Dr. Mennatullah Siam

Curriculum Vitae

Mail	mennatul@ualberta.ca, menna.seyam@gmail.com
Homepage	https://msiam.github.io/homepage/
Research	Computer Vision, Machine Learning, Robotics, Foundation Models, Responsible AI.
Title	PhD, PEng.

ACADEMIC POSITIONS

Tenure-track Assistant Professor 7/23-5/25

Engineering and Applied Sciences, University of Ontario Institute of Technology, Canada, Full time.

Image and Video Understanding (IVU) Lab

Affiliate Assistant Professor 3/24-5/25

Computer Science, University of British Columbia, Canada.

EDUCATION

PhD in Computing Science 2015-2021

University of Alberta

Under supervision of Professor Martin Jagersand.

Thesis Title: learning video object segmentation from limited labeled data.

Thesis Nominated for Department Award.

MSc. in Communication and Information Technology 2010-2013

Nile University

Under supervision of Dr. Mohamed ElHelw.

Thesis Title: Robust Target Detection and Tracking in UAV Imagery.

BSc. in Computer Science 2006-2010

Ainshams University

Excellent with Honours.

Second in my batch (from around 300 students).

PUBLICATIONS

The two main journals of computer vision are the International Journal on Computer Vision (IJCV) and the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) with acceptance rates below 30%. The three main conferences are the IEEE International Conference on Computer Vision (ICCV), the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) and the European Conference on Computer Vision (ECCV). These three conferences are very selective in general less than 25% of the articles submitted are accepted and their proceedings play a role which is as important as journals. The two main conferences in robotics are the IEEE Conference on Robotics and Automation (ICRA) and the IEEE Conference on Intelligent Robots and Systems (IROS). According to Google Scholar my h-index is 19 with total citations more than 2K.

Journal Publications and Theses:

[1] **Mennatullah Siam**. "Learning Video Object Segmentation from Limited Labelled Data". In: *PhD thesis, University of Alberta* (2021).

[2] **Mennatullah Siam**. "Temporal Transductive Inference for Few-Shot Video Object Segmentation". In: *International Journal of Computer Vision*. 133 (2025), pp. 4465–4482. DOI: 10.1007/s11263-025-02390-x.

[3] Matthew Kowal, **Mennatullah Siam**, Md Amirul Islam, Neil D.B. Bruce, Richard P. Wildes, and Konstantinos G. Derpanis. "Quantifying and Learning Static vs. Dynamic Information in Deep Spatiotemporal Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).

[4] Abdul-Hakeem Omotayo, Ashery Mbilinyi, Lukman Ismaila, Houcemeddine Turki, Mahmoud Abdien, Karim Gamal, Idriss Tondji, Yvan Pimi, Naome A Etori, Marwa M Matar, ..., and **Mennatullah S**. "The State of Computer Vision Research in Africa". In: *Journal of Artificial Intelligence Research, Fairness and Bias in AI Special Issue* 81 (2024), pp. 43–69.

[5] Clifford Broni-Bediako, Junshi Xia, Jian Song, Hongruixuan Chen, **Mennatullah Siam**, and Naoto Yokoya. "Generalized Few-Shot Semantic Segmentation in Remote Sensing: Challenge and Benchmark". In: *IEEE Geoscience and Remote Sensing Letters* (2024).

Conference Publications:

[1] Rezaul Karim, He Zhao, Richard P. Wildes, and **Mennatullah Siam**. "MED-VT: Multiscale Encoder-Decoder Video Transformer with Application to Object Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

[2] **Mennatullah Siam**, Boris N. Oreshkin, and Martin Jagersand. "AMP: Adaptive Masked Proxies for Few-Shot Segmentation". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 5249–5258.

[3] Raghav Goyal, Wan-Cyuan Fan, **Mennatullah Siam**, and Leonid Sigal. "TAM-VT: Transformation-Aware Multi-scale Video Transformer for Segmentation and Tracking". In: *Winter Conference on Applications of Computer Vision* (2025).

[4] Rayat Mir Hossain, **Mennatullah Siam**, Leonid Sigal, and Jim Little. "Visual Prompting for Generalized Few-shot Segmentation: A Multi-scale Approach". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.

[5] Abdul-Hakeem Omotayo, Mai Gamal, Eman Ehab, Gbetondji Dovonon, Zainab Akinjobi, Ismaila Lukman, Houcemeddine Turki, Mahmod Abdien, Idriss Tondji, Abigail Oppong, Yvan Pimi, Karim Gamal, and **Mennatullah Siam**. "Towards a Better Understanding of the Computer Vision Research Community in Africa". In: *Equity and Access in Algorithms, Mechanisms, and Optimization*. 2023.

[6] Matthew Kowal, **Mennatullah Siam**, Md Amirul Islam, Neil D.B. Bruce, Richard P. Wildes, and Konstantinos G. Derpanis. "A Deeper Dive Into What Deep Spatiotemporal Networks Encode: Quantifying Static vs. Dynamic Information". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 13999–14009.

[7] **Mennatullah Siam***, Naren Doraiswamy*, Boris N. Oreshkin*, Hengshuai Yao, and Martin Jägersand (* equally contributing). "Weakly Supervised Few-shot Object Segmentation using Co-Attention with Visual and Semantic Embeddings". In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. 2020, pp. 860–867.

[8] **Mennatullah Siam**, Chen Jiang, Steven Lu, Laura Petrich, Mahmoud Gamal, Mohamed Elhoseiny, and Martin Jagersand. "Video object segmentation using teacher-student adaptation in a human robot interaction (HRI) setting". In: *Proceedings of the International Conference on Robotics and Automation*. 2019, pp. 50–56.

[9] Masood Dehghan*, Zichen Zhang*, **Mennatullah Siam***, Jun Jin, Laura Petrich, and Martin Jagersand (* equally contributing). "Online object and task learning via human robot interaction". In: *Proceedings of the International Conference on Robotics and Automation*. 2019, pp. 2132–2138.

[10] **Mennatullah Siam**, Sara Elkerdawy, Mostafa Gamal, Moemen Abdel-Razek, Martin Jagersand, and Hong Zhang. "Real-Time Segmentation with Appearance, Motion and Geometry". In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2018, pp. 5793–5800.

[11] **Mennatullah Siam***, Mostafa Gamal*, Moemen Abdel-Razek*, Martin Jagersand, and Senthil Yogamani (* equally contributing). "RTSeg: Real-time Semantic Segmentation Comparative Study". In: *Proceedings of the IEEE International Conference on Image Processing* (2018).

[12] **Mennatullah Siam**, Heba Mahgoub, Mohamed Zahran, Senthil Yogamani, Martin Jagersand, and Ahmad El-Sallab. "MODNet: Moving Object Detection Network with Motion and Appearance for Autonomous Driving". In: *Proceedings of the IEEE International Conference on Intelligent Transportation Systems* (2018).

[13] **Mennatullah Siam**, Sara Elkerdawy, Martin Jagersand, and Senthil Yogamani. "Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges". In: *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*. 2017, pp. 1–8.

[14] Eslam Mohamed, Mahmoud Ewaisha, **Mennatullah Siam**, Hazem Rashed, Senthil Kumar Yogamani, Waleed Hamdy, Mohamed El-Dakdouky, and Ahmad El Sallab. "Monocular Instance Motion Segmentation for Autonomous Driving: KITTI InstanceMotSeg Dataset and Multi-Task Baseline". In: *Proceedings of the IEEE Intelligent Vehicles Symposium*. 2021, pp. 114–121.

[15] Sepehr Valipour, **Mennatullah Siam**, Martin Jagersand, and Nilanjan Ray. "Recurrent Fully Convolutional Networks for Video Segmentation". In: *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. 2017.

[16] Abhineet Singh, **Mennatullah Siam**, and Martin Jagersand. "Unifying Registration based Tracking: A Case Study with Structural Similarity". In: *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. 2017.

[17] Sepehr Valipour, **Mennatullah Siam**, Eleni Stroulia, and Martin Jagersand. "Parking-stall vacancy indicator system, based on deep convolutional neural networks". In: *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. IEEE. 2016, pp. 655–660.

[18] **Mennatullah Siam**, Sepehr Valipour, Martin Jagersand, and Nilanjan Ray. "Convolutional gated recurrent networks for video segmentation". In: *2017 IEEE international conference on image processing (ICIP)*. IEEE. 2017, pp. 3090–3094.

[19] Bjarne Großmann, **Mennatullah Siam**, and Volker Krüger. "Comparative evaluation of 3D pose estimation of industrial objects in RGB pointclouds". In: *Proceedings of the International Conference on Computer Vision Systems*. 2015, pp. 329–342.

[20] **Mennatullah Siam**, Abhineet Singh, Camilo Perez, and Martin Jagersand. "4-DoF tracking for robot fine manipulation tasks". In: *2017 14th Conference on Computer and Robot Vision (CRV)*. IEEE. 2017, pp. 329–336.

[21] **Menna Siam**, Ramy ElSayed, and Mohamed ElHelw. "On-board multiple target detection and tracking on camera-equipped aerial vehicles". In: *Proceedings of the IEEE International Conference on Robotics and Biomimetics*. 2012, pp. 2399–2405.

[22] **Mennatullah Siam** and Mohammed ElHelw. "Tracking ground targets from a UAV using New PN constraints". In: *IFAC Proceedings Volumes* 46.30 (2013), pp. 19–25.

Preprint and Under-review Publications:

[1] Mir Rayat Imtiaz Hossain, **Mennatullah Siam**, Leonid Sigal, and James J Little. "The Power of One: A Single Example is All it Takes for Segmentation in VLMs". In: *arXiv preprint arXiv:2503.10779* (2025).

[2] **Mennatullah Siam**. "PixFoundation 2.0: Do Video Multi-Modal LLMs Use Motion in Visual Grounding?" In: *arXiv preprint arXiv:2509.02807* (2025).

[3] **Mennatullah Siam**. "PixFoundation: Are We Heading in the Right Direction with Pixel-level Vision Foundation Models?" In: *arXiv preprint arXiv:2502.04192* (2025).

[4] Mai Gamal, Mohamed Rashad, Eman Ehab, Seif Eldawlatly, and **Mennatullah Siam**. "System Identification of Neural Systems: Going Beyond Images to Modelling Dynamics". In: *arXiv preprint arXiv:2402.12519* (2024).

[5] Rezaul Karim, He Zhao, Richard P. Wildes, and **Mennatullah Siam**. "A Unified Multiscale Encoder-Decoder Transformer for Video Segmentation". In: *arXiv preprint arXiv:2304.05930* (2023).

[6] **Mennatullah Siam**, Alex Kendall, and Martin Jagersand. "Video Class Agnostic Segmentation with Contrastive Learning for Autonomous Driving". In: *arXiv preprint arXiv:2105.03533* (2021).

[7] **Mennatullah Siam**, Konstantinos G. Derpanis, and Richard P. Wildes. "Multiscale Memory Comparator Transformer for Few-shot Video Segmentation". In: *arXiv preprint arXiv:2307.07812*. 2023.

[8] Leila Cheshmi and **Mennatullah Siam**. "Multiscale Video Transformers for Class Agnostic Segmentation in Autonomous Driving". In: *arXiv preprint arXiv:2508.14729* (2025).

Workshop Publications:

[1] Abduljaleel Adejumo, Faegheh Yeganli, Clifford Broni-bediako, Aoran Xiao, Naoto Yokoya, and **Mennatullah Siam**. "A Vision Centric Remote Sensing Benchmark". In: *Eval-FoMo-2 Workshop (Non-Archival), CVPR* (2025).

[2] **Mennatullah Siam**. "PixFoundation: Are We Heading in the Right Direction with Pixel-level Vision Foundation Models?" In: *Eval-FoMo-2 Workshop (Non-Archival), CVPR* (2025).

[3] Mai Gamal, Mohamed Rashad, Eman Ehab, Seif Eldawlatly, and **Mennatullah Siam**. "System Identification of Neural Systems: Going Beyond Images to Modelling Dynamics". In: *Neuro AI Workshop Short Paper, NeurIPS* (2024).

[4] Hesham Ali, Idriss Tondji, and **Mennatullah Siam**. "Two-stage Joint Transductive and Inductive learning for Nuclei Segmentation". In: *Machine Learning for Health Symposium, Findings Track (Non Archival)* (2023).

[5] Matej Kristan, Jiri Matas, Pavel Tokmakov, et al. "The Second Visual Object Tracking Segmentation VOTS2024 Challenge Results". In: *ECCV Workshops (7)*. 2024.

[6] **Mennatullah Siam**, Konstantinos G. Derpanis, and Richard P. Wildes. "Temporal Transductive Inference for Few-Shot Video Object Segmentation". In: *Machine Learning for Autonomous Driving Workshop in Neurips*. 2021.

[7] **Mennatullah Siam**, Alex Kendall, and Martin Jagersand. "Video Class Agnostic Segmentation Benchmark for Autonomous Driving". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. June 2021, pp. 2825–2834.

[8] **Mennatullah Siam**, Boris Oreshkin, and Martin Jagersand. "Adaptive Masked Weight Imprinting for Few-Shot Segmentation". In: *Learning from Limited Labelled Data ICLR Workshop* (2019).

[9] **Mennatullah Siam**, Mostafa Gamal, Moemen Abdel-Razek, Senthil Yogamani, Martin Jagersand, and Hong Zhang. "A comparative study of real-time semantic segmentation for autonomous driving". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2018, pp. 587–597.

[10] **Mennatullah Siam**, Heba Mahgoub, Mohamed Zahran, Senthil Yogamani, Martin Jagersand, and Ahmad El-Sallab. "Motion and Appearance Based Multi-Task Learning Network for Autonomous Driving". In: *Machine Learning for Intelligent Transportation Neurips Workshops* (2017).

[11] **Mennatullah Siam** and Mohammed Elhelw. "Enhanced target tracking in uav imagery with pn learning and structural constraints". In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2013, pp. 586–593.

PATENTS

- Mennatullah Siam, Senthil Yogamani, Ahmad ElSallab, and Heba Mahgoub. "Motion and Appearance Based Multi-Task Learning of Motion Segmentation and Vehicle Detection", `https://worldwide.espacenet.com/publicationDetails/biblio?CC=DE&NR=102018114229&KC=&FT=E&locale=en_EP#`. (pending)

RESEARCH GRANTS

Total Funds: ∼$314,000 (CAD)

- NSERC Discovery Grant, "Learning Pixel-level Video Understanding". $120,000 (CAD)
- NSERC Discovery Grant Amendment. $35,000 (CAD)
- NSERC Discovery Grant Launch Supplements. $12,500 (CAD)

- NSERC Alliance International, Catalyst Grant, "Assessing the Suitability of Datasets in Foundation Models Development: Foundational Models for Remote Sensing A Case Study". $25,000 (CAD)
- Digital Research Alliance of Canada, Resources for Research Group. $44,214 (CAD)
- Startup Funds, Ontario Tech University. $60,000 (CAD)
- Mitacs Globalink, "Adapting Foundation Models to 3D Segmentation of Medical Data", UBC. $6,000 (CAD)
- Google Research Credits. $6,000 (USD)
- OpenAI Research Credits. $2,000 (USD)

RESEARCH SUPERVISION

Research Scientist

- Faegheh Yeganli, UBC (Supervision)

MSc/MEng Students

- Abduljaleel Adejumo, AMMI-AIMS (Supervision, 2024-2025)
- Leila Cheshmi, Ontario Tech (Supervision, 2023-2025)

Interns

- Mai Gamal, German University in Cairo (PhD Student) (Summer'23, Summer'24)
- Omid Reza Heidari, Concordia University (MSc Student) (Winter'25)

AWARDS AND RECOGNITIONS

- Outstanding Reviewer in ICCV 2023 [1].
- VISTA Postdoctoral Fellowship ($110,000). 2021-2023.
- Alberta Innovates Technology Futures Graduate Scholarship (PhD) ($94,500). 2017-2019.
- Alberta Graduate Excellence Scholarship ($12,000). 2019.
- Verna Tate Graduate Scholarship in Science, nominated from Computing Science Department, University of Alberta ($10,000). 2019.
- KUKA Innovation Finalist Award. 2018.
- TAM-VT, UBC students work, VOTS Challenge, 5th place, ECCV, 2024.

[1] https://iccv2023.thecvf.com/outstanding.reviewers-118.php

INVITED TALKS

- (Keynote) "Learning Scene and Video Understanding with Limited Labelled Data". Black in AI workshop, co-located with Neurips, 2022.
- "From Scene to Video Understanding, What to Consider". University of British Columbia, Canada, 2022.
- "From Scene to Video Understanding, What to Consider". Dr. Alice Othmani Lab, France (remote), 2023.
- "Few-shot Learning Tutorial." Samsung AI, Canada, 2022.
- "Image and Video Class Agnostic Segmentation". Huawei, Canada, 2021.
- "Image and Video Class Agnostic Segmentation". York University, Canada, 2021.
- "On the Intersection of Few-shot and Video Object Segmentation." Doctoral Consortium, CVPR. Online, 2021.

TEACHING EXPERIENCE

Instructor SOFE4620U Winter 2024
Ontario Tech University
Machine Learning and Data Mining (72 students).

Instructor SOFE2715U Winter 2024
Ontario Tech University
Data Structures (100 students).

Instructor ELEE2110 Fall 23,24
Ontario Tech University
Discrete Mathematics for Engineers (96 students).

Instructor CIT 690 Spring 2022
Nile University
Computer Vision for Graduate students (20 students).

Co-Instructor MM 805 Winter 2021
University of Alberta
Computer Vision and 3DTV for Master students in Multimedia Program (26 students).

ACADEMIC SERVICE

- Primary Organizer of PixFoundation: Pixel-level Vision Foundation Models Workshop in CVPR 2025 [2].
- Area chair in IEEE CVF Winter Conference on Applications of Computer Vision (WACV), 2024, 2025, 2026.
- Supporting Organizer of African Computer Vision Summer School (ACVSS), Nairobi, Kenya, 2024 [3].

[2] https://sites.google.com/view/pixfoundation
[3] https://sites.google.com/view/acvss/past-editions/2024/

- Speaker in the Introduction to Computer Vision Tutorials, preparation to ACVSS, remote, 2024 [4].
- Organizer of 3^{rd} Workshop on Learning with Limited Labelled Data for Image and Video Understanding (L3D-IVU) in CVPR 2024 [5].
- Organizer of 2^{nd} Workshop on Learning with Limited Labelled Data for Image and Video Understanding (L3D-IVU) in CVPR 2023 [6].
- Organizer of Learning with Limited Labelled Data for Image and Video Understanding (L3D-IVU) Workshop in CVPR 2022, [7].
- Technical committee member in Medical Image Learning with Limited and Noisy Data (MILLand) Workshop in MICCAI 2022, 2023 [8].
- Organizer CV4Africa workshop, Deep Learning Indaba, Accra, Ghana, 2023.
- Organizer and mentor in Black in AI social, CVPR, Vancouver, Canada, 2023.
- Guest Editor of special issue in Remote Sensing Journal on Autonomous Driving [9], 2022.
- Reviewer in ICRA, IROS, ECCV, WACV, CVPR, ICCV, NeurIPS, ICLR (for the last six years).
- Reviewer in IJCV, TPAMI, Pattern Recognition Letters, IEEE Intelligent Systems, IEEE Transactions on Robotics.
- Program Committee Member in Machine Learning for Autonomous Driving Workshop - Neurips 2020, 2021.

TEACHING, RESEARCH AND INDUSTRY EXPERIENCE

I have strong experience in industry and academia across both Computer Vision and Robotics research. For the later I have worked across three robotic platforms including: (i) robotic arms (e.g., KUKA robotic arm in my PhD and UR in Aalborg Copenhagen), (ii) self driving cars (during my internship in Wayve), and (iii) unmanned aerial vehicles (quadcopters VTOLs during my MSc). I have also experience working with and contributing to various robotic simulators including Carla simulator (during my internship in Wayve) and ROS Gazebo, and experience working in with ROS and multiple DL frameworks, e.g., Pytorch. Finally, I have experience mentoring a multitude of undergrad, MSc, PhD students and interns across both Computer Vision and Robotics.

Lecturer-Term 2/22-7/22

Nile University, Information Technology and Computer Science, Remote, Part time.

Teaching Computer Vision CIT-690 Course

Postdoctoral Fellow 7/21-7/23

York University, Lassonde School of Engineering, Full time.

[4] https://www.youtube.com/watch?v=6_WbhllqKRk&list=PL4jKsHbreeuDK-QNIYnB8M1qPoTqdzr1y
[5] https://sites.google.com/view/l3divu2024
[6] https://sites.google.com/view/l3d-ivu-2023
[7] https://sites.google.com/view/l3d-ivu/
[8] https://zghada90.wixsite.com/milland/committee
[9] https://www.mdpi.com/journal/remotesensing/special_issues/75B73YS791

CVIL York Lab. Supervised by Professor Richard Wildes

Postdoc Researcher 3/22-2/23
Vector Institute, Full time.

Machine Learning Engineer Intern 6-12/2020
Wayve Ltd, London, UK.

Research Intern 8/19-4/20
Huawei Research, Edmonton, Canada.

Software Engineering Intern 6-9/2018
Autonomous Driving Team, Nvidia Corporation, Santa Clara, US.

Software Engineering Intern 5-7/2017
Valeo Vision Systems, Ireland and DL Research, Egypt.

Research Assistant 2014-2015
Aalborg Copenhagen University, Mechanical Engineering, Denmark.

Software Engineering Intern 2012-2013
Sony Stuttgart Technology Center, Germany

VOLUNTEERING EXPERIENCE

Super Volunteer 2021, 2022
WiML Neurips 2021, 2022
Volunteering in helping out workshop organizers in different tasks.

Mentor 2019-2020
Black in AI
Helped as a mentor in the mentoring program within BAI.

Co-Founder 2022
Ro'ya CV4Africa Initiative
Volunteered to start a 'computer vision for Africa' community as part of Deep Learning Indaba. Our community slack has more than 400 members. [10]

[10] https://x.com/RoyaCV4Africa