Alaaeldin El-Nouby

alaaelnouby@gmail.com | aelnouby.github.io | jxpBMwwAAAAJ

Education

Meta AI and École Normale Supérieure *Paris, France*

PhD in Computer Science *August 2020 - Present*

- Advisors: Hervé Jegou, Ivan Laptev, and Natalia Neverova.
- Research: Vision Transformers, Self-Supervised Learning, Language-Vision Pre-training, Image Compression, Image Retrieval.

University of Guelph *Guelph, Ontario, Canada*

MASc in Engineering Systems and Computing *September 2017 - September 2019*

- Advisors: Graham Taylor, Rozita Dara
- GPA: 4.0
- Awards: Vector Institute Research Grant, CEPS Dean's Scholarship

Ain Shams University *Cairo, Egypt*

BEng Computers and Systems Engineering *September 2011 - June 2016*

- Advisor: Hazem Abbas
- Grade: Distinction
- Bachelor Thesis: Smart Airport Surveillance System (Action Recognition, Unattended Object Detection, Tracking)

Experience

FAIR, Meta AI *Paris, France*

PhD Resident *August 2020 – Present*

- Advised by Hervé Jegou and Natalia Neverova at FAIR. Conducting research on Vision Transformers, Self-supervised learning, Language-vision pre-training, Image Retrieval and Image Compression. My research at FAIR has resulted in 12 papers, 6 of which as a primary author.
- Anticipated Graduation: June-August 2023

Apple Inc. *Cupertino, CA, US*

Research Intern *February 2019 – May 2019*

- Worked as a part of the Video Engineering Group under the supervision of Josh Susskind and Shuangfei Zhai, conducting research on video understanding. An early version of this project won the "Best Poster Award" at the *Large Scale Holistic Video Understanding workshop, ICCV2019*.

Microsoft Research *Montreal, Canada*

Research Intern *August 2018 – November 2018*

- Worked on using Generative models and dialogues for text to image in a sequential multi-step manner. The internship resulted in GeNeVA, a novel image generation method published at *ICCV 2019*.

Vector Institute / University of Guelph *Guelph, ON, Canada*

Graduate Research Assistant *September 2017 – September 2019*

- Advised by Dr. Graham Taylor at University of Guelph as a student researcher at Vector Institute.
- Conducted research on learning better features for video understanding and Human activity recognition.
- Research was funded by the DeepVision project.

Microsoft Advanced Technology Lab Cairo | NLP Team *Cairo, Egypt*

Research Software Development Engineer *August 2016 – August 2017*

- Worked on developing a task oriented virtual agent for customer support. My focus was on developing engineering pipelines that automates the extraction of answers from different sources on the web (e.g. web pages, click through logs).

Publications

- OmniMAE: Single Model Masked Pretraining on Images and Videos - Rohit Girdhar*, Alaaeldin El-Nouby*, Mannat Singh*, Kalyan Vasudev Alwala*, Armand Joulin, Ishan Misra*. Under Review
- Are Large-scale Datasets Necessary for Self-Supervised Pre-training? - Alaaeldin El-Nouby*, Gautier Izacard*, Hugo Touvron, Ivan Laptev, Hervé Jegou, Edouard Grave. Under Review
- XCiT: Cross-Covariance Image Transformers - Alaaeldin El-Nouby, Hugo Touvron, Mathilde Caron, Piotr Bojanowski, Matthijs Douze, Armand Joulin, Ivan Laptev, Natalia Neverova, Gabriel Synnaeve, Jakob Verbeek, Hervé Jegou. NeurIPS 2021
- LeViT: a Vision Transformer in ConvNet's Clothing for Faster Inference - Ben Graham, Alaaeldin El-Nouby, Hugo Touvron, Pierre Stock, Armand Joulin, Hervé Jégou, Matthijs Douze. ICCV 2021

- Tell, draw, and repeat: Generating and modifying images based on continual linguistic instruction - Alaaeldin El-Nouby, Shikhar Sharma, Hannes Schulz, Devon Hjelm, Layla El Asri, Samira Ebrahimi Kahou, Yoshua Bengio, Graham W Taylor. ICCV 2019
- Resmlp: Feedforward networks for image classification with data-efficient training - Hugo Touvron, Piotr Bojanowski, Mathilde Caron, Matthieu Cord, Alaaeldin El-Nouby, Edouard Grave, Gautier Izacard, Armand Joulin, Gabriel Synnaeve, Jakob Verbeek, Hervé Jégou. TPAMI
- Three things everyone should know about Vision Transformers - Hugo Touvron, Matthieu Cord, Alaaeldin El-Nouby, Jakob Verbeek, Hervé Jégou. ECCV 2022
- Augmenting Convolutional networks with attention-based aggregation - Hugo Touvron, Matthieu Cord, Alaaeldin El-Nouby, Piotr Bojanowski, Armand Joulin, Gabriel Synnaeve, Hervé Jégou. arXiv 2021
- Training vision transformers for image retrieval - Alaaeldin El-Nouby, Natalia Neverova, Ivan Laptev, Hervé Jégou. arXiv 2021
- Skip-Clip: Self-Supervised Spatiotemporal Representation Learning by Future Clip Order Ranking - Alaaeldin El-Nouby, Shuangfei Zhai, Graham W Taylor, Joshua M Susskind. ICCV 2010 workshop - Best poster Award
- Real-Time End-to-End Action Detection with Two-Stream Networks - Alaaeldin El-Nouby, Graham W Taylor. CRV 2018

Academic Service

Reviewing ICCV'21, CVPR'22, ECCV'22, NeurIPS'22 SSL worhhsop, TPAMI (2021-Present).

Teaching CIS*3110 Operating Systems - *University of Guelph*, CSE*121 Fundamentals of Programming - *Ain Shams University*

Interns Co-Advised Federico Baldassarre (with Hervé Jegou) and Joao Maria Janeiro (with Jakob Verbeek) while at FAIR.

Invited Talks

- **Vector Institute / University of Guelph** - Masked Image Modeling for Visual Representation Learning
- **Transformers for Vision workshop, CVPR'22** - Are Large-scale Datasets Necessary for Self-Supervised Pre-training?
- **Max Planck Institute / Tübingen AI Center** - Are Large-scale Datasets Necessary for Self-Supervised Pre-training?
- **Computer Vision Group (CVG), University of Bern** - Are Large-scale Datasets Necessary for Self-Supervised Pre-training?
- **Large Scale Holistic Video Understanding workshop, CVPR'21** - Training Vision Transformers for Image Retrieval
- **KTH Royal Institute of Technology** - Training Vision Transformers for Image Retrieval
- **Microsoft Research Montreal** - Sequential Scene Understanding and Generation
- **DeepVision workshop, Simon Fraser University** - Real-Time End-to-End Action Detection with Two-Stream Networks
- **Twenty Billion Neurons** - Real-Time End-to-End Action Detection with Two-Stream Networks