HVD Newsletters

#3 – March 2024

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

Harvesting Visual Data: enabling computer vision in unfavourable data scenarios

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

Third semester progress report

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

Task 1.2 has delayed again the publication of D1.2 "Evaluation frameworks: datasets, metrics and benchmarks" to June 2024 in order to incorporate the datasets related to several publications under review.

The R&D activities within WP2 have advanced properly in their three tasks. D2 "Technologies for low availability of annotated data", due January 2024 has been delayed to June 2024 also due to the different publications under review..

Task 2.1 (Real data in absence of annotations) has produced several advances in the development and evaluation of supervised, unsupervised, auto-supervised, incremental learning, fusion of models and foundational models adaptation methods.

Task 2.2 (Creation and use of synthetic data) has developed simulators for training autonomous driving vehicles in urban environments and pose estimation of satellites in open space. It has created synthetic datasets using generative algorithms based on diffusion.

Task 2.3 (Model profiling) has studied the explainability of image classification analyzing the decision margin, analyzed the results of different auto-supervised models, analyzed models in the domain of tumors, detection of out-of-training distributions, identified social biases in discriminative visual models and studied the AIAct.

Within WP3 "Management, dissemination and use cases", deliverable D3.2v2 "Results Report" has been published March 2024, after being delayed from their original publication schedule (June 2023). T3.1 (Management) keeps on weekly checking project progress; T3.2 (Communication) updates regularly the project web page, and has

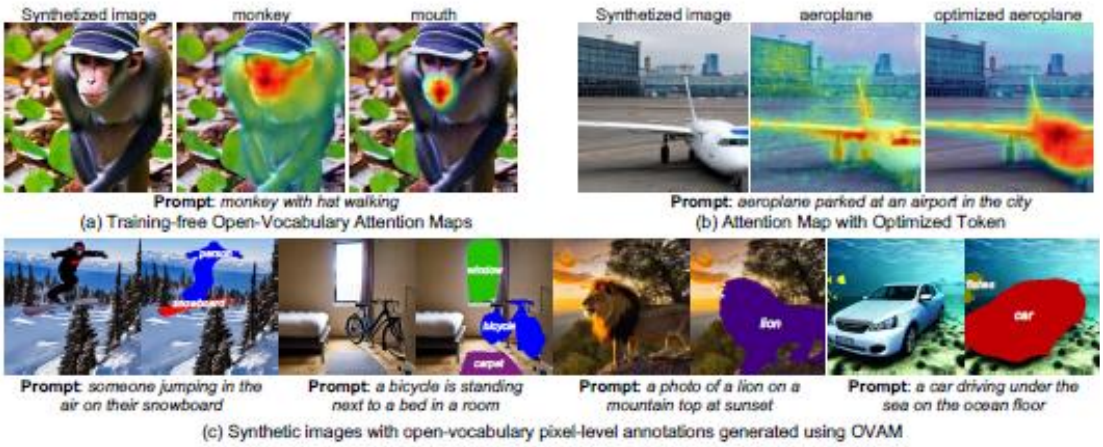
published this third HVD Newsletter; T3.3 (Dissemination) has produced several papers that are under review; whilst within T3.4 (Use cases and technology transfer) two project proposals have been submitted (one with SENER and another with BDEO).

Third semester results

Conferences

Pablo Marcos-Manchón, Roberto Alcover-Couso, Juan Carlos SanMiguel, José M. Martínez, "Open-Vocabulary Attention Maps with Token Optimization for Semantic Segmentation in Diffusion Models", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle (USA), Jun. 2024. (Arxiv https://arxiv.org/abs/2403.14291)

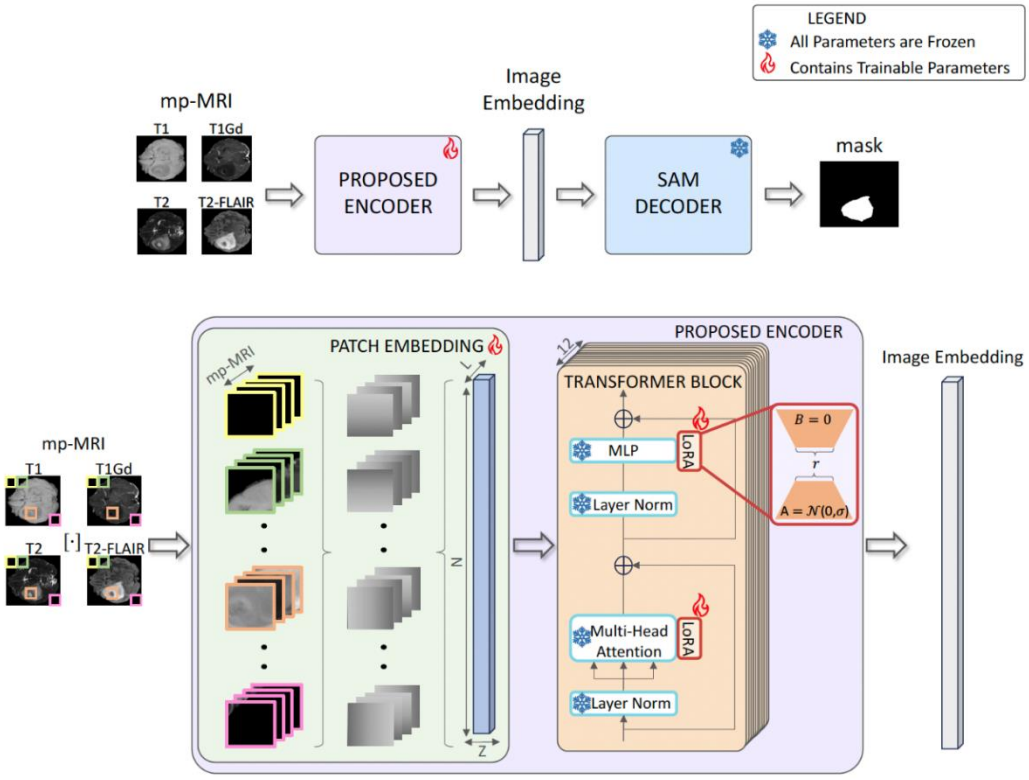
Abstract: Diffusion models represent a new paradigm in text-to-image generation. Beyond generating high-quality images from text prompts, models such as Stable Diffusion have been successfully extended to the joint generation of semantic segmentation pseudo-masks. However, current extensions primarily rely on extracting attentions linked to prompt words used for image synthesis. This approach limits the generation of segmentation masks derived from word tokens not contained in the text prompt. In this work, we introduce Open-Vocabulary Attention Maps (OVAM)–a training-free method for text-to-image diffusion models that enables the generation of attention maps for any word. In addition, we propose a lightweight optimization process based on OVAM for finding tokens that generate accurate attention maps for an object class with a single annotation.



We evaluate these tokens within existing state-of-the-art Stable Diffusion extensions. The best-performing model improves its mIoU from 52.1 to 86.6 for the synthetic images' pseudo-masks, demonstrating that our optimized tokens are an efficient way to improve the performance of existing methods without architectural changes or retraining.

Cecilia Diana Albelda, Roberto Alcover; Álvaro García Martín, Jesús Bescós Cano, "How SAM Perceives Different mp-MRI Brain Tumor Domains?", 4th Workshop on Domain adaptation, Explainability, Fairness in AI for Medical Image Analysis (DEF-AI-MIA) in conjunction with the IEEE Computer Vision and Pattern Recognition Conference (CVPR), Seattle (USA), Jun. 2024.

Abstract: Gliomas, among the deadliest forms of cancer, are brain tumors that present a significant challenge due to their rapid progression and resistance to treatment. Effective and early diagnosis is critical for improving patient prognosis. Deep learning, particularly through large-scale vision models like Segment Anything Model (SAM), offers a new pathway for tumor segmentation. This study seeks to address the primary challenge of adapting SAM for mp-MRI brain scans, which typically encompass multiple imaging modalities not fully utilized by standard three-channel vision models. We demonstrate that leveraging all available MRI modalities achieves superior performance compared to the standard mechanism of repeating a MRI scan to fit the input embedding.



Proposed Encoder. We propose to modify the patch embedding layer, so that it accounts for the all the MRI modalities, allowing for a seamless integration of the information. Then, we employ LoRAs to tune Multi Layer Perceptron blocks (MLP) and Attention (Q,K,V embedding) layers of the transformer blocks.

Our research also focuses on parameter-efficient tuning of SAM to effectively train the model while minimizing resource usage, showcasing significant improvements when evaluated across multiple datasets. Finally, we expose how SAM perceives differences across varied brain tumor domains by visually analyzing the features extracted on each of them.

Software

Open-Vocabulary Attention Maps with Token Optimization for Semantic Segmentation in Diffusion Models

https://github.com/vpulab/ovam

Implementation of our paper "Open-Vocabulary Attention Maps with Token Optimization for Semantic Segmentation in Diffusion Models", CVPR, 2024.

SAM Adaptation for mp-MRI Brain Tumor Segmentation

https://github.com/vpulab/med-sam-brain

Implementation of our paper "How SAM Perceives Different mp-MRI Brain Tumor Domains?", DEF-AI-MIA Workshop in conjunction with CVPR, 2024.

Master thesis

Cataloging and monitoring vegetaion in urban environments using geo-positioned images, Rafael López García (advisor: Marcos Escuero-Viñolo), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Visual Signal Processing, Univ. Autónoma de Madrid, Dec. 2023.

Abstract: This Master Thesis explores the use of advanced image processing tools for cataloging and characterizing vegetation in urban environments. Sequences captured by a geopositioned camera on a moving vehicle are analyzed using semantic segmentation algorithms, plant species classification, vegetation density determination. This allows for a detailed categorization of vegetation, significantly improving the understanding of its distribution and density in urban settings. Such advanced segmentation is essential for assessing the health of the urban ecosystem and planning environmental management interventions. The reidentification of vegetation in images captured at different times enable long-term monitoring of vegetation evolution, allowing observation of growth trends, responses to urban interventions, and seasonal changes. Additionally, the use of a plant species classifier is investigated, a key tool for enhancing the specificity of the catalog. This classifier not only facilitates the identification and differentiation of vegetation but also provides valuable data for biodiversity and conservation studies. Obtaining vegetation density based on position provides information about the spatial distribution and abundance of plant species in urban environments. This knowledge is crucial for understanding the ecological dynamics of these areas, including aspects such as biodiversity, ecosystem health, and their ability to provide environment improvements, such as climate regulation, air quality improvement, and the promotion of recreational and aesthetically pleasing spaces. Moreover, this analysis allows for the identification of areas with vegetation deficiency, thus guiding urban planning and

environmental management efforts towards improving vegetation coverage, which is crucial for urban well-being and sustainability. The ultimate goal is to develop an application that processes information derived from these analytical techniques, offering a useful and accessible tool for environmental management and informed decision-making. This project, which aims to create an updated and comprehensive vegetation register in urban environments like the UAM campus, includes a vegetation catalog that integrates spatio-temporal data.

Multi-Camera Multi-Target Tracking of Vehicles using Graph Neural Netowrks, Héctor Mejía Vallejo (advisor: Juan Carlos San Miguel), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Ciencia de Datos, Univ. Autónoma de Madrid, Feb. 2024.

Abstract: Multi-object multi-camera tracking involves accurately identifying and locating objects over time across multiple cameras concurrently. Despite advancements in the field, current works still rely on traditional cost matrices between pairs of tracklets to retrieve multi-camera tracks. These methods often struggle with preserving track consistency, resulting in suboptimal tracking performance. Moreover, graphs conveniently serve as data structures designed to model relationships within data and are ideally suited for track association. Thus, this thesis builds upon prior research on multi-camera multi-object vehicle tracking that frames the matching of single-camera tracklets as a link prediction problem on a graph, utilizing Graph Neural Networks (GNNs) as the model to predict these links. Specifically, a baseline architecture is built as a two-stage system comprising single-camera tracking and inter-camera association. For single-camera trajectories, the estimations used are provided by the dataset and correspond to the TNT tracker and SSD detector. Subsequently, the inter-camera association is performed using a GNN to predict links between pairs of single-camera embeddings corresponding to the same identity, followed by a connected components approach to retrieve global multi-camera trajectories. Then, a post-processing routine ensures that the number of single-camera tracklets predicted to belong to the same vehicle does not exceed the number of available cameras. To enhance the performance of the baseline, a graph sampling scheme is employed to randomly remove edges connecting vehicles with different identities until the desired ratio of positive-negative edges is achieved. This step aims to balance the training of the GNN. Additionally, embedding galleries consisting of a fixed number of re-identification embeddings from detections in single-camera tracklets are proposed as an alternative to the average embeddings. The features from these galleries are then combined using custom encoders to generate a single optimal embedding to feed the GNN. The baseline and all the proposed improvements are evaluated using the CityFlowV1 dataset from the Nvidia AI City Challenge 2020, where a Bayesian approach for hyperparameter optimization is applied to search for the optimal tracker between the baseline and the proposed improvements. Finally, The results outperform state-of-the-

art works by obtaining 85.69% IDF1 for tracking and 97.67% F1-score for link prediction. By comparison, the previous best-performing tracker achieved 79.87% IDF1.

Graduate thesis

Estado actual del EU-AI Act: evolución, contenido, y perspectivas (**Current state of the EU-AI Act: evolution, content, and prospects**), Olivia Merezeanu (advisor: Marcos Escudero-Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jan. 2024.

Abstract: This Bachelor's Thesis discusses the first regulatory framework for artificial intelligence in the European Union. The European Commission proposed that artificial intelligence systems must meet certain requirements before being able to provide their service to users. These requirements include that AI systems must ensure safety and respect the current legislation regarding fundamental rights and EU principles. They aim to provide a clear and predictable regulatory environment that offers legal security. Likewise, the regulatory framework aims to enhance governance and ensure the effective implementation of the law, as well as to foster a single market by promoting lawful, safe, and reliable usage.

Currently, artificial intelligence is booming and plays a significant role in digital transformation. It is increasingly present in various areas of daily life, triggering a wide range of economic and social benefits across all aspects and activities of society. However, this rapid technological evolution brings forth new risks, hence the need to strike a balance between technological development and the security of European citizens.

In this work we will explain the historical evolution of this regulatory framework since its inception, as well as the articles outlined in the "ALTAI" document and the ethical guidelines that artificial intelligence systems must adhere to. Possible obstacles and consequences of such regulations will also be analyzed. Finally, a comparison will be made between the European Union's regulations and the regulations currently under development in China, the United States, and Australia.

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

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN