Welcome to the 2nd OpenDR Project Newsletter !

Newsletter 2.0: First release of the toolkit and current status

Almost everything we hear about 0artificial intelligence today is thanks to deep learning (DL). Deep learning has achieved tremendous performance jumps in the last decade in several Computer Vision (CV) and Machine Learning (ML) tasks, achieving in many cases super-human performance. However, DL cannot be currently fully exploited in robotics scenarios due to a number of barriers.

Learning Curve Barrier

DL has a **steeper learning curve** than traditional CV and ML methods

Computational Complexity Barrier

DL requires vast amounts of computational power and **energy**

Static Perception Barrier

DL is applied on **static environments** and **does not exploit spatial or temporal embodiment**

The need for an open deep learning toolkit that contains easy to train and deploy real-time, lightweight, Robot Operating System (ROS) compliant deep learning models for robotics is evident. This is where the **OpenDR** project enters.

What is OpenDR?

OpenDR "Open Deep Learning for Robotics Toolkit", is a EU 2020 Project which was launched on January 2020 and aims to develop a modular, open and non-proprietary toolkit for core robotic functionalities by harnessing deep learning to provide advanced perception and cognition capabilities, meeting in this way the general requirements of robotics applications in the areas of healthcare, agri-food and agile production. The **OpenDR** project is coordinated by the Aristotle University of Thessaloniki, Greece and will be running throughout December 2023 with a total budget of 6.6 Million Euros.

OpenDR will enable real-time robotic visual perception on high-resolution data and enhance the robotic autonomy exploiting lightweight deep learning for deployment on robots and devices with limited computational resources. In addition, it aims to propose, design, train and deploy models that go beyond static computer vision and towards active robot perception, providing deep human-centric and environment active robot perception, as well as enhanced robot navigation, action and manipulation capabilities.

OpenDR's expected impact is to improve the technical capabilities in robotics by providing easily deployable, efficient and novel Deep Learning tools, as well as to lower the technical barriers by providing a modular and open platform for developing Deep Learning for Robotics tools. Concerning industry, the project's expected impact is to enable a greater range of applications in agri-food, healthcare robotics and agile production, as well as to strengthen the competitiveness of companies by lowering the cost to access robotics-oriented Deep Learning tools.

OpenDR Consortium

OpenDR consortium is a very good mix of 8 partners from 7 European Countries: 2 companies working in various fields of robotics, one company working in the field of robotics simulations, and 5 Universities that join the project with 4 robotics laboratories and 3 deep learning and computer/robot vision laboratories.

Aristotle University of Thessaloniki (AUTH) is the largest university in Greece, established in 1925. AUTH coordinates the project and leads the organization of dissemination activities. AUTH is focusing its research on deep human centric active perception and cognition, where it contributes on deep person/face/body part active detection/recognition and pose estimation, deep person/face/body part tracking, human activity recognition, as well as social signal analysis and recognition. AUTH is also leading the research in object detection/recognition and semantic scene segmentation and contributes to other areas such as evaluation and benchmarking activities of the project.

Tampere University (TAU) is Finland's second-largest university with 20.000 students and 330 professors. TAU participates with two labs/groups namely the Laboratory of Signal Processing at the Department of Computing Sciences and the Cognitive Robotics Group at the Department of Automation Technology and Mechanical Engineering. TAU is leading the research in deep human centric active perception and cognition, working mainly on deep speech and biosignals analysis and recognition, and contributes to deep person/face/body part active detection/recognition and multi-modal human centric perception and cognition as well as in a number of other topics. TAU also contributes on defining the agile production use case requirements and specifications and on the integration of OpenDR to this use case.

University of Freiburg (ALU-FR) is one of Germany's leading research institutions with an international reputation in many fields. ALU-FR is leading the research in deep environment active perception and cognition. ALU-FR is focusing its research on Deep SLAM and 3D scene reconstruction, as well as on deep navigation. It also contributes on developing methodologies for deep planning.

Aarhus University (AU), Denmark participates in OpenDR with two groups, namely the Data-Driven Analytics Group and the Artificial Intelligence in Robotics Group, both belonging to the Section of Electrical and Computer Engineering. AU is leading work on 2D/3D Object localization and tracking and works on sensor information fusion, as well as object detection/recognition and semantic scene segmentation and understanding. AU also contributes to a number of areas such as deep person/face/body part active detection/recognition, deep person/face/body part tracking, deep planning, etc.

Delft University of Technology (TUD) is the oldest and largest technical university in the Netherlands. TUD is leading/organizing the research activities on deep action and control, deep planning, as well as deep navigation. Furthermore, TUD is also leading and undertaking the research activities on human robot interaction. Finally, it leads and organizes the toolkit evaluation and benchmarking activities of the project.

Cyberbotics (CYB) is a Swiss spin-off company from EPFL, which has been developing the Webots robot simulator since 1998. CYB is leading efforts of defining the toolkit's requirements and specifications. CYB is also working on developing simulation environments and collecting data. Finally, it also leads on toolkit integration by collecting and integrating all the OpenDR modules developed by the partners.

PAL Robotics (PAL) is a Spanish SME that provides robotic products and services. PAL is organizing and coordinating the toolkit integration, as well as the use cases integration activities. PAL also contributes on defining the healthcare robotics use case requirements and specifications and works on the integration of OpenDR Toolkit to this use case, as well as on its evaluation.

Agro Intelligence APS (AGI), Denmark is organizing and coordinating the toolkit evaluation, as well as the use cases specific toolkit evaluation activities. AGI also contributes on defining the agri-food use case requirements and specifications and works on the integration and evaluation of OpenDR Toolkit in this specific use case.

Work Performed in the 2nd Year

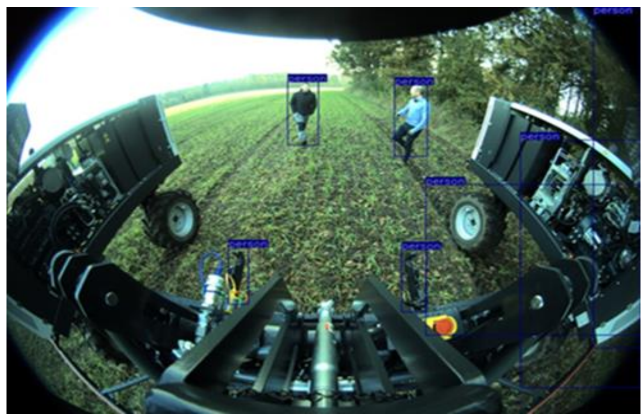
1st Public Release of the Toolkit

Following months of development, integration and debugging as well as countless videoconferencing sessions, the **first official public release of OpenDR** is finally accessible in GitHub, as well as through pip and Docker Hub! OpenDR provides an intuitive and easy-to-use **Python interface**, a **C API** for selected tools, **a wealth of usage examples and supporting tools**, as well as **ready-to-use ROS nodes**. The toolkit provides more than 20 methods, for human pose estimation, face detection, recognition, facial expression recognition, semantic and panoptic segmentation, video and skeleton based action recognition, image, multimodal and point cloud based object detection, 2D and 3D object tracking, speech command recognition, heart anomaly detection, navigation for wheeled robots, and grasping. A set of data generation utilities, a hyperparameter tuning tool and a framework to easily apply RL both in simulation and real robotics applications are also included. All methods and their parameters are thoroughly documented, demonstration examples are available to showcase their functionality, and continuous integration tests ensure both consistency of the code and that no conflicts arise between the different tools. At the same time, OpenDR is built to support Webots Open Source Robot Simulator, while it also extensively follows industry standards, such as ONNX model format and OpenAI Gym Interface. We look forward to receiving your feedback, bug reports and suggestions for improvements!

Deep Human Centric Active Perception and Cognition

Throughout the second year of the project AUTH, AU and TAU have continued their work on the human centric tools and algorithms for the OpenDR project, bringing the team closer to the realization of the powerful, flexible and efficient robotics toolkit.

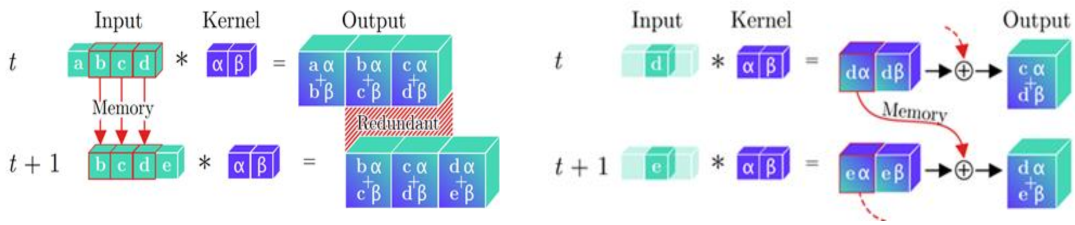
AUTH has made additional contributions to the development of active perception in OpenDR. Specifically, a controller based on deep reinforcement learning, capable of issuing movement commands such that the face recognition confidence from an on-board camera image is maximized. Another work tackled active perception by predicting the best view angles for face recognition via synthetic image generation.



Human detection example on an image collected from the Robotti platform. Both true and false detections are evident, highlighting the need to adapt existing models to the particularities of agriculture use case.

AUTH also investigated artificial data generation to expand on the existing datasets, demonstrating significant performance improvements when both real and synthetic data are used for training DNN models. In addition, AUTH proposed a novel DNN-based Non-Maximum Suppression (NMS) method for the person detection task. The method reformulates NMS for object detection as a sequence-to-sequence problem and proposes a novel DNN-based architecture for classifying a ROI as “correct” or “potentially suppressed”. Finally, AUTH jointly with AGI evaluated the performance of standard human detection models in the practical setting of the OpenDR agriculture use case, using visual data collected from

Robotti agricultural robot, highlighting the need to tackle the domain adaptation problem within the project.



Standard 3D convolution operation (on the left) and proposed continual convolution (on the right). The proposed operation avoids redundant computations and requires less memory

To facilitate efficient video processing, AU has designed continual convolution – a replacement for the standard 3D convolution that allows fast per-frame predictions, while being able to reuse the exact same weights. Continual convolutions can thus be incorporated into existing models without fine tuning, unlike other similar operations from the literature. AU applied continual convolution in its own graph convolutional models for human activity recognition, yielding an increase in throughput from 7.0 times to 19.9 times with the same prediction accuracy. AU also proposed to extend its spatio-temporal graph models from skeleton-based human action recognition to landmark-based facial expression recognition, achieving accuracy comparable with state-of-the-art approaches, while using models of lower complexity.

TAU has started working on human centric multimodal perception and cognition, proposing a novel attention-based methodology where the neural models learn to ignore less relevant portions of the inputs. The developed attention mechanism developed was incorporated into TAU methods, allowing them to learn a joint spatio-temporal mask instead of independent separate masks. A tool for multimodal hand gesture recognition from RGBD data input was also developed.

Deep Environment Active Perception and Cognition

Partners AU, ALU-FR, AUTH, TAU and TUD worked towards providing a multitude of lightweight deep learning methods for deep environment active perception and cognition in various fields.

In the field of object detection/recognition and semantic scene segmentation, AUTH proposed an end-to-end trainable pseudo-active sensory refinement method that works by applying a number of neural transformation layers on the sensor data. This allows for refining the sensory input, without having to reacquire the sensor data. Moreover, AUTH developed a single-stage self-distillation method for improving the performance of any deep neural model, in an online fashion, allowing for training compact yet effective classification models. Furthermore, AUTH worked on improving object recognition methods, by incorporating label embedding criteria into the learning objective of lightweight neural networks, capturing both general class similarities as well as instance specific resemblances between samples. AU proposed a type of neural networks called Variational Neural Networks, which can be used for the estimation of the deep learning model uncertainty. TAU proposed Sparse Representation Matching, a method to transfer intermediate knowledge obtained from one Convolutional Neural Network to another one by using sparse representation learning. TAU also conducted research towards better understanding of learned feature spaces in image classification tasks in deep learning. It is expected that models achieving good performance converge to feature spaces with highly-separable compact classes. TAU found that better compactness is not necessarily associated with better performance.

In the field of object localization and tracking, AU worked towards 3D object detection based on spherical projection images of point cloud data. The proposed method is able to regress the 3D size, rotation and position offset of an object. AU proposed methods for 3D object detection and tracking, based on spherical projection images and Siamese PointPillars, respectively. AUTH developed a 2D tracking methodology that focuses on active object tracking. The tracker receives an input visual observation and directly outputs the most appropriate control actions to follow and keep the target in its field of view. ALU-FR integrated their panoptic segmentation approach for images, EfficientPS, into the toolkit and extended it to LiDAR-based point clouds. Additionally, ALU-FR improved their work towards monocular vision-based localization in 3D maps.

In the field of sensor information fusion TAU has continued its work on the multimodal sensor fusion framework for manipulation tasks. The formulation of the representation learning task was extended with additional objective functions to support input reconstruction and cross-modal compensation, enabling the system to react to corrupted inputs. Additionally, TAU is working on optimizing the structures of the fusion modules and the feature encoder as a whole, using neural architecture search. Finally, TUD has developed efficient sensor fusion strategies in the context of object detection for RGB and IR/depth sensors for harsh lighting conditions. The proposed fusion strategies learn to exploit sensor redundancy in extreme lighting conditions by intelligently determining the scalar/mask weights for the dominant sensor modality. A novel data augmentation scheme has been proposed to mimic such extreme conditions.

Deep Robot Action and Decision Making

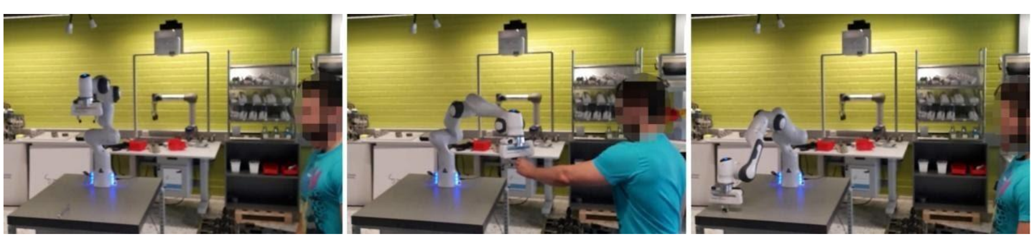
Partners AU, ALU-FR, TUD, TAU worked on the design of novel navigation, planning, and control algorithms, thus contributing to the state of the art and to the key project objectives.

AU proposed a novel end-to-end path planning algorithm for a team of multiple aerial-ground robots, targeting the agriculture use case. The main operations in the field are handled by the ground vehicle whereas the aerial robot is responsible for re-planning a collision-free trajectory for the ground robot in case of obstacles. The method was successfully tested in simulations and real-world experiments.

ALU-FR developed a novel approach for mobile navigation and manipulation. Given an arbitrary end-effector motion, reinforcement learning is used to control the robot base to output actions that ensure that the end-effector motions remain kinematically feasible. The method allows the use of a novel dense reward signal for long horizon tasks and shows strong zero-shot generalisation to unseen tasks and motions.

TUD finalized their work on the design of a novel and efficient model-based agent that learns a latent Koopman representation from images and is robust to possible distractions. This representation allows the method to plan efficiently using linear control methods, such as linear model predictive control. In addition, TUD developed EAGER, an engine agnostic gym environment for robotics tasks that will be integrated in the OpenDR toolkit.

TAU investigated robot grasping models for handling industrial objects in the agile production use case. Their single demonstration grasping has been integrated in Webots and in the OpenDR toolkit. In addition, a human-robot interaction scenario was developed that enables the collaboration between human and robot based on commands given by a person.



Different stages of the hand-over scenario by speech command-based coordination.

Simulation Environments and Data

AUTH finalized an annotated mixed (real along with synthetic) image dataset for human-centric perception tasks. The dataset was generated by using deep learning-generated 3D human models to populate real background images with humans in various poses and positions. In addition, starting from hand-crafted 3D human models it generated a dataset of textured SMPL-D parametric statistical body models, and made it available for direct use in simulations in Webots.

Animatable 3D human models based on the parametric SMPL+D body model.

Simulation is an important factor in the development of the OpenDR toolkit. It allows for the rapid setup of scenarios in which to test the algorithms being developed, to easily generate the data necessary for the training and testing of the models as well as to easily disseminate the results of this work.

Desktop rendering, old web rendering engine and current web rendering engine.

CYB extended the capabilities of the Webots simulator in a number of ways: ameliorating the rendering of simulations run on the cloud through porting its web rendering engine to the web using WebAssembly, converting its entire library to the default ROS coordinate system for better compatibility, adding functionalities needed in DL applications, and improving ROS/ROS2 compatibility, providing now 16 ROS2 packages.

Toolkit Integration for the Use Cases

Integration of toolkit functionalities in robotics platforms that will be used in the three project use cases has already started. PAL collaborated with AUTH in order to integrate tools such as face and mask detection and active or static face recognition inside the Jetson TX2 embedded in the TIAGo robot.

OpenDR toolkit code running in the Jetson TX2 embedded in TIAGo.

TAU has worked to integrate several tools developed in the OpenDR toolkit to the agile production use case, including a single demonstration grasping model and the speech recognition model. A human-robot collaboration scenario was developed that concerns the assembly of a Diesel engine, including tasks for both a robot (Franka Emika) and a human and the handling of various components.

Agile production use case scenario combining the single demonstration object grasping model and the speech recognition model.

The integration of the OpenDR tools for the agriculture use case will start at the beginning of 2022 and field tests will be conducted later in the same year. AGI has been in the process of improving and integrating the SLAM plant row guidance into the vision system of its Robotti agricultural robot and also building the intra row prototype and testing it.

Dissemination

Dissemination and communication are taken very seriously in OpenDR. Indeed, the consortium undertook numerous efforts in various directions to attract interest in the project findings and results. The project website and its social media accounts (Facebook, Twitter, LinkedIn, YouTube) receive frequent updates and posts regarding new publications, project news, forthcoming events etc. Virtually all project publications but also videos and slides from conference presentations are available on the official website.

Similar to the first year, the consortium managed to generate a high volume of publications. Fourteen papers were presented or accepted in high quality, well-established international conferences (including ICME, IROS, ICIP, EUSIPCO etc.) and twelve papers were published or accepted in scientific journals (including the highly influential IEEE TIP, IEEE TRO and IJCV). In total, project partners have published so far 50 conference and journal papers, one edited book and numerous preprints.

A particularly important dissemination activity was a workshop on "Open Deep Learning Toolkit for Robotics: Towards Democratizing Artificial Intelligence" that was organized at IROS 2021. The workshop was well attended and included talks from consortium members and prominent roboticists. Other significant dissemination activities included organization of a conference special session, co-organization of an IEEE TAI special issue on AI for robotics, invited talks, participation in events like ERF 2021 or GSoC 2021 and so on. The release of the first public version of the toolkit in M24 will signal the start of a new series of activities that will focus on attracting relevant parties in getting to know and use it.

opendr.eu | @OpenDR | OpenDR H2020 Research Project | opendr-eu/opendr | OpenDR

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under Grant agreement No *871449*.