



Middle European
interdisciplinary
master's programme in
Cognitive Science

MEi:CogSci Learning Contract for the Mobility Semester



Erasmus+

1 Student Information

Student Name	Jelena Epifanic
Home University	Comenius University Bratislava
Student ID Number (Home University)	1355888
Degree Programme Code (Home University)	242100
Host University	University of Vienna

This learning contract ensures that the ECTS credits the MEi:CogSci-student acquires at the host university will be recognised at the home university. In order to make this contract valid, please follow the procedure below:

A) Preparation phase

1. **Planning of studies and courses at the host university:** Student fills out the semester contract in negotiation with local coordinator.
2. **Negotiation of Special Topic of Interest Module(s)/Mobility Project:** The student negotiates the *special topic of interest* (i.e., a cognitive phenomenon) they want to study and how (i.e., a combination of courses, lab work, self-study, literature used) with the supervisor.
3. **Concrete plan of the project:** The student specifies the work-plan for the module (elements of module, milestones, deliverables, dates,...).
4. **Acknowledgement:** The supervisor checks the contract and gives their OK;
 - a. The **student sends the LC to the local coordinators at the home and host university** (+ cc to the supervisor)
 - i. with the agreement sentence: "I agree to this learning contract"
 - ii. as a **.pdf only**
 - iii. adding their name to the title of the document, e.g. **SurnameName_LC_Mobility**
 - iv. with an email head of this format only: LC_ < student surname, first name> _ <supervisor surname>
 - b. **Supervisor acknowledges that they accept the proposal by replying to the email (reply to all).**
5. **Approval by the home university:** The local coordinator at the home university approves it or requests changes (go back to step 2)

B) Mobility phase

6. **In case of changes in project/planned courses:** the student has to inform the coordinators at the host and home universities immediately.
7. After finishing the project, the supervisor grades, signs and stamps the document.
8. Graded, signed and stamped Learning Contract is sent to the coordinator of the host university **within the specified deadline**.

C) Grading & recognition phase

8. **Final grading & recognition:** Original signed contract & certificates/transcripts are returned to coordinator at home university for grade recognition after the project has been finished.

2 Semester Contract

S-I-CS New Trends in Cognitive Science Module: 10 ECTS				
Course Title	Course Type	ECTS	Grade (host)	Grade (home)
New Trends in Cognitive Science	Seminar	6	1	
MEi:CogSci Journal Club	Seminar	4	1	
Module Grade				

S-I-PJ Special Topic of Interest (Project) Module: 20 ECTS				
Project Title	Supervisor	ECTS	Grade (host)	Grade (home)
Exploring the Unknown: Incremental Learning Strategies for Open World Object Detection	Markus Vincze Ao.Univ.Prof. Dipl.-Ing. Dr.techn.	20		
Module Grade				

Date, Stamp & Signature of Local Coordinator
at **Host** University

Date, Stamp & Signature of Local Coordinator
at **Home** University

3 S-I-PJ Special Topic of Interest (Project) Module

Learning Outcomes*

Subject specific

- Advanced knowledge and understanding of a phenomenon from the perspective of at least two disciplines

Methodological

- Ability to approach a phenomenon in an interdisciplinary manner

Generic/Instrumental

- Ability to write and follow a project plan

Systemic

- Interdisciplinary work/thinking
- Project-oriented work and organisational skill
- Critical evaluation of approaches & methods
- Quick orientation & navigation in mother and/or novel complex field
- Change of viewpoint/perspectives (intellectual mobility)
- Phenomenon-oriented thinking
- Problem-solving abilities

*as defined in the MEi:CogSci curriculum

3.1 S-I-PJ Special Topic of Interest (Project) Module – Project Specifications

3.1.1 General Project Information

Title of Specialisation Project	Supervisor	ECTS
Exploring the Unknown: Incremental Learning Strategies for Open World Object Detection	Markus Vincze Ao.Univ.Prof. Dipl.-Ing. Dr.techn.	20

3.1.2 Summary of Topic/Phenomenon (3000-4000 characters)

TOPIC OF SPECIALISATION

Design and implementation of 3D open-world detector

PHENOMENON & (PERSONAL) GOALS

Continuous learning has been a problem of great interest in the field of AI. For real-world model applications and further progress towards more general AI, it is important that these systems are able to efficiently process streams of information, autonomously learn multiple tasks, deal with uncertainty and the main prerequisite is to be able to continuously learn and accumulate previous knowledge, mirroring the way biological systems operate. Humans, for example, learn from a few examples, distill essential knowledge, and recall it when necessary. Additionally, a significant portion of knowledge acquisition occurs through intervention and observation. Recently, a team of researchers proposed a complex model of a "lifelong learning machines" that incorporates essential lifelong learning features observed in biological systems [1]. These features are: transfer and adaptation, overcoming catastrophic forgetting, exploiting task similarity, task-agnostic learning, noise tolerance, resource efficiency and sustainability.

Despite significant progress in robotics over the past few decades, there remains a substantial gap between current solutions and the development of a more general agent capable of autonomously exploring and incrementally learning from its environment. The complexity of the problem at hand is such that current solutions address only a specific aspect of this broad model within a limited domain. An illustrative example is the challenge of incremental learning in the field of computer vision. The Open World Object Detection (OWOD) [2] is problem introduced by K. J. Joseph et al. where the model learns new, previously "unknown" classes incrementally. Over the past few years, there have been efforts to develop 2D image detectors capable of functioning in an open-world context, enabling them to acquire knowledge about new object classes without forgetting the previously learned ones [2, 3, 4].

The importance of developing such robust systems lies in the ever-increasing demand for robots that can function effectively in dynamic, open-world environments. Whether a robot operates with limited knowledge or possesses a more comprehensive understanding, many tasks performed by these robots, like tidying up, rely heavily on accurate detection of changes in their surroundings.

Given that our primary focus is on Human-Robot Interaction (HRI), our objective is to develop a system capable of detecting and classifying objects within a scene. In our pursuit of creating a more precise and robust computer vision system designed for object manipulation by a service robot, the necessity for a 3D Detector becomes apparent. Importantly, we aim to achieve this using semantic knowledge acquired during model training, with the potential to expand it by learning to recognize new object instances. This approach will pave the way for incorporating other modalities in the future for HRI.

References

- [1] D. Kudithipudi et al., 'Biological underpinnings for lifelong learning machines', Nature Machine Intelligence, vol. 4, no. 3, pp. 196–210, Mar. 2022.
- [2] K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, 'Towards Open World Object Detection', in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 5826–5836.
- [3] A. Gupta, S. Narayan, K. J. Joseph, S. Khan, F. S. Khan, and M. Shah, 'OW-DETR: Open-world Detection Transformer', arXiv [cs.CV]. 2022.
- [4] S. Ma et al., 'CAT: LoCalization and IdentificAtion Cascade Detection Transformer for Open-World Object Detection', arXiv [cs.CV]. 2023.

3.2 Project Plan

The parties are aware that the project has to be finished by **16.02.2024**.

Information on deadlines at host and home universities is available on the MEi:CogSci websites.

3.2.1 Project Steps

Literature Research			Total Working Hours (WH)/ECTS: 75 / 3		
Working-package (WP)	Start – End	WH / ECTS	Activities	Resources required	Milestones (M)
WP Initial1	October/2023 - October/2023	50 / 2	Gaining an overview of relevant studies and related work on Continuous learning in biological and artificial systems	Computer, access to literature, journals, libraries, discussion time with supervisor	M1 Lit1
WP Initial2	October/2023 - October/2023	25 / 1	Creating annotated bibliography, writing notes	Computer, access to literature, journals, libraries	M1 Lit2

Conceptualisation			Total WH/ECTS: 50 / 2		
Working-package (WP)	Start – End	WH / ECTS	Activities	Resources required	Milestones (M)
WP Concept1	October/2023 - October/2023	25 / 1	Conceptualising a project goal based on identified studies/literature	Computer, consolidation time with supervisor	M2 Theses1
WP Concept2	October/2023 - October/2023	25 / 1	Defining model architecture and evaluation criteria	Computer, consolidation time with supervisor	M2 Theses1

Programming/Implementing the Model			Total WH/ECTS: 200 / 8		
Working-package (WP)	Start – End	WH / ECTS	Activities	Resources required	Milestones (M)
WP Program1	November/2023 - November/2023	50 / 2	Setting up development environment, integrating existing modules	Computer, access to tools, programming environment, frameworks, consolidation time with supervisor	M3 Mod.1
WP Program2	November/2023 - December/2023	75 / 3	Implementing defined architecture	Computer, access to tools, programming environment, frameworks, consolidation time with supervisor	M3 Mod.2

Programming/Implementing the Model					Total WH/ECTS: 200 / 8
Working-package (WP)	Start – End	WH / ECTS	Activities	Resources required	Milestones (M)
WP Program3	December/2023 - December/2023	75 / 3	Training the model and iterative adaption of the proposed architecture	Computer, access to tools, programming environment, frameworks, consolidation time with supervisor	M3 Mod.3

Running and Analysing the Model					Total WH/ECTS: 100 / 4
Working-package (WP)	Start – End	WH / ECTS	Activities	Resources required	Milestones (M)
WP Test1	January/2024 - January/2024	50 / 2	Evaluating the model on defined dataset and fine tuning of the proposed architecture	Computer, access to tools, programming environment, frameworks, consolidation time with supervisor	M4 Mod.1
WP Test2	January/2024 - January/2024	50 / 2	Evaluating the model on new test dataset and fine tuning of the proposed architecture	Computer, access to tools, programming environment, frameworks, consolidation time with supervisor	M4 Mod.2

Project Documentation					Total WH/ECTS: 75 / 3
Working-package (WP)	Start – End	WH / ECTS	Activities	Resources required	Milestones (M)
WP Docu1	October/2023 - February /2024	37.5 / 1.5	Short weekly reports on project status	Computer, consolidation time with supervisor	M4 Docu start
WP Docu2	February/2024 - February /2024	37.5 / 1.5	Final report on project written	Computer, consolidation time with supervisor	M4 Docu end

3.2.2 Project Milestones

Mile-stone	Result/"Product" and/or Deliverables
M1 Lit1	Relevant studies searched and the material of interest read
M1 Lit2	Bibliography annotated and literature notes on relevant studies finished
M2 Theses1	Concrete problem for project identified, project proposal written
M2 Theses2	Conceptual architecture defined
M3 Mod.1	Working environment set up, frameworks installed, existing modules integrated, datasets prepared
M3 Mod.2	Defined architecture implemented
M3 Mod.3	Model trained
M4 Mod.1	Evaluation of trained model and fine-tuning
M4 Mod.2	Testing the model performance on new test data
M4 Docu start	Short weekly reports on project status written
M4 Docu end	Final report and 7-page paper on project written (end of project documentation)

3.3 Short Project Report (~1 page, 3000-5000 characters)

Exploring the Unknown: Incremental Learning Strategies for Open World Object Detection

To better comprehend the complexity of lifelong learning problems in artificial systems, we first summarized recent advances in this domain, including theoretical strategies, biological foundations of lifelong learning, and practical applications. In real-world settings, implementing these features poses specific challenges; thus, in our work, we focused on the problem of incremental learning in the field of computer vision. Specifically, we focused on the Open World Object Detection (OWOD) [1] problem, where the model learns new, previously "unknown" classes incrementally. First, we assessed the state-of-the-art solutions using the open-world protocol [1] for 2D detection. Following the evaluation, we proposed strategies for the possible advancement of 2D detectors of interest for the future phase of the project. To address the 3D Open World object detection problem, we proposed the Prompting 3D World (Pro3DW) framework. Through these steps, we aim to contribute to current efforts to bridge the gap between theoretical understanding and practical implementation in lifelong learning for visual systems.

Lifelong Learning - Theoretical Foundations

Continual learning is a subject of significant research interest, spanning both biological and artificial systems [2], [3]. In artificial systems, continual learning involves learning from dynamic data distributions, mirroring learning in real-world scenarios. Presently, in its most basic understanding, lifelong learning focuses on the problem of catastrophic forgetting, where acquiring new knowledge interferes with previously learned knowledge, leading to performance degradation due to different data distributions [4], [5]. Wang et al. [2] systemized approaches to solving the continual learning problem into five groups: (i) regularization-based approach; (ii) replay-based approach; (iii) optimization-based approach; (iv) representation-based approach; (v) architecture-based approach. Regularization approaches, inspired by neuroscience models, specifically neuroplasticity, introduce strategies for constraining and penalizing the update of neural weights [6]. Neurogenesis [7], another mechanism observed in biological systems, has inspired dynamic architectures [8]. The brain exhibits the ability to generalize effectively by statistically extracting information from distinct episodic events. Episodic replay and Complementary Learning Systems (CLS) theory [9] have served as important inspirations for continual learning models [10]. In our solution, we mostly rely on the use of foundation models pre-trained on large-scale data corpora [11], [12], for downstream tasks [13]. This approach facilitates knowledge transfer and is a promising method for addressing continual learning challenges. Systemizing the challenges of continual learning with reference to biologically inspired solutions leads to comprehensively understanding the complexity of lifelong learning problems in artificial systems.

Open world detection problem

In our pursuit of creating a precise and robust computer vision system for object manipulation by a service robot, we aimed to address OWOD detection in both 2D and 3D image domains, combining the strengths of both approaches. While 2D detection has a longer history and more available data, 3D detection is vital for precise robot navigation in real-world environments. Our objective was to explore the potential of large-scale pre-trained models in addressing OWOD, both in 2D and 3D object detection tasks. Motivated by recent advancements in transformers-based architectures, especially in the 2D visual domain, we plan to leverage multi-modal foundation models like Grounding DINO [14] and OWLv2 [15] for 2D detection. For the 3D detection task, we intend to build upon the solutions proposed by Fan et al. [16], specifically the Promptable Object Pose Estimation (POPE) [73], which also utilizes pre-trained large-scale 2D foundation models. The framework Pro3DW we proposed for addressing the 3D OWOD problem adopts the same pipeline as the authors of the POPE solution.

Initially, we plan to utilize an off-the-shelf 2D object detector, specifically the pre-trained Grounding-DINO [14] or OWLv2 [15] models (both evaluated on selected datasets and fine-tuned if necessary), for generating object prompts and object classifications. Further, the object prompts generated by the 2D detector will be forwarded as the source view to POPE [16] to generate 3D bounding boxes. The implementation and evaluation of this solution will be addressed in the next phase of our project.

Experiments

For the evaluation, OWLv2 specifically OWLv2 CLIP B/16 ST+FT and Grounding DINO models are initialized with pre-trained weights. The experiment is implemented on split of MS-COCO dataset [17] and Pascal VOC [18]. Eighty classes are divided into four tasks. For each task, models are text prompted with the name of classes introduced in that task. In every next task 20 additional text prompts of classes' names are added summing up to 80 in the last, 4th task. Evaluation metrics included standard mean average precision (mAP) for known classes, while recall serves as the primary metric for detecting unknown objects. This choice stems from the absence of annotations for all possible unknown object instances in the dataset, as noted in previous works [78].

Results

Preliminary results obtained using the OWOD evaluation protocol on Grounding-DINO [14] or OWLv2 [15] did not meet our expectations in comparison to state-of-the-art OWOD models [1], [19], [20]. Further investigation is

necessary to determine whether employing different prompt techniques or fine-tuning the models would improve performance. To better understand the issue, we conducted qualitative analysis. We observed that the OWLv2 detector managed to detect instances of small objects, which could be one of the reasons for the suboptimal evaluation performance. We are confident that by refining parameters such as threshold score confidence and conducting further fine-tuning of the models, we can effectively address this challenge.

Conclusion

Insight into the theoretical foundations of the continual learning problem provides valuable understanding of the complexity involved. Continual learning, a major challenge, benefits from studying biological models, especially in addressing non-stationary data and temporal correlations. Neuroscience and cognitive science have therefore provided invaluable inspiration for the development of continual learning solutions.

With the idea of addressing a small aspect of this problem—Open World Detection in both 2D and 3D image domains—we investigated the utilization of pre-trained grounding models for this purpose. Preliminary results obtained using the OWO evaluation protocol did not meet our expectations in comparison to state-of-the-art OWO models, so further investigation is needed to determine whether to employ different prompt techniques or fine-tune the models. Despite the initial results, we believe this direction shows promise for further employment of these models for the purpose of 2D Open World Object Detection and 3D image-based Open World Object Detection problem, especially due to the successful integration of different modalities for our final goal of enhancing Human-Robot Interaction (HRI).

References

- [1] K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, 'Towards Open World Object Detection', in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 5826–5836.
- [2] L. Wang, X. Zhang, H. Su, and J. Zhu, 'A Comprehensive Survey of Continual Learning: Theory, Method and Application', arXiv [cs.LG]. 2024.
- [3] D. Kudithipudi et al., 'Biological underpinnings for lifelong learning machines', Nature Machine Intelligence, vol. 4, pp. 196–210, 03 2022.
- [4] R. M. French, 'Catastrophic forgetting in connectionist networks', Trends in cognitive sciences, vol. 3, no. 4, pp. 128–135, 1999.
- [5] M. De Lange et al., 'A continual learning survey: Defying forgetting in classification tasks', IEEE transactions on pattern analysis and machine intelligence, vol. 44, no. 7, pp. 3366–3385, 2021.
- [6] J. Kirkpatrick et al., 'Overcoming catastrophic forgetting in neural networks', Proceedings of the national academy of sciences, vol. 114, no. 13, pp. 3521–3526, 2017.
- [7] N. Urban and F. Guillemot, 'Neurogenesis in the embryonic and adult brain: same regulators, different roles', Frontiers in Cellular Neuroscience, vol. 8, 2014.
- [8] A. A. Rusu et al., 'Progressive neural networks', arXiv preprint arXiv:1606.04671, 2016.
- [9] D. Kumaran, D. Hassabis, and J. L. McClelland, 'What learning systems do intelligent agents need? Complementary learning systems theory updated', Trends in cognitive sciences, vol. 20, no. 7, pp. 512–534, 2016.
- [10] G. M. Van de Ven, H. T. Siegelmann, and A. S. Tolias, 'Brain-inspired replay for continual learning with artificial neural networks', Nature communications, vol. 11, no. 1, p. 4069, 2020.
- [11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', arXiv [cs.CL]. 2019.
- [12] A. Radford et al., 'Learning Transferable Visual Models From Natural Language Supervision', arXiv [cs.CV]. 2021.
- [13] X. Han et al., 'Pre-trained models: Past, present and future', AI Open, vol. 2, pp. 225–250, 2021.
- [14] S. Liu et al., 'Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection', arXiv [cs.CV]. 2023.
- [15] M. Minderer, A. Gritsenko, and N. Houlsby, 'Scaling Open-Vocabulary Object Detection', arXiv [cs.CV]. 2023.
- [16] Z. Fan et al., 'POPE: 6-DoF Promptable Pose Estimation of Any Object, in Any Scene, with One Reference', arXiv [cs.CV]. 2023.
- [17] T.-Y. Lin et al., 'Microsoft COCO: Common Objects in Context', arXiv [cs.CV]. 2015.
- [18] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, 'The pascal visual object classes challenge: A retrospective', International journal of computer vision, vol. 111, pp. 98–136, 2015.
- [19] A. Gupta, S. Narayan, K. J. Joseph, S. Khan, F. S. Khan, and M. Shah, 'OW-DETR: Open-world Detection Transformer', arXiv [cs.CV]. 2022.
- [20] S. Ma et al., 'CAT: LoCalization and IdentificAtion Cascade Detection Transformer for Open-World

Final grade for the project

____ / ____

Host Grade / Home Grade

(see grade conversion matrix on last page)

Date, Stamp & Signature of Supervisor (Host University)

Grade Conversion Matrix

BRAT		BUD		LJUB		VIE		ZAG	
A	výborne (excellent)	5	jeles (excellent)	10	odlično (excellent)	1	sehr gut (excellent)	5	odličan (excellent)
B	vel'mi dobre (very good)	4	jó (good)	9	prav dobro (very good)	2	gut (good)	4	vrlo dobar (very good)
C	dobre (good)	4	jó (good)	8	prav dobro (very good)	2	gut (good)	4	vrlo dobar (very good)
D	uspokojivo (satisfactory)	3	Közepes (fair)	7	dobro (good)	3	befriedigend (satisfactory)	3	dobar (good)
E	dostatočne (sufficient)	2	Elégséges (satisfactory)	6	Zadostno (sufficient)	4	genügend (sufficient)	2	dovoljan (satisfactory)
F	nedostatočne (insufficient)	1	Elégtelen (fail)	5	nezadostno (insufficient)	5	nicht genügend (insufficient)	1	nedovoljan (insatisfactory)