HVD Newsletters

#1 – April 2023

PID2021-125051OB-I00 HVD (2022-2025)

Harvesting Visual Data: enabling computer vision in unfavourable data scenarios

http://www-vpu.eps.uam.es/projects/HVD/

Introducing HVD

Computer Vision (CV) is relied on for the ongoing revolution of Artificial Intelligence (AI) applications based on visual information processing. Together with the AI field, Computer Vision is currently booming, being the focus of unprecedent initiatives and funding efforts targeting to boost a transformational technology. The main reason for this explosive growth has been the development of Machine Learning schemes (mainly Deep Convolutional Neural Networks-CNNs) that are allowing to reach success rates similar to the human ones in problems requiring the analysis of visual information (e.g., object detection or semantic segmentation [1][2]). In fact, the multitude of reports (ocde.org or European Commission), that declare Artificial Intelligence (AI) as the currently most impacting General-Purpose Technology (GPT), rely on Computer Vision for the current revolution of applications based on visual information processing.

The time is now, as effective solutions based on CV are not in a visible future, but are currently a reality, and are already transforming and defining the world we live in. The deep learning (DL) revolution, unprecedented in the field of CV, is taking place both in the scientific and industrial fields, thanks to the broad and open availability of research ideas to confront many of the most demanded applications but also enabling new ones for the public good. However, the adaptation of existing ideas, canonical models, methods, and training procedures, encoded to specific domains has revealed new limitations, problems and challenges that endanger their fundamental promise of universality and, if unattended, may risk the society rights protected by the EU law.

Some of these problems are intrinsic to the use of AI and Deep CNNs. Our proposal in this project is to delve into some of the limitations of these approaches when applied to visual signals, namely, issues related to the availability of annotated data. Data is the condition that enables nowadays CV models. Despite having received substantially less attention than algorithms, methods, or models, increasing efforts are being made to provide large-scale accessible repositories, e.g., by encouraging fair data access and

data-sharing. Our main goal is to reduce the huge dependency that AI and Deep CNNs algorithms have on the availability of large-scale annotated training data.
With respect to annotated data availability, as the current state-of-the-art solutions are mainly based on supervised learning approaches, their success requires the availability of large human-annotated datasets (such as ImageNet [3]) which hinges on a large amount of expensive supervision, time, and effort. Moreover, for some sensitive domains, such as traffic collision, annotations and content are rarely available, which in practice prevents the creation of large and complete datasets. Some applications require a continuous addition of data or annotations, while conserving the previously encoded model knowledge. For example, due to the inclusion of additional tasks or additional target classes. Besides, these trained systems are usually tailored to a specific task and cannot be adapted to other tasks without re-training. To cope with these issues, a large research effort has been devoted to achieve systems that can: model the intrinsic patterns in the data without (fully) leveraging human labeling; adapt continuously the training process to the availability of new additional data; and extrapolate useful training information using complementary synthetic datasets for which an automatic annotation can be obtained. In this direction, we propose (1) to explore the capabilities of training methods using real data in absence of annotations through non-supervised and self-supervised approaches; and (2) to explore the creation and use of synthetic data to complement the training processes.

First sept-mester progress report

The provisional announcement of project approval arrived June 2022, but the official project approval was not published till November 15th, allowing to start the official opening of the project at the university. Whilst waiting for the notification of definitive approval, we have been working in the lines described in the proposal, but without being able to work with financial support.

Within WP1, Task T1.1 started evaluating the required infrastructure acquisitions and launching the purchase. Due to the delay in official approval of the project and the regulations at our university with respect to contracts during December and January, we were not able to order the new hardware till February and was received in April. We are completing the installation of one Server with 8 NVIDIA Graphical Processing Units (A40 model). This new hardware will allow training state-of-the-art algorithms and developing the tasks of the project.
Task 1.2 has just started to analyse the evaluation framework requirements for the project and compiling the different datasets to be used as well as organizing the development of new ones.

The R&D activities within WP2 are organized in three tasks:

- Task 2.1 (Real data in absence of annotations) has currently four main activity areas:
    - In the area of Semantic Segmentation, we are working in a Curriculum-based Learning (CL) strategy for a gradual introduction of semantic classes in terms of increasing difficulty, and in a novel framework for label down-sampling via soft-labelling that better conserves label information after down-sampling. It fully aligns soft-labels with image data to keep the distribution of the sampled pixels, producing reliable annotations for under-represented semantic classes, allowing training competitive models at lower resolutions.
    - To reduce the domain gap between train and test sets, we are exploring unsupervised domain adaptation, generating automatically pseudo-labels from the testing data. We are applying this method to fine-tune ReID models using pseudo-labels obtained by clustering, and for spacecraft pose estimation by model consensus.
    - We have shown that the linear probing (training a linear classifier on top of the frozen feature extractor) in order to estimate transfer performance in SSL approaches, is not very strongly correlated with the performance of the models finetuned end-to-end. The latter is often the final objective in transfer learning, and we are proposing a way to obtain a significantly better proxy task by unfreezing and jointly finetuning batch normalization layers together with the classification head.
    - We are also exploring Soft-labels and Class-weighting for unsupervised domain adaptation in different application scenarios.
- Task 2.2 (Creation and use of synthetic data) has currently three main activity areas:
    - We have developed a modified CARLA simulator designed with LiDAR semantic segmentation in mind, by adding new classes and modifying already existing objects to align better with their counterparts from real datasets such as SemanticKITTI. Using this tool, we propose SynthmanticLiDAR, a synthetic dataset for semantic segmentation on LiDAR imaging, designed to be similar to SemanticKITTI.
    - In the scope of semantic segmentation of urban scenes, we are investigating the potential of latent diffusion models, specifically using the Stable Diffusion architecture, for generating synthetic datasets to aid in the task of semantic segmentation for urban scenes.
    - We are exploring how to enhance synthetic datasets used for 3D LiDAR point clouds segmentation by incorporating precise intensity signals to the synthetically generated point clouds, with the aim of achieving more accurate and powerful models.
- Task 2.3 (Model profiling) just started February 2023 and just started February 2023 and we are currently collecting and arranging state of the art approaches on model explainabiity and identification of biasing features.

With respect to management, dissemination and use cases, the scope of WP3: T3.1 (Management) is weekly checking project progress; T3.2 (Communication) launched the project web page on September 1st, 2022, and, after the incorporation of two new Observing partners (see below), is looking for additional new Observing partners and has published this first HVD Newsletter (with one month delay); T3.3 (Dissemination) has managed to publish several papers (mostly based on preliminary tasks related to HVD activities) and is working towards "D3.1 Data Management Plan" scheduled June 2023; whilst T3.4 (Use cases and technology transfer) won't start till the second year of the project.

New Registered Observers: Deimos and Sener

Deimos joined as Registered Observer in March 2023.

Deimos is the technology branch of Elecnor Group. Elecnor is a large Spanish group specialised in integrated management and promotion of projects and infrastructure development. Founded in 1958, Elecnor is traded in Madrid Stock Exchange and currently covers a total of 12 sectors within four large business areas: Networks and Infrastructures, Telecommunications and IT, Renewable Energies, and Concessions. Deimos, which employs more than 400 people in Spain, Portugal, Romania and United Kingdom, has a proven track record in the space sector, its core business, being today one of the leading suppliers of space systems in Europe. Founded in 2001, Deimos acquired a growing specialization in the engineering studies and turnkey solutions in the following areas:

- Mission Analysis and Systems Engineering: Earth Observation, Science, Exploration, etc.
- Flight Systems and on-board software.
- Advanced systems for GNC (Guidance, navigation and control)
- GNSS (Global Navigation Satellite Systems) systems and receivers
- Space Surveillance, Space Weather & NEOs
- Ground Segment Systems
- Satellite Control Systems
- Payload Data Ground Segment
- Design of Navigation, Integrity and Performance Algorithms
- Satellite navigation-based location services

Besides attendance to HVD Industry Days and being informed of project progress, Deimos is interested in validation of our solutions within their areas of interest, and in analyzing future technology transfers to their products and research lines: Inisight4EO for video processing on board earth observation satellites, APPIDE for the automatic inspection of photovoltaic plants, and vision-based navigation technologies, scenarios where image quality represents a challenge.

Sener, through his Aerospace division, joined as Registered Observer in April 2023.

The Sener group develops innovative engineering and technology solutions in different markets, with the purpose of contributing to the progress of society in a sustainable environment, through their commitment to their customers, people and the environment. Their customers and partners recognize them for closeness, trust, professionalism, and spirit of collaboration.

In all their activity, they are distinguished by innovative work, which has managed to expand sectors and geographical areas in a continuous search for new horizons. To do this, they develop high-tech projects that respond to the needs of customers, solving the most complex challenges in the most effective way.

SENER Aerospace and Defense participates in numerous cutting-edge international projects: in space exploration, astronomy, and defense technology. A passion for precision that has led them to place SENER technology even on Mars.

Additionally to HVD Industry Days attendance and follow up of project progress, SENER Aerospace is interested in validation of our algorithms within their areas of interest, and in analyzing future technology transfers to their products and research lines for Autonomous Navigation Systems and Image-based Medical Diagnostic.

First sept-mester results

Journals

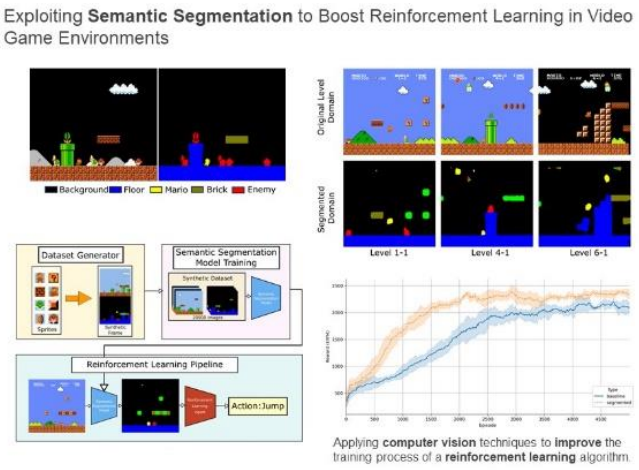
Javier Montalvo, Álvaro García-Martín, and Jesús Bescós, **Exploiting semantic segmentation to boost reinforcement learning in video game environments**, Multimedia Tools and Applications, 82, 10961-10979, March 2023. (DOI 10.1007/s11042-022-13695-1).

Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

This work is part of the preliminary tasks related to the HVD project.

In this work we explore enhancing performance of reinforcement learning algorithms in video game environments by feeding it better, more relevant data. For this purpose, we use semantic segmentation to transform the images that would be used as input for the reinforcement learning algorithm from their original domain to a simplified semantic domain with just silhouettes and class labels instead of textures and colors, and then we train the reinforcement learning algorithm with these simplified images. We have conducted different experiments to study multiple aspects: feasibility of our proposal, and potential benefits to model generalization and transfer learning. Experiments have been performed with the Super Mario Bros video game as the testing environment. Our results show multiple advantages for this method. First, it proves that using semantic segmentation enables reaching higher performance than the baseline reinforcement learning algorithm without modifying the actual algorithm,

and in fewer episodes; second, it shows noticeable performance improvements when training on multiple levels at the same time; and finally, it allows to apply transfer learning for models trained on visually different environments. We conclude that using semantic segmentation can certainly help reinforcement learning algorithms that work with visual data, by refining it. Our results also suggest that other computer vision techniques may also be beneficial for data prepoccessing.

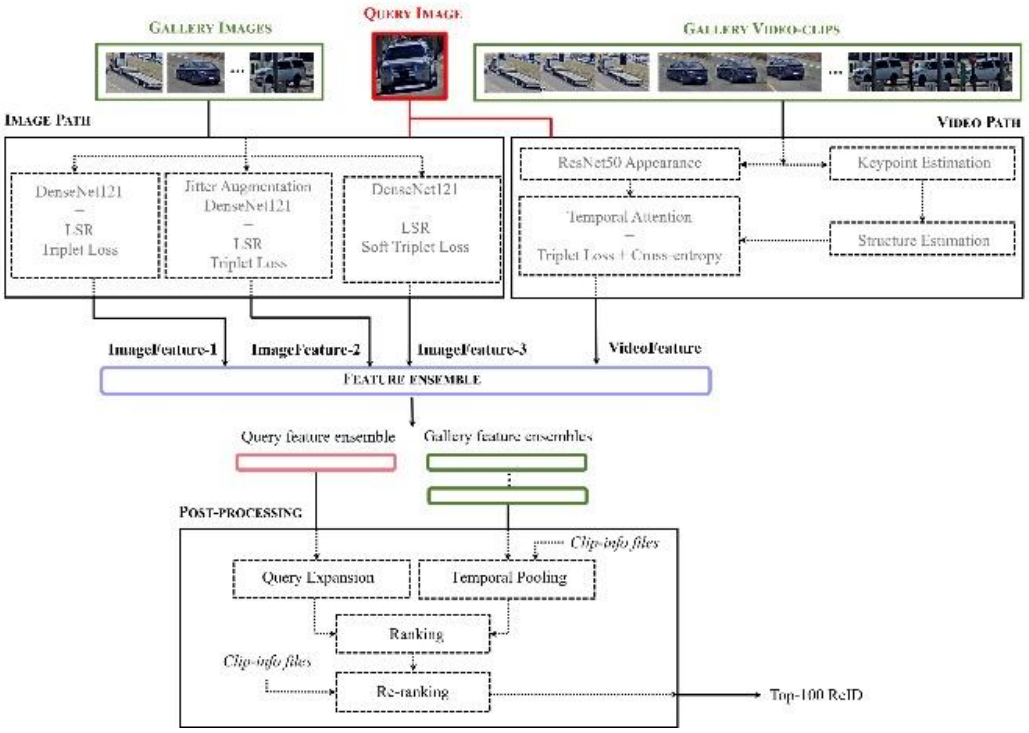


Paula Moral, Álvaro García-Martín, José M. Martínez, and Jesús Bescós, **Enhancing vehicle re-identification via synthetic training datasets and re-ranking based on video-clips information**, Multimedia Tools and Applications, Online 21 March 2023. (DOI 10.1007/s11042-023-14511-0).

Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

This work is part of the preliminary tasks related to the HVD project.

Vehicle re-identification (ReID) aims to find a specific vehicle identity across multiple non-overlapping cameras. The main challenge of this task is the large intra-class and small inter-class variability of vehicles appearance, sometimes related with large viewpoint variations, illumination changes or different camera resolutions. To tackle these problems, we proposed a vehicle ReID system based on ensembling deep learning features and adding different post-processing techniques. In this paper, we improve that proposal by: incorporating large-scale synthetic datasets in the training step; performing an exhaustive ablation study showing and analyzing the influence of synthetic content in ReID datasets, in particular CityFlow-ReID and VeRi-776; and extending post-processing by including different approaches to the use of gallery video-clips of the target vehicles in the re-ranking step. Additionally, we present an evaluation framework in order to evaluate CityFlow-ReID: as this dataset has not public ground truth annotations, AI City Challenge provided an on-line evaluation service which is no more available; our evaluation framework (see below) allows researchers to keep on evaluating the performance of their systems in the CityFlow-ReID dataset.

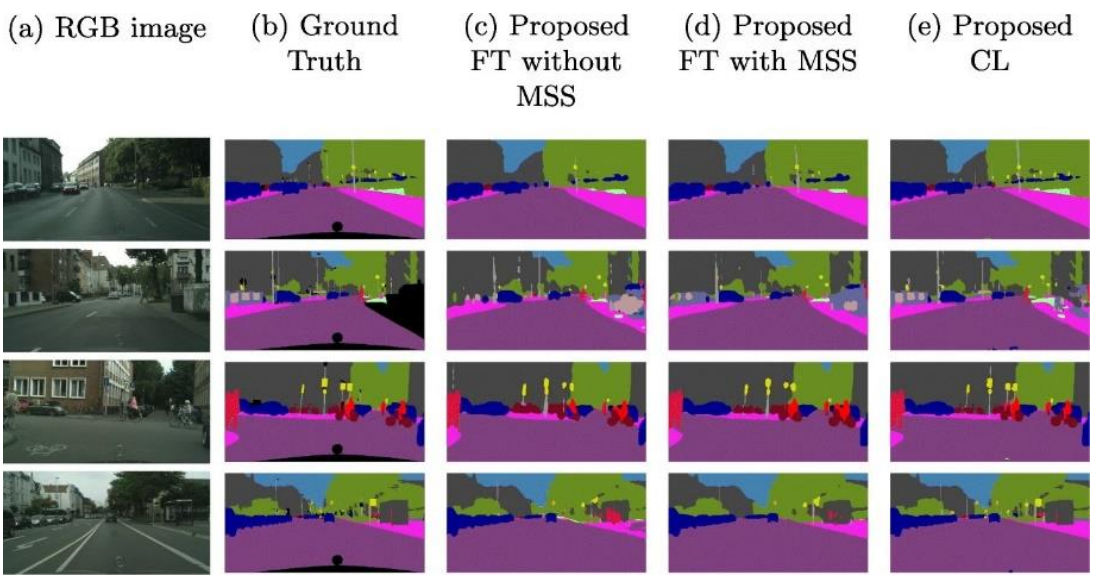


Roberto Alcover-Couso, Juan Carlos SanMiguel, Marcos Escudero-Viñolo, and Álvaro García-Martín, **On exploring weakly supervised domain adaptation strategies for semantic segmentation using synthetic data**, Multimedia Tools and Applications, Online 11 March 2023. (DOI 10.1007/s11042-023-14662-0).

Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

This work is part of the preliminary tasks related to the HVD and SEGA-CV projects.

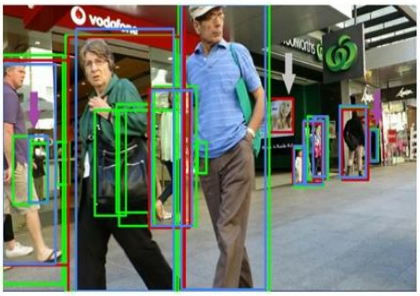
Pixel-wise image segmentation is key for many Computer Vision applications. The training of deep neural networks for this task has expensive pixel-level annotation requirements, thus, motivating a growing interest on synthetic data to provide unlimited data and its annotations. In this paper, we focus on the generation and application of synthetic data as representative training corpuses for semantic segmentation of urban scenes. First, we propose a synthetic data generation protocol, which identifies key features affecting performance and provides datasets with variable complexity. Second, we adapt two popular weakly supervised domain adaptation approaches (combined training, fine-tuning) to employ synthetic and real data. Moreover, we analyze several backbone models, real/synthetic datasets and their proportions when combined. Third, we propose a new curriculum learning strategy to employ several synthetic and real datasets. Our major findings suggest the high-performance impact of pace and order of synthetic and real data presentation, achieving state of the art results for well-known models. The results by training with the proposed dataset outperform popular alternatives, thus demonstrating the effectiveness of the proposed protocol.

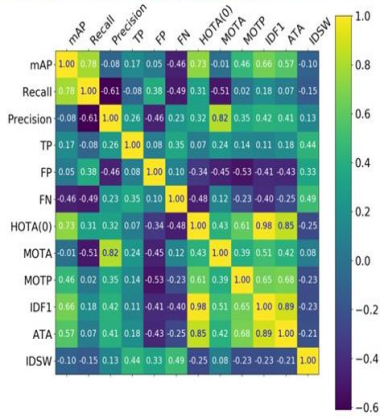


Conferences

Juan Carlos SanMiguel, Jorge Muñoz, Fabio Poiesi, "Detection-aware multi-object tracking evaluation", in Proc. of the 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Madrid, Spain, 2022, pp. 1-8, doi: 10.1109/AVSS56176.2022.9959412

Abstract: How would you fairly evaluate two multi-object tracking algorithms (i.e. trackers), each one employing a different object detector? Detectors keep improving, thus trackers can make less effort to estimate object states over time. Is it then fair to compare a new tracker employing a new detector with another tracker using an old detector? In this paper, we propose a novel performance measure, named Tracking Effort Measure (TEM), to evaluate trackers that use different detectors. TEM estimates the improvement that the tracker does with respect to its input data (i.e. detections) at frame level (intra-frame complexity) and sequence level (inter-frame complexity). We evaluate TEM over well-known datasets, four trackers and eight detection sets. Results show that, unlike conventional tracking evaluation measures, TEM can quantify the effort done by the tracker with a reduced





correlation on the input detections. Its implementation will be made publicly available online.

Software

Experimental Framework for MOT

https://github.com/vpulab/MOT-evaluation

This repository is an experimental framework for Multi Object Tracking analysis. Multiple detectors, trackers, and evaluation metrics can be found, and, also, documentation allowing to implement further models in the framework.
Software related to Detection-aware multi-object tracking evaluation, AVSS 2022

Exploiting Semantic Segmentation To Boost Reinforcement Learning In Video Game Environments

https://github.com/vpulab/Semantic-Segmentation-Boost-Reinforcement-Learning

Pytorch implementation of our paper Exploiting semantic segmentation to boost reinforcement learning in video game environments, MTAP 2023

VPULab-CityFlow-ReID-Evaluation-Framework

https://github.com/vpulab/VPULab-CityFlow-ReID-Evaluation-Framework

This Evaluation Framework allows researchers to keep on evaluating the performance of their approaches and the ones from the State-of-the-Art over the CitiFlow-ReID dataset. As this dataset has not public ground truth annotations, AI City Challenge provided an on-line evaluation server during the challenge, which is no more available.
Software related to Enhancing vehicle re-identification via synthetic training datasets and re-ranking based on video-clips information, MTAP 2023

On exploring weakly supervised domain adaptation strategies for semantic segmentation using synthetic data

https://github.com/vpulab/WSDA_semantic/

Pytorch implementation of our paper On exploring weakly supervised domain adaptation strategies for semantic segmentation using synthetic data, MTAP 2023.

Graduate thesis

Clasificación auto-supervisada de imágenes mediante aprendizaje contrastivo (**Auto-supervised image classficiation based on contrastive learning**), Alejandro

Guzmán Huerta (advisor: Miguel Ángel García García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería Informática, Univ. Autónoma de Madrid, Feb. 2023.

Abstract: In this 2022 the word of the year has been "artificial intelligence", according to FundéuRAE and is that increasingly, this technology is being used for the development of various applications, causing curiosity and general interest of the population. It is noted that in recent years, this and many other technologies have advanced by leaps and bounds since 1950 by the hand of Alan Turing, the father of computer science and artificial intelligence (AI).

The use of artificial intelligence (AI) can currently be found in different fields. It is present in the health area, with systems that allow us to identify tumors, diseases or various pathologies. Weather forecasts for the detection of temperatures, rainfall, etc. For this reason, very specific professionals are needed in each specific branch. These professionals will carry out the classification of the data, which will then be introduced into the artificial intelligence so that it learns to differentiate the different patterns that may exist. Similarly, there is a set of models based on self-supervised learning, which consists of obtaining unlabeled data and using it to train these models to train themselves.

That is why in this TFG we research the basis of self-supervised learning, specifically the SimClr model, comparing the performance with different algorithms (AlexNet, ResNet, VGG, MobileNet, SqueezeNet). All these, based on convolutional neural networks, using two databases (Imagewoof and Imagenette). These databases are extracted from ImageNet which, as we will see in this paper, is a very popular and influential database, which is composed of different types of images. To develop it, Pytorch has been used as a Machine Learning library. Nowadays it is widely used to develop projects on neural networks, it gives the possibility of using GPUs instead of CPUs so that the training is done in a more agile way. On the other hand, it has different ecosystems that are tools or frameworks, in this case Pytorch-lightning has been used allowing to build models in a more dynamic way. Finally, Neptune.ia has been used to compare, record, and organize the results obtained.

The HVD (PID2021-125051OB-I00) project is supported by

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN