Continual Test-time Adaptation for Object Detection

	Team	Overall	mAP	mAP_drop
1 ★	YSVnL	49.0	49.3	0.2
2 ★	USTC-BIVLab	43.3	60.4	8.7

	Team	Overall	mAP	mAP_drop
3 ★	DicaLab	32.1	40.8	4.3
4	JFFF	31.4	38.9	3.8
5	Continual learners	28.5	39.9	5.7
6	NKU_TTA	27.0	40.8	6.9
7	JFF	26.7	39.4	6.3
8	YM	25.1	39.3	7.1
	(Baseline: No adaptation)	24.2	38.8	7.3

Challenge B - Continual Test-time Adaptation for Semantic Segmentation

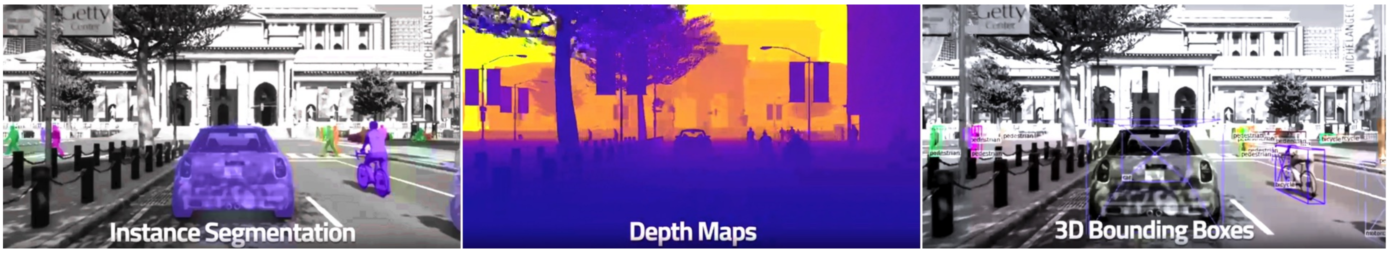
	Team	Overall	mIoU	mIoU_drop
1 ★	PKU_BLV_LAB	74.3	84.6	5.2
2 ★	ContinualBD	69.4	80.5	5.5
3 ★	Continual learners	38.7	71.5	16.4
4	NKU_CTTA	14.8	68.9	27.1
	(Baseline: No adaptation)	18.0	71.9	26.9

Challenge C - Robust RAW Object Detection

	Team	Overall	AP50	AP75	mAP	R@50@95
1 ★	IPIU-XDU	299.3	88.6	76.3	67.3	67.1
2 ★	X_ROD	292.9	87.3	74.7	66.3	64.7

	Team	Overall	AP50	AP75	mAP	R@50@95
3 ★	DLUT_IIAU	253.9	76.1	63.9	57.3	56.6
4	QYK	81.0	28.6	19.2	18.2	15.1

Challenge A - Multitask Learning for Robustness



Goal

In this challenge, we aim to develop **multitask learning (MTL)** models that are robust to different domains. We believe that the ability to learn a unified model for multiple 2D and 3D tasks simultaneously is a key component of robustness. The tasks include a combination of monocular 2D and 3D vision tasks:

- 2D Instance segmentation
- Monocular depth estimation
- 3D object detection

Data usage

All participants will use the SHIFT Dataset for developing their MTL methods.

- **Training:** IMAGES DEPTH MAPS (8BIT) INSTANCE ANNOTATIONS 3D DETECTION ANNOTATIONS
- **Online evaluation (Val Phase):** IMAGES
- **Online evaluation (Test Phase):** IMAGES

Here, we provide the depth maps in **8bit** PNG format to reduce the file size, whereas the original data are in **24bit** PNG format. The 255 values in the depth maps correspond to 80m, which is the maximum depth used also in the evaluation.

Evaluation protocol

The participants will be evaluated on the **test split** of the SHIFT dataset. The metrics for each task are as follows:

- 2D Instance segmentation: mAP

- The evaluation will be performed following COCO's instance segmentation protocol.

- Monocular depth estimation: SILog
  - The evaluation will be performed following KITTI's scale-invariant log difference metric.
  - Only the points whose ground truth depth within the valid range (0.01m - 80m) will be evaluated.

- 3D object detection: mAP, mTPS
  - The evaluation will be performed following a simplified NuScene protocol.
  - For the mean average precision (mAP) metrics, we define a match by considering the 2D center distance on the ground plane, same as NuScenes. The match distance thresholds used are {0.5, 1, 2.0} meters.
  - For the mean true positive scores (mTPS) metric, the distance threshold used is 1.0 meter. The mTPS is 1 minus the mean of the Average Translation Error (ATE), Average Scale Error (ASE), and Average Orientation Error (AOE).

The final score will be the average of the scores of all the tasks.

Computational resources

We recommend the participants to have access to at least 4 GPUs for developing the models. The data will occupy around 50GB of disk space. You will need more disk space for storing the HDF5 files.

Codes

A baseline method for this challenge can be found at Mask3D-SHIFT. Thanks to @Fan Chengxiang and @Jiabei Xiao for their efforts.

Submission

Please use the submit your results at the VIS Eval server.

Challenge B - Continual Test-time Adaptation

Goal

Adapting to the continuously changing environment plays a curial role in safety for all the vision models. In this challenge, we aim to develop **test-time adaptation (TTA)** methods that can adapt the model in the sequence of domain shifts continuously. Here, we are interested in two fundamental vision tasks separately:

- Semantic segmentation

- Object detection

These two tasks are evaluated and ranked separately. But, the participants are encouraged to develop TTA methods that can be applied to both tasks.

Data usage

All participants will use the SHIFT Dataset for developing their TTA methods.

- **Training of source models:** IMAGES SEMANTIC MASKS DETECTION ANNOTATIONS
- **Developing your test-time adaptation methods:** VIDEOS SEMANTIC MASKS DETECTION ANNOTATIONS
- **Online evaluation (Val Phase):** VIDEOS
- **Online evaluation (Test Phase):** VIDEOS

For a fair comparison, the participants are NOT allowed to use any other data for training or test-time adaptation.

Evaluation protocol

The TTA methods will be evaluated in a per-sequence manner, i.e., the model must be evaluated on each sequence separately and be reset to source model before the next sequence. The evaluation metrics will be the average performance across all the sequences.

Computational resources

We recommend the participants to have access to at least 2 GPUs for developing the models. The data will occupy around 50GB of disk space. If you have limited computational resources, you can use our pretrained models for developing your TTA methods. The pretrained models can be found at the following code repository.

Codes

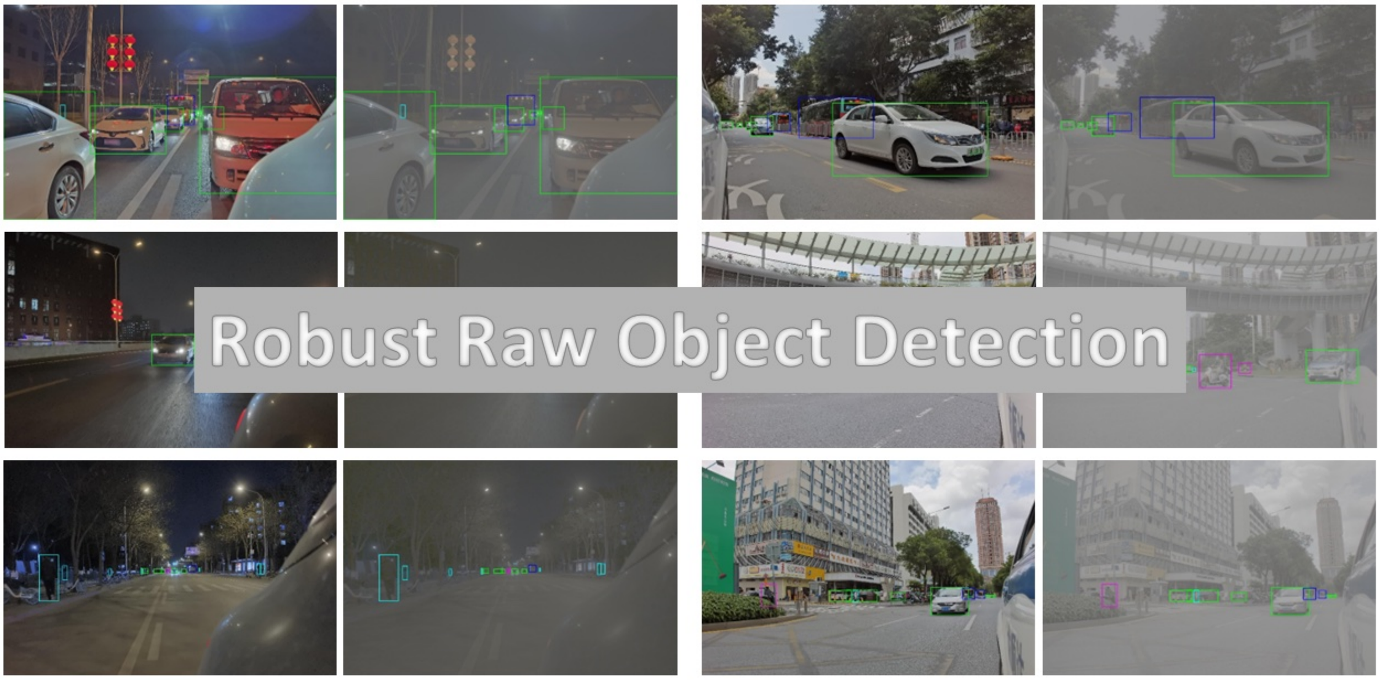
We provide reference implementations of the baseline methods and the evaluation codes:

- **Object detection**: github.com/SysCV/shift-detection-tta.
- **Semantic segmentation**: github.com/zwbx/SHIFT-Continual_Test_Time_Adaptation.

Submission

Please use the submit your results at the VIS Eval server.

Challenge C - Robust RAW Object Detection



Goal

In this challenge, we aim to develop Raw Object Detectors that are robust to different scenes, i.g., lighting conditions. Traditional object detection methods take 8-bit sRGB images as inputs. However, sRGB images suffer from poor imaging quality, especially in extreme cases, e.g., low light conditions or scenes with strong bright / dark contrast. On the other hand, 24-bit RAW data, which is not degraded by any processor, contains more information than those sRGB conterparts. We believe such RAW images possibly lead to a better detection performance.

Making one step furthur, we want raw detectors that can handle both day and night scenes. Different lighting conditions, to some extent, are a powerful data augmentation. Detectors that are available for various lighting conditions is a key component of robustness.

Participants should get detection results for both day and night scene with a single model.

Data usage

We provide ROD dataset for participants to develop their Raw Detectors. Additional training data is allowed.

- **Training Phase:** IMAGES AND ANNOTATIONS
- **Val Phase:** IMAGES
  Validation annotations will be released when the challenge is over.
- **Test Phase:** Testing data will NOT provided to participants. Participants should submit inference codes and also well-trained weights. Organizers will run the code and get the final ranking.

Evaluation protocol

The participants will be evaluated on the test split of the ROD dataset. Evaluation metrics include AP50, AP75, mAP, Rp95.

- For AP50, AP75 and mAP: We follow this repo to calculate.

- For Rp95: It is defined as the highest Recall when Precision is greater or equal to 0.95. IOU is set to be 0.5.

The final score will be the average of all four metrics.

Computational resources

We recommend the participants to have access to at least 4 GPUs for developing the models. It generally takes 3 days for training 300 epochs. The data will occupy around 250 GB of disk space.

Codes

Please refer to this repo for the baseline methods and evaluation codes.

Submission

Please use the submit your results at the VIS Eval server.

Organizers

Tao Sun
Challenge Session Chair
ETH Zürich

Mattia Segu
Challenge Session Chair
ETH Zürich

Chengxiang Fan
Multitask Robustness
Zhejiang University

Jiabei Xiao
Multitask Robustness
Zhejiang University

Wenbo Zhang
Domain Continuous Learning
University of Adelaide

Jingyang Peng
RAW image Object Detection
Huawei Noah's Ark Lab

Copyright © VCL Workshop Organizing Team 2023 .