# Anurag Arnab

*Staff Research Scientist, Google DeepMind*

*anurag.arnab@gmail.com*
https://anuragarnab.github.io

## Summary

I am a Staff Research Scientist at Google DeepMind working on multimodal understanding and generation. I enjoy solving challenging problems and envisage putting my artificial intelligence and software engineering skills to use in a manner that is beneficial to society.

## Education

2015–2019 **DPhil in Information Engineering (Computer Vision)**, *University of Oxford*, United Kingdom
Research focused on deep learning and integrating probabilistic graphical models into neural networks. I published at leading venues about this line of work and its applications to semantic-, instance- and panoptic-segmentation respectively.
Supervisor: Professor Philip Torr

2011–2014 **BSc (Eng) Electrical and Computer Engineering**, *University of Cape Town*, South Africa
Graduated as the top student in the entire Engineering faculty of approximately 500 students.

## Experience

Sept. 2019 – present **Research Scientist**, *Google DeepMind*, Grenoble, France
Working with Cordelia Schmid on multimodal understanding and generation. I have developed numerous state-of-the-art algorithms which have landed in Gemini and "Nano Banana", published at leading venues, filed 16 patents, and transferred technology internally to products such as YouTube, Photos, Cloud and Lens.

June - Oct. 2018 **Research Scientist Intern**, *DeepMind*, London, United Kingdom
Worked with Andrew Zisserman. A paper about the project was subsequently published at CVPR 2019.

June - Sep. 2017 **Software Engineering Intern (Research)**, *Google*, Zürich, Switzerland
Worked on large-scale image retrieval. The neural network I developed surpassed the previous system by a large margin and was subsequently put into production as part of Google Lens.

Jan. - July 2015 **Visiting Intern**, *Torr Vision Group*, University of Oxford, United Kingdom
Visiting intern at Prof. Philip Torr's research group where I subsequently did my PhD. The internship culminated in a first-author publication at BMVC.

## Selected Publications

Full list on *Google Scholar*. More than 13 000 citations, and h-index of 34, at time of writing.

**Anurag Arnab***, Ahmet Iscen*, Mathilde Caron, Alireza Fathi, Cordelia Schmid. Temporal Chain of Thought: Long-Video Understanding by Thinking in Frames. *NeurIPS*, 2025

Gemini Team. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *Technical Report*, 2024

Xingyi Zhou*, **Anurag Arnab***, Shyamal Buch, Shen Yan, Austin Myers, Xuehan Xiong, Arsha Nagrani, Cordelia Schmid. Streaming Dense Video Captioning. *CVPR*, 2024

Otniel-Bogdan Mercea, Alexey Gritsenko, Cordelia Schmid, **Anurag Arnab**. Time-Memory- and Parameter-Efficient Visual Adaptation. *CVPR*, 2024

Google Research. Scaling Vision Transformers to 22 Billion Parameters. *ICML*, 2023

Michael S. Ryoo, Keerthana Gopalakrishnan, Kumara Kahatapitiya, Ted Xiao, Kanishka Rao, Austin Stone, Yao Lu, Julian Ibarz, **Anurag Arnab**. Token Turing Machines. *CVPR*, 2023

**Anurag Arnab***, Mostafa Dehghani*, Georg Heigold, Chen Sun, Mario Lučić, Cordelia Schmid. ViViT: A Video Vision Transformer. *ICCV*, 2021

**Anurag Arnab**, Chen Sun, Cordelia Schmid. Unified Graph Structured Models for Video Understanding. *ICCV*, 2021

Kuang-Huei Lee*, **Anurag Arnab***, Sergio Guadarrama, John Canny, Ian Fischer*. Compressive Visual Representations. *NeurIPS*, 2021

**Anurag Arnab***, Carl Doersch*, Andrew Zisserman. Exploiting Temporal Context for 3D Human Pose Estimation in the Wild. *CVPR*, 2019

Qizhu Li*, **Anurag Arnab***, Philip H.S Torr. Weakly- and Semi-Supervised Panoptic Segmentation. *ECCV*, 2018

**Anurag Arnab**, Ondrej Miksik, Philip H.S Torr. On the Robustness of Semantic Segmentation Models to Adversarial Attacks. *CVPR*, 2018

**Anurag Arnab**, Philip H.S Torr. Pixelwise Instance Segmentation with a Dynamically Instantiated Network. *CVPR*, 2017

**Anurag Arnab**, Sadeep Jayasumana, Shuai Zheng, Philip H.S Torr. Higher Order Conditional Random Fields in Deep Neural Networks. *ECCV*, 2016

## Awards and Honours

- **Epic Kitchens Action Recognition Challenge (CVPR 2022)** Member of winning team.
- **Outstanding Reviewer Award**. Recognised as an outstanding reviewer for the NeurIPS 2019, ECCV 2020 and NeurIPS 2021 conferences.
- **Clarendon Scholarship**, 2015-2018. Awarded to only two graduate students applying to the Department of Engineering Science at Oxford on the basis of academic merit.
- **Engineering Council of South Africa Medal of Merit**, 2014. Awarded to the best student graduating with the degree of BSc (Eng).
- **ESKOM Award**, 2014. Awarded to the best Engineering graduate at the University of Cape Town over the 4-year degree curriculum, out of approximately 500 students.
- **City of Cape Town Corporation Silver Medal**, 2013. Best engineering student across third year, out of approximately 500 students.
- **City of Cape Town Corporation Bronze Medal**, 2012. Best engineering student across second year, out of approximately 500 students.
- **Electrical and Computer Engineering Class Medal**. Best student in degree programme in each year of study from 2011 through 2014.
- **Class Medals** Awarded for achieving the highest marks in an individual course. Obtained seven medals in the following undergraduate courses:
  - Linear Algebra and Differential Equations *class size of about 500 students*
  - Operating Systems and C++ *class size of about 60*
  - Computer Science 2: Concurrency, Computer Architecture and Human-Computer Interaction *class size of about 150*
  - Object Oriented Programming *class size of about 300*
  - Digital Systems *class size of about 30*
  - Electrical Engineering I *class size of about 160*
  - Africa: Culture, Identity and Globalisation *class size of about 160*

## Invited Talks

Dec. 2022 **Google visits POSTECH**, *Pohang*, South Korea
Presented research on large-scale video understanding at POSTECH and GIST universities in Korea.

June 2022 **CVPR Workshop: Holistic Video Understanding**, *New Orleans*, United States
Presented research on large-scale video understanding with transformers (CVPR 2022, NeurIPS 2021, ICCV 2021) at the CVPR conference.

June 2022 **CVPR Workshop: Epic Kitchens Challenge**, *New Orleans*, United States
Presented our winning entry for the Action Recognition challenge.

June 2021 **CVPR Workshop: Learning from Limited and Imperfect Data (L2ID)**, *Online*, United States
Presented research on transformers (ICCV 2021) and learning from weak supervision (ECCV 2020) in the context of video understanding at the CVPR conference.

June 2021 **Structured Models for Video Understanding**, *Online*, UNIST, South Korea
Invited talk at Ulsan National Institute of Science and Technology (UNIST). Presented work on transformer- and graph-structured-models (both published at ICCV 2021) for video understanding.

January 2021 **Visual Intelligence Seminar**, *Online*, Fudan University, China
Presented work on action recognition (ECCV 2020), video segmentation (IROS 2020) and 3D human pose estimation (CVPR 2019).

January 2020 **Scene Understanding with Deep Structured Models**, *University of Warsaw*, Poland
Guest lecture at a Computer Vision graduate class at the Universtiy of Warsaw.

June 2019 **CVPR Workshop: Learning from Imperfect Data**, *Long Beach*, United States
Presented my work on weakly-supervised 3D human pose estimation (CVPR 2019) and panoptic segmentation (ECCV 2018).

Aug. 2017 **Computer Vision and Geometry Group**, *ETH Zürich*, Switzerland
Presented my work on semantic- and instance segmentation (CVPR 2017, BMVC 2017, ECCV 2016).

Oct. 2016 **ECCV Tutorial: Deep Learning Meets Model Optimization and Statistical Inference**, *Amsterdam*, Netherlands
Presented my ECCV and BMVC 2016 papers about incorporating mean-field inference of Conditional Random Fields into neural networks for semantic- and instance segmentation respectively.

Feb. 2016 **Vision and Learning Seminar**, *Online*, Nankai University, China
Presented my BMVC 2015 paper which incorporated audio information into semantic segmentation, as part of a larger presentation of our research group's work.

Nov. 2015 **Centre for Vision, Speech and Signal Processing**, *University of Surrey*, United Kingdom
Invited to give a seminar about my BMVC 2015 paper about audiovisual semantic segmentation.

## Programming Competencies

| | |
|---|---|
| Fluent in | Python (readability at Google) |
| Previous use | C++, C, CUDA, MATLAB, Lua, Java, C#, HTML, Javascript, CSS |
| Libraries | JAX, PyTorch, Tensorflow, Flax, Apache Beam, OpenCV, Boost, NumPy, Flask, Bootstrap |
| Github | `https://github.com/anuragarnab` |
| Open Source | Developer of `Scenic` library for Computer Vision research using `JAX`. |

## Service and Outreach

2017 - present **Peer reviewing for conferences and journals**
Area Chair for CVPR, NeurIPS, ICLR and ICCV.
Action Editor for TMLR.
Three outstanding reviewer awards at NeurIPS and ECCV.

July 2024, 2025 **Guest Lecture at African Computer Vision Summer School**, Kenya and Rwanda
Invited lecture on deep learning architectures at both 2024 and 2025 editions of the summer school. Students voted my lecture the best in the summer school.

Nov. 2018 - present **Mentor**, *Black in AI*, Remote
I have advised eight mentees, and helped them get accepted to graduate school or begin industry-careers in computer vision and machine learning.

November 2022 **Guest Lecture at Deep Learning Indaba X**, Tanzania
Invited lecture on multitask learning at summer school.

June 2021 **Guest Lecture at Deep Learning Indaba X**, Tanzania
Invited lecture on transformer models in computer vision at summer school.

2013 - 2016 **Developer**, *Hyperion Development*, South Africa
Member of a student-run organisation which provided charitable teaching and training in computer programming. Primary role was developing the content of a free online course teaching Python, and also tutoring this course. Over 10 000 South African students took this course, and it was funded by grants from The Turing Trust, Facebook, Google and the Python Software Foundation among others.