HVD Newsletters

#2 – September 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/

Second semester progress report

The project has been running properly during this semester, with some minor adjustments to the workplan.

Within WP1, Task T1.1 has completed the acquisition and installation of new hardware, fulfilling M1.1 "Acquisition of required hardware for WP3 and WP3": one Server with 8 NVIDIA Graphical Processing Units (A40 model). D1.1 "System Infrastructure" has been published with all the details with respect to the current system infrastructure available for the project.
Task 1.2 has delayed the publication of D1.2 "Evaluation frameworks: datasets, metrics and benchmarks" to January 2024 in order to incorporate the datasets created within the project fulfilling also the Data Management Plan. Therefore, M1.2 "Generation of datasets for WP2" has also been delayed.

The R&D activities within WP2 have advanced properly in their three tasks and has started to organize D2 "Technologies for low availability of annotated data", due January 2024 (as well as milestones M2.1 "First version of non-supervised and self-supervised algorithms"; M2.2 "First version of continual lifelong learning algorithms"; M2.3 "First version of creation and use of synthetic datasets"; and M2.4 "First version of model profiling").

Task 2.1 (Real data in absence of annotations) has produced several advances. From one side, in the development and evaluation of supervised and unsupervised algorithms. Also, auto-supervised and incremental learning methods have been proposed.

Task 2.2 (Creation and use of synthetic data) has developed simulators for training autonomous driving vehicles in urban environments. Additionally, it has studied the explainability of generative algorithms based on diffusion for the creation of synthetic datasets.

Task 2.3 (Model profiling) has studied the explainability of image classification analyzing the decision margin.

Within WP3·"Management, dissemination and use cases", deliverables D3.1v1 "Data Management Plan" and D3.2v1 "Results Report" have been published September 2023, after being delayed from their original publication schedule (June 2023). D3.1 will be updated with each new dataset published.
T3.1 (Management) keeps on weekly checking project progress; T3.2 (Communication) updates regularly the project web page, is looking for additional new Observing partners and has published this second HVD Newsletter; T3.3 (Dissemination) has produced several papers that are under review; whilst T3.4 (Use cases and technology transfer) is just starting during this month of September 2023.

New Registered Observers: Servicio de Radiodiagnóstico - Hospital Universitario La Princesa/Facultad de Medicina UAM

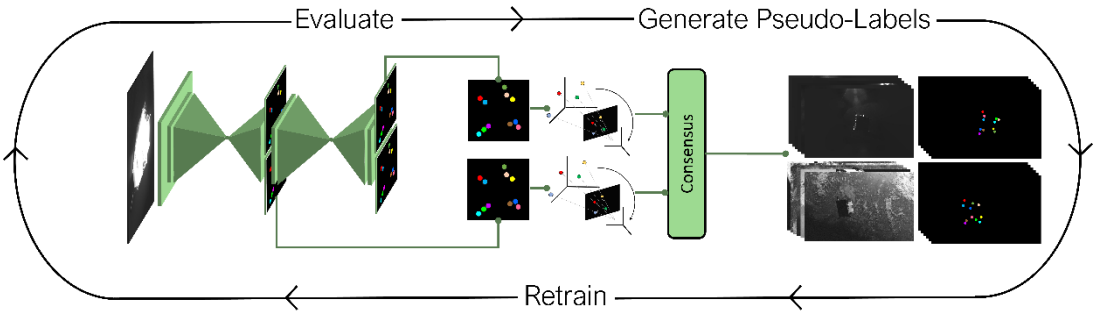
Servicio de Radiodiagnóstico del Hospital Universitario La Princesa (Radiodiagnosis Service of La Princesa University Hospital) joined as Registered Observer in July 2023.
The University Hospital of La Princesa combines its 170 years of history with the modernity of a centre of the XXI century in which it offers assistance of the highest quality, MIR teaching and research at the highest level with a great impact factor that places it among the best Spanish hospitals, all without losing its commitment to its raison d'être; the patients you serve.
The Radiodiagnosis service has long been characterized by being a service specially integrated with the rest of the Hospital, under a modern concept of Clinical Radiology, pioneer in the organization by "Organ and System" that has always allowed a correct integration with the rest of the Hospital, allowing a high-level specialization. A Diagnostic Imaging Service involves technical support of the different imaging methods, knowledge of the most appropriate test for each occasion, the analysis of the findings in these tests and the correct interpretation of them. In recent years the specialty has become a vital part of medicine, at the centre of the diagnosis of most patients. The variety of techniques that are handled and the complexity of many of them, in many cases specific to different pathologies, has demonstrated the convenience of an organization in organs and systems, which brings a greater benefit to the different medical surgical specialties.

Second semester results

Journals

Juan Ignacio Bravo Pérez-Villar, Álvaro García-Martín, Jesús Bescós, Marcos Escudero-Viñolo, **Spacecraft Pose Estimation: Robust 2D and 3D-Structural Losses and Unsupervised Domain Adaptation by Inter-Model Consensus**, IEEE Transactions on Aerospace and Electronic Systems, Online 21 August 2023. (DOI 10.1109/TAES.2023.3306731).

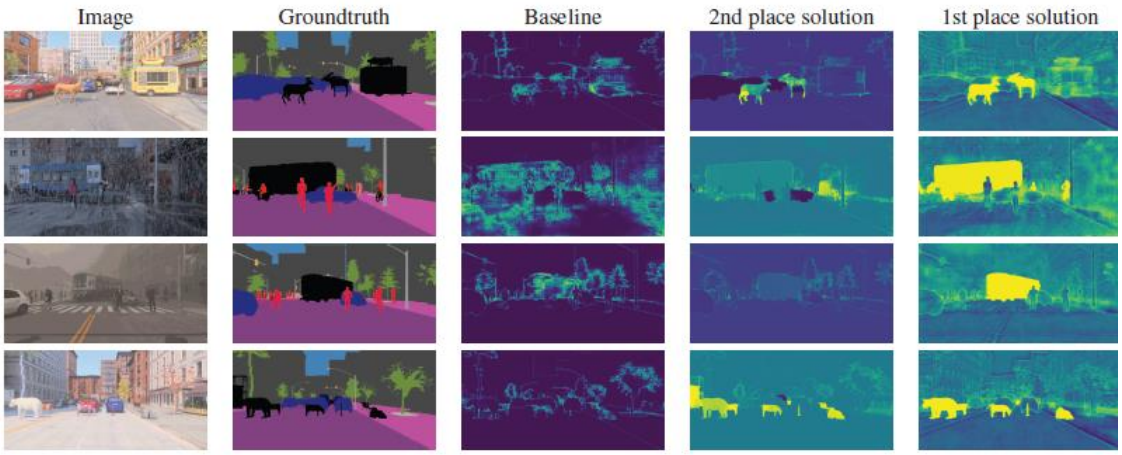
The accurate estimation of spacecraft pose is crucial for missions involving the navigation of two spacecraft in close proximity. Supervised algorithms are currently the state-of-the-art approach for spacecraft pose estimation. However, the absence of training data acquired in operational scenarios poses a challenge for the supervised algorithms. To address this issue, computer-aided simulators have been introduced to solve the issue of data availability but introducing a large gap between the training domain and test domain. We here describe an algorithm for unsupervised domain adaptation with robust pseudo-labelling by model consensus. Moreover, the proposed method incorporates 3D structure into the spacecraft pose estimation pipeline to provide robustness against high illumination shifts between domains. Our solution has ranked second in the two categories of the 2021 Pose Estimation Challenge (SPEC2021) organised by the European Space Agency and the Stanford University, achieving the lowest average error over these two categories. Our solution is available at: https://github.com/JotaBravo/spacecraft-uda



Conferences

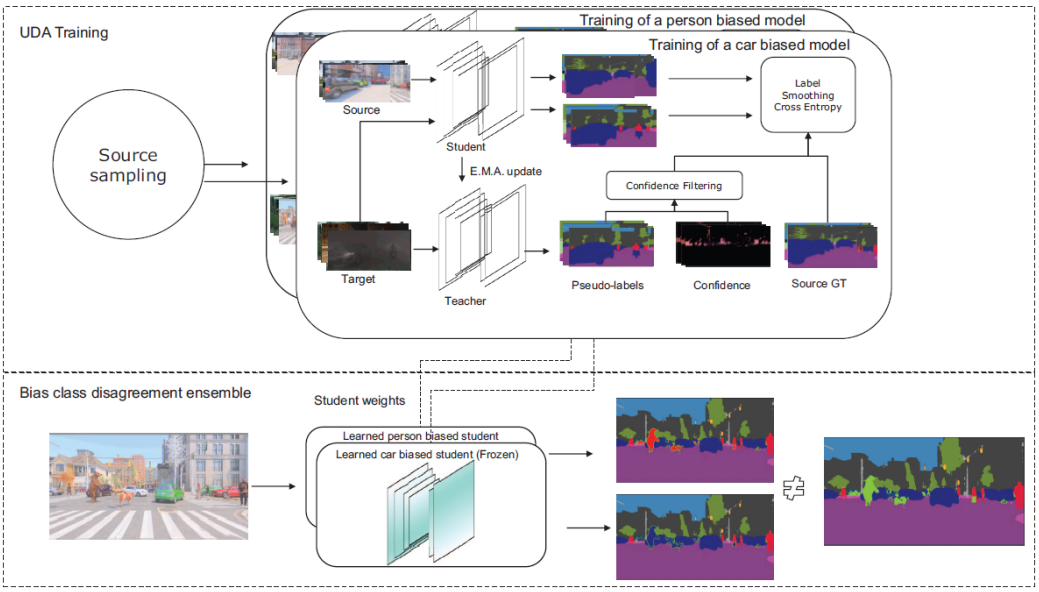
Xuanling Yu et al., "The Robust Semantic Segmentation UNCV2023 Challenge Results", IEEE International Conference on Computer Vision Workshops 2023 (ICCV), Workshop on Uncertainty Quantification for Computer Vision, Paris (France), Oct. 2023, pp. 4618-4628.

Abstract: This paper outlines the winning solutions employed in addressing the MUAD uncertainty quantification challenge held at ICCV 2023. The challenge was centered around semantic segmentation in urban environments, with a particular focus on natural adversarial scenarios. The report presents the results of 19 submitted entries, with numerous techniques drawing inspiration from cutting-edge uncertainty quantification methodologies presented at prominent conferences in the fields of computer vision and machine learning and journals over the past few years. Within this document, the challenge is introduced, shedding light on its purpose and objectives, which primarily revolved around enhancing the robustness of semantic segmentation in urban scenes under varying natural adversarial conditions. The report then delves into the top-performing solutions. Moreover, the document aims to provide a comprehensive overview of the diverse solutions deployed by all participants. By doing so, it seeks to offer readers a deeper insight into the array of strategies that can be leveraged to effectively handle the inherent uncertainties associated with autonomous driving and semantic segmentation, specifically within urban environments.



Roberto Alcover, Juan C. SanMiguel, Marcos Escudero, "Biased Class disagreement: detection of out of distribution instances by using differently biased semantic segmentation models", IEEE International Conference on Computer Vision Workshops 2023 (ICCV), Workshop on Uncertainty Quantification for Computer Vision, Paris (France), Oct. 2023, pp. 4580-4588.

Abstract: Autonomous driving heavily relies on accurate under-standing of the surrounding environment, which is facilitated by semantic segmentation models that classify each pixel in an image. However, training these computer vision models using available datasets often fails to capture the diverse conditions and objects that can be encountered during a trip. Adverse weather conditions and the presence of Out-of-Distribution (OOD) instances, such as wild animals and debris, are common challenges in autonomous driving. Unfortunately, current models struggle to perform well in unseen conditions.

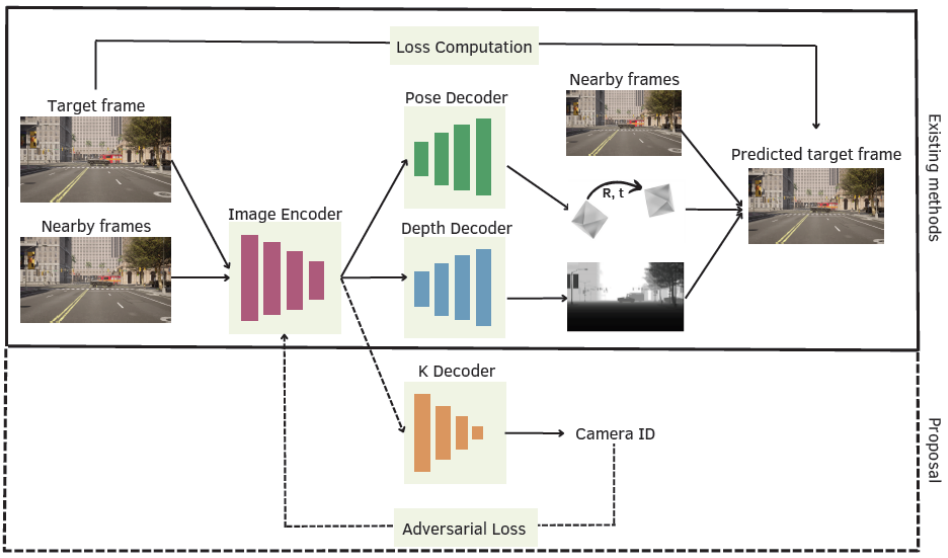


To address these limitations, this paper proposes a comprehensive approach that integrates uncertainty quantification and bias reinforcing within the framework of Unsupervised Domain Adaptation (UDA). Our approach leverages multiple models with diverse biases, aiming to assign high-confidence predictions to OOD instances by mapping them to the selected prior semantic category. Extensive evaluations on the MUAD dataset demonstrate the effectiveness of our approach in improving performance and robustness against OOD instances. Notably, our approach achieves outstanding results, securing the first position in the MUAD challenge.

Cecilia Diana Albelda, Juan Ignacio Bravo Pérez-Villar, Javier Montalvo, Álvaro García Martín, Jesús Bescós Cano, "Self-Supervised Monocular Depth Estimation on Unseen Synthetic Cameras", 26th Iberoamerican Congress on Pattern Recognition (CIAPR), Coimbra (Portugal), Nov. 2023.

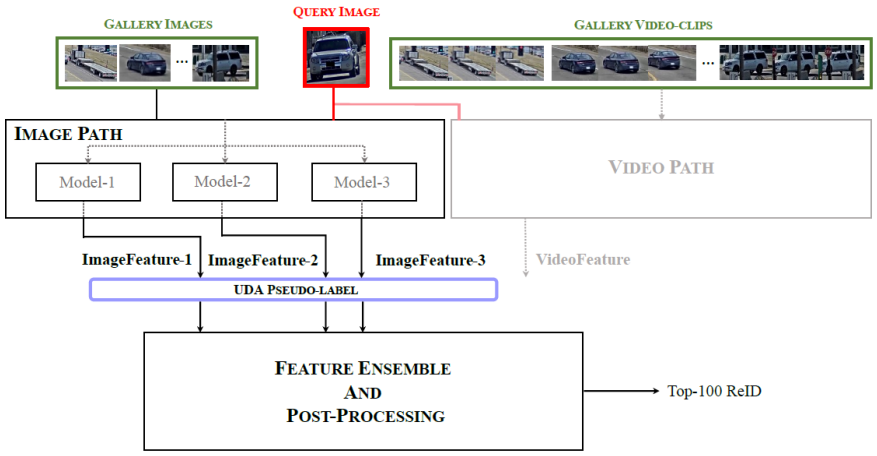
Abstract: Monocular depth estimation is a critical task in computer vision, and self-supervised deep learning methods have achieved remarkable results in recent years. However, these models often struggle on camera generalization, i.e. at sequences captured by unseen cameras. To address this challenge, we

present a new public custom dataset created using the CARLA simulator, consisting of three video sequences recorded by five different cameras with varying focal distances. This dataset has been created due to the absence of public datasets containing identical sequences captured by different cameras. Additionally, it is proposed in this paper the use of adversarial training to improve the models' robustness to intrinsic camera parameter changes, enabling accurate depth estimation regardless of the recording camera. The results of our proposed architecture are compared with a baseline model, hence being evaluated the effectiveness of adversarial training and demonstrating its potential benefits both on our synthetic dataset and on the KITTI benchmark as the reference dataset to evaluate depth estimation.



Paula Moral, Alvaro Garcia-Martin, Jose M. Martinez, "Vehicle Re-Identification based on unsupervised domain adaptation by incremental generation of Pseudo-labels", 26th Iberoamerican Congress on Pattern Recognition (CIAPR), Coimbra (Portugal), Nov. 2023.

Abstract: The main goal of vehicle re-identification (ReID) is to associate the same vehicle identity in different cameras. This is a challenging task due to variations in light, viewpoints or occlusions; in particular, vehicles present a large intra-class variability and a small inter-class variability. In ReID, the samples in the test sets belong to identities that have not been seen during training. To reduce the domain gap between train and test sets, this work explores unsupervised domain adaptation (UDA) generating automatically pseudo-labels from the testing data, which are used to fine-tune the ReID models. Specifically, the pseudo-labels are obtained by clustering using different hyperparameters and incrementally due to retraining the model a number of times per hyperparameter with the generated pseudo-labels. The ReID system is evaluated in CityFlow ReID-v2 dataset.



Software

Spacecraft Pose Estimation: Robust 2D and 3D-Structural Losses and Unsupervised Domain Adaptation by Inter-Model Consensus

https://github.com/JotaBravo/spacecraft-uda

Implementation of our paper Spacecraft Pose Estimation: Robust 2D and 3D-Structural Losses and Unsupervised Domain Adaptation by Inter-Model Consensus, IEEE Trans on AES, 2023.

Master thesis

Self-supervised monocular Depth estimation and visual odometry on unseen synthetic cameras, Cecilia Diana Albelda (advisor: Juan Ignacio Braco Pérez-Villas; lecturer: Álvaro García-Martín), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Visual Signal Processing, Univ. Autónoma de Madrid, Jun. 2023.

Abstract: Visual odometry and monocular depth estimation models have been taking more presence in the computer vision field recently. There are many self-supervised deep learning methods devoted to these tasks that achieve state-of-the-art results. However, they tend to underperform when it comes to generalize to sequences taken by new cameras.

On the one hand, one of the major challenges for unseen camera generalization is the absence of a suitable dataset that enables fair comparisons between identical sequences taken by different cameras. To address this problem, we have created a custom dataset using the CARLA simulator [1]. This dataset comprises three different video sequences, each one captured by 5 specific cameras of diverse focal distances.

On the other hand, we propose an approach that makes use of adversarial training in order to make the model invariant to changes within the internal camera parameters. Hence, we enable depth and pose estimation independently from the camera used to record the sequence.

The results obtained using the proposed architecture are compared to those achieved by a baseline one. Moreover, it has also been performed an evaluation process of the adversarial learning effect, therefore demonstrating the potential gains of our approach.

Specifically, this project explores the improvement of actual self-supervised monocular depth estimation and visual odometry models by increasing their robustness under changes in the camera calibration parameters through adversarial training.

Deep Learning modelling od intensity signals for synthetic Lidar data generation, Andrés Gómez-Caraballo Yélamos (advisor: Javier Montalvo Rodrigo; lecturer: Pablo Carballeria López), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Visual Signal Processing, Univ. Autónoma de Madrid, Jun. 2023.

Abstract: Autonomous driving field has gradually become a central field of research in computer vision and deep learning. Technologies such as LiDAR have an essential role in that type of systems, since they enable deep understanding of the vehicle surroundings. LiDAR data must be processed by semantic segmentation systems which allow for a detailed interpretation of the environment through advanced object detection and classification.

However, training high-performing semantic segmentation systems requires large amounts of labeled data, which are hard to obtain. In this context, synthetic data generation has emerged as a promising solution. Nevertheless, this approach presents a major issue, since synthetic datasets are not capable of generating data that are identical to real ones, as they may differ in appearance or fail to provide certain types of information, such as remissions, which have been proved to positively impact the performance of semantic segmentation systems.

Therefore, this Master's Thesis explores a deep learning approach to obtain accurate 3D point cloud intensity information. For this purpose, we propose a network architecture based on SqueezeSegV3 [1], which is capable of predicting remission values for a real-world 3D point cloud. To achieve this objective, we introduce suitable evaluation metrics and visualization tools for analyzing the performance of the developed models.

Throughout this project, we have conducted a series of experiments to determine the optimal architecture and configuration to achieve the proposed objective. The results obtained in these experiments have been analyzed and compared

with each other to draw the final conclusions. These results allow us to affirm that it is possible to train a model for predicting remissions. Likewise, we have observed a slight general improvement in models' performance when incorporating class semantic information, whose potential impact can be studied in depth in the future.

Synthetic Data Generation uUsing Latent Diffussion Models for Semantic Segmentation of Urban Scenes, Pablo Marcos Manchón (advisor: Juan Carlos San Miguel Avedillo), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Visual Signal Processing, Univ. Autónoma de Madrid, Jun. 2023.

Abstract: This master's thesis investigates the use of text-to-image Latent Diffusion Models (LDM) for generating synthetic datasets in semantic segmentation tasks. Specifically, it focuses on their application in urban scenarios, where the scarcity of annotated data motivates the use of synthetic data.

The research centers around Diffusion Attentive Attribution Maps (DAAM), an existing explainability method used to attribute the influence of each part of a text prompt to regions in a generated image produced by an LDM.

Two extensions of DAAM are proposed. Firstly, "Open Vocabulary DAAM" is introduced, enabling the construction of attribution maps for arbitrary texts, regardless of whether they were used as prompts for generating the synthetic images. Secondly, "Linear DAAM" is presented as a simplified version that facilitates the generation of attribution maps for individual words. These modifications facilitate the use of this method for object segmentation based on semantic descriptions.

To address the challenge of selecting the most appropriate word to semantically describe an object, an optimization strategy in the text-embedding space is proposed. This approach aims to identify the most accurate words for describing target regions, thereby enhancing the precision of segmentation masks.

To validate the proposed methodology, a series of experiments were conducted on a dataset generated using Stable Diffusion. The results confirm the effectiveness of optimized tokens in segmenting objects across diverse images, thereby emphasizing the valuable semantic information contained within these tokens.

This work contributes to the research problem in two main aspects. Firstly, it deepens the explainability of LDMs through the development of "Open Vocabulary DAAM," a tool with the potential to analyze learned semantic relationships, potential biases, and synthesis mechanisms. Secondly, it advances research on Open Vocabulary-based segmentation models by proposing a

strategy for searching descriptive words for an object, resulting in improved segmentation masks without the need for model retraining.

Although these findings are preliminary, they strongly highlight the potential of attention maps in object segmentation. Moreover, they provide a solid foundation for future research in this field.

Design and Evaluation of a Dataset for Detection and Geopositioning of Urban Elements, Juan José Löpez Castilla (advisor: Álvaro García Martín), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Visual Signal Processing, Univ. Autónoma de Madrid, Sept. 2023.

Abstract: The proliferation of deep learning has led to remarkable advancements in computer vision, particularly in the field of object detection. This could be used for key urban infrastructure tasks like urban planning or infrastructure monitoring. The thesis addresses the critical need for curated and comprehensive geopositioned datasets tailored to urban elements, in order to enhance the accuracy and applicability of object detection models in urban contexts.

The research centers on the design and generation of a geopositioned image dataset comprising diverse urban elements. The dataset covers an array of objects commonly found in urban landscapes, including street signs, waste containers, and pedestrian crossings. It includes variation in lighting conditions, weather, and occlusions, which are characteristic of urban settings. The dataset is annotated with bounding boxes to facilitate the training and evaluation of deep learning models.

To prove the dataset's utility, a comparative evaluation of object detection models is conducted. State-of-the-art deep learning architectures for real time object detection, such as YOLOv5 and YOLOv7, are trained and fine-tuned using the urban dataset. The performance metrics are analyzed in order to draw conclusions from the models and the dataset. The results demonstrate the capability to accurately detect the urban elements included, but also showcases the areas of improvement, like making that detection more specific.

By addressing the scarcity of geopositioned urban datasets, this thesis contributes to the advancement of object detection capabilities in complex urban environments. The methodologies and insights presented herein offer resources for researchers and anyone aiming to develop deep learning models for a wide range of urban applications.

Graduate thesis

Clasificación auto-supervisada de imágenes mediante aprendizaje no-contrastivo (**Auto-supervised image classficiation based on non-contrastive learning**), Franciso Lerma Martínez (advisor: Miguel Ángel García García), Trabajo

Fin de Grado (Graduate Thesis), Grado en Ingeniería Informática, Univ. Autónoma de Madrid, May 2023.

Abstract: Nowadays, machine learning is an essential tool in the performance of many tasks such as artificial vision, voice assistants or autonomous driving. This is due to the great technological development in hardware in the last decade and the proliferation of data thanks to the Internet. However, most of this data is not labelled, making supervised learning not suitable. Labelling data is a manual and costly task that requires a lot of time and, sometimes, professional knowledge. Hence, self-supervised learning arises, which allows training feature extractors capable to learn useful and robust representations of unlabelled data. These self-supervised feature extractors can be included in other neural networks thanks to transfer learning, facilitating the performance of other supervised tasks and reducing the amount of labelled data required.

In this Graduate thesis, we study supervised learning using convolutional neural networks and we study noncontrastive self-supervised learning based on simple Siamese networks, known as SimSiam. This method maximizes the similarity between two positive augmentations of one unlabelled image, without collapsing to constant or trivial outputs.

The main objective of this paper is to implement SimSiam in Pytorch and to pre-train the feature extractors of the convolutional neural networks ResNet, VGG, AlexNet, SqueezeNet y MobileNet with this method. Subsequently, the results obtained from the supervised classification with these convolutional neural networks will be compared with the results obtained from the self-supervised classifiers, built on the feature extractors pre-trained with SimSiam. Experimental tests will be performed on the Imagenette and ImageWoof datasets, which are subsets of 10 ImageNet classes. Finally, we show the results and conclusions obtained from the experimental tests for each convolutional neural network and which are more suitable for this type of self-supervised learning.

Clasificación auto-supervisada de imágenes mediante contrste de momento (**Auto-supervised image classficiation based on momentum contrast**), Víctor Sánchez de la Roda Núñéz (advisor: Miguel Ángel García García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería Informática (Programa de Doble Grado Ing. Informática y Matemáticas), Univ. Autónoma de Madrid, Jul. 2023.

Abstract: The year 2023 is shaping up to be the one introducing Artificial Intelligence, or more commonly called AI, into everyday life. We can't stop hearing this term in the media and on social networks. Chat-GPT is probably the main culprit, a chatbot trained through both supervised and reinforcement learning.

Throughout this work, we will delve into another area of AI, namely deep learning. This discipline dates back to the 1950s, [1], with the perceptron. A

simple model that led to the development of a technique that replicated human neuronal behavior, hence the models were named neural networks. Specifically, we will focus on the use of Convolutional Neural Networks, applied to image classification. The objective of these is to develop sets of common features among images of the same class. In the context of machine learning, a class is one of the n sets that will make up the scope of the experiment. For example, we could take tennis balls, basketballs, and soccer balls as classes. We would be facing a problem with 3 classes.

Throughout the work, we will review the development of deep learning, show how we can measure the performance of the networks in the task they are being trained on, and explore a new methodology, called Momentum Contrast, recently developed to train these networks in a self-supervised manner, that is, without the need to tell the network which predefined class in our problem the element it has just seen belongs to.

Contribuciones a la simulación de sistemas multi-cámara basado en UNITY (**Contributions to UNITY-based multi-camera system simulation**), Miguel Bellido González del Alba (advisor: Juan C. SanMiguel), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2023.

Abstract: In this Bachelor's Thesis, our main objective is to study semantic segmentation tasks, with a primary focus on the domain gap and how it impacts the usage of different datasets in our experiments. Firstly, at the state-of-art section, we will study the various characteristics of a neural network and aim to develop a comprehensive understanding of semantic segmentation tasks. including all necessary properties for the successful implementation of our system. Next, we will design a system capable of achieving the objectives set forth in this work. We will develop a program using the DeeplabV3+ architecture, which is known to be suitable for our specific task. We will use two different backbones and work with three dataset to conduct a thorough analysis of the domain gap. We will define and conduct several experiments, and subsequently present the results obtained from these experiments. There will be three experiments in total. The first experiment will involve an initial implementation with each domain treated independently, meaning training and testing with the same dataset. The second experiment will analyze the effect of the domain gap by training with one dataset and testing with a different one. The third experiment will focus on analyzing the impact of introducing a higher or lower percentage of real data during the training phase. Finally, we will draw conclusions regarding the work conducted and explore potential solutions or suggestions for future research and improvements.

Interpretabilidad de Redes Neuronales de clasificación de imágenes mediante el cálculo de sus márgenes (**Auto-supervised image classficiation based on momentum contrast**), Anisha Anil Sujanani Sujanani (advisor: Kirill Sirotkin; lecturer: Marcos Escudero Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Sept. 2023.

Abstract: In this bachelor's Thesis, the interpretation of using neural networks for image classification through margin calculation will be analysed. The margin in the context of machine learning refers to the distance or separation between the classification decisions of a model and the decision boundary. It is a measure of the model's confidence or certainty in its predictions. In particular, the margins of different models trained through Self-Supervised learning will be analysed, which is a learning approach that does not require labels as they are automatically generated by defining a pretext task. Analysing the margin distribution can provide information about the quality and performance of the model, as well as possible areas of improvement in terms of accuracy and confidence in the classifications. The relationship between these margins and the classification effectiveness will be studied, as well as a deeper exploration of the different techniques applied in Self-Supervised learning.

The set of points that divides the input space (in image classification, the space defined by all the images) into labelled regions is known as the decision space of a classifier, and the boundaries between regions are referred to as decision boundaries. The margin for each class can be understood as the distance from the samples of each class to their decision boundaries. Describing how a classifier creates such boundaries and studying the margins of each class is crucial for interpretability.

Currently, deep learning is part of many projects, demonstrating the existing potential in this field. Deep learning techniques are characterized using large amounts of training data. For example, various studies in the healthcare sector are determined using neural networks. However, obtaining such data is not always easy, and sometimes it may be insufficient, leading to deviations from the expected results. In these situations, it is advisable to resort to alternatives, particularly the use of Self-Supervised learning techniques.

In this work we will analyze models trained by means of self-supervised learning techniques. In particular, the decision frontiers learned by these models, when adapted to an image classification task, will be studied. The obtained results confirm that the self-supervised technique used has an impact on the learned decision frontiers. Moreover, the experimental results suggest that the margin can be a complementary measure to the prediction confidence to aid model interpretability.

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

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN