ARTIFICIAL INTELLIGENCE

DM 630/Leonardo S.p.A. - Procedural Learning from ego/exocentric video and multimodal signals for aeronautical applications

Funded By	LEONARDO S.p.A. (Roma) [P.iva/CF:00881841001] MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584] Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	TOMMASI TATIANA - tatiana.tommasi@polito.it
Contact	TOMMASI TATIANA - tatiana.tommasi@polito.it PISTILLI FRANCESCA - francesca.pistilli@polito.it
Context of the research activity	This research will focus on deep learning methods able to understand and reason about procedural activities. With the current abundance of human-activity videos, procedural learning from videos has become an emerging topic. However, effectively learning and abstracting procedural knowledge remains a significant challenge. This approach is particularly valuable in extending applications to professional fields, such as aeronautics, where procedural videos are scarce and instructions are often formalized in alternative formats, like textbooks. In these scenarios, weakly supervised learning is crucial for leveraging limited annotated data to build robust procedural models. This research will investigate the major bottlenecks in procedural learning within realistic settings, with application on aeronautical. Progetto finanziato dal PNRR a valere sul DM 630/2024 - CUP: E14D24002330004
	Procedural learning consists of identifying the key steps to perform a task, determining their logic and temporal order, and forecasting the following steps to fulfill a given goal from a specific status. This is a growing research topic with several real-world applications. Currently, procedural learning literature focuses on learning from video demonstrations of human activities [1,2,3]. Existing methods generally rely on egocentric videos with their unique action-centric perspective or on instructional videos along with proper text

Objectives	descriptions [4,5]. However, it is essential to shift from approaches developed for a single scenario to more versatile, general-purpose models, able to learn broader procedural knowledge across varied environments. Existing datasets, such as Ego4D [6], encompass a variety of scenarios, making them a promising starting point for developing more generalizable and transferable models. Promoting atomic action abstraction has the potential to generate reusable knowledge that can be efficiently exploited not only in different environments and conditions (i.e. in case of domain shift) but also to learn novel tasks, as already demonstrated for human activity recognition and reasoning from egocentric videos [7]. This research aims at pushing the boundaries of such representation to more complex applications involving articulated procedures. This strategy is particularly valuable in extending applications to professional fields, such as aeronautics, where data is scarce. In this specific scenario, procedural videos are limited or not available and instructions are often formalized in alternative formats, like textbooks. By using weakly supervised learning techniques and algorithms developed in different settings it could be possible to transfer procedural learning to these challenging scenarios. Therefore, we aim at investigating novel methods for procedural learning, enabling procedural knowledge abstraction and reutilization. This research will provide an impact on both fundamental and applied research for a variety of computer vision and robotics applications. It will find applications in the aeronautical scenario, facing limited data availability by developing robust procedural models, able to tackle noise and domain shift (e.g. day/night illumination condition, different cockpit models), and creating novel testbenches of simulated pilot's operational environment with ego/exo videos and additional multimodal signals from the pilot cockpit. This research is part of an industrial collaboration with Leonardo Spa and it will involve a 6-month internship with the company. It is expected that the scientific results of the project will be reported at top computer vision, robotics and machine learning conferences (IEEE CVPR, IEEE ICCV, ECCV, IEEE IROS, IEEE ICRA, NeurIPS, ICML). At least one journal publication is expected on one of the following international journals: IEEE PAMI, IJCV, CVIU. [1] Bansal, Siddhant, Chetan Arora, and C. V. Jawahar. "My view is the best view: Procedure learning from egocentric videos." ECCV, 2022. [2] Lv, Zhaoyang, et al. "Aria Everyday Activities Dataset." arXiv preprint arXiv:2402.13349 (2024). [3] Song, Yale, et al. "Ego4d goal-step: Toward hierarchical understanding of procedural activities." NeurIPS 2024. [4] Afouras, Triantafyllos, et al. "Ht-step: Aligning instructional articles with how-to videos." NeurIPS 2024 [5] Zhong, Yiwu, et al. "Learning procedure-aware video representation from instructional videos and their narrations." CVPR 2023. [6] Grauman, Kristen, et al. "Ego4d: Around the world in 3,000 hours of egocentric video." CVPR 2022. [7] Peirone, Simone Alberto, et al. "A Backpack Full of Skills: Egocentric Video Understanding with Diverse Task Perspectives." CVPR 2024.
	Strong knowledge of linear algebra, calculus, probability are prerequisites. The candidate is required to have a good understanding of machine learning,

Skills and competencies for the development of the activity	deep learning, and computer vision concepts. The candidate is expected to have strong programming skills (Python) and familiarity with at least one recent deep learning framework (PyTorch or Tensorflow). The candidate is expected to be proactive and capable of autonomously studying and reading the most recent literature. English fluency, both oral and written, is required.