Samuel Tonks

https://samueltonks.com/

Experienced Machine Learning (ML) researcher with proven expertise in a broad range of deep learning based approaches to medical image translation tasks with peer-reviewed publications at AI for bioimaging conferences. Demonstrable experience in designing and implementing, full SOTA AI/ML solutions, with high-resolution images at industrial scale within GSK. Affiliated with The Alan Turing Institute, the UK's leading institute of Artificial Intelligence (AI) as a previous enrichment student and ongoing core developer for Scivision [1] an open-source computer vision (CV) package.

EXPERIENCE

Deep Learning at Marine Biological Lab (DL@MBL)[2]

Teaching Associate

Recruited to provide expertise on generative modelling approaches to image translation.

- **Responsibilities**: Develop repository for students to explore generative modelling approach to image translation for virtual nuclei and cytoplasm staining. Partner with industry colleagues from Chan Zuckerberg Biohub to deliver session on regression vs generative modelling approaches for image translation tasks.
- Achievements: Developed end-to-end image translation repository for students to learn about generative approaches and train a model from scratch.

The Alan Turing Institute

Core Developer

Awarded Turing enrichment scheme placement award [4] a 12 month program for Doctoral students to spend time at the Turing. I then continued my affiliation with the Turing as a core developer of Scivision, an open-source computer vision package designed at the Turing.

- **Responsibilities**: Core developer of Scivision[5] an open-source computer vision package developed at The Alan Turing Institute. Sourced new models and data sets, designed and developed code for improving automation and deployment, testing frameworks and docker implementations.
- Achievements: Implemented several CI/CD methods to improve maintenance of Scivision catalog. Designed and implemented docker solutions for multiple Scivision models resulting in a considerable reduction in setting up of model environments for different user systems.

GSK Pharma R&D Bioimaging Analytics

University Worker (Ph.D)

Nov 2020 - Present Collaboration as part of Ph.D research focused on industrial application of deep learning based approaches to virtual staining within GSK Pharma R&D Bioimaging Analytics. Implemented end-to-end (E2E) deep learning based solution for generating label-free fluorescence microscopy from unstained bright-field images.

- **Responsibilities**: Explore, design, develop and maintain E2E pipeline with SOTA generative modelling approaches for image translation applications such as virtual fluorescence and virtual h& staining on large datasets (5 million + images). Design relevant and meaningful quantitative methods to evaluate the performance of these methods tailored to the specific real-world application. Set up and conduct large-scale experiments to test hypotheses and drive product development. Partner with stakeholders within Bioimaging from different departments and time zones to define AI solutions to biological problems.
- Achievements: Designed and implemented E2E pipeline utilising advanced generative models with a focus on meaningful evaluation metrics for virtual fluorescence microscopy resulting in significant efficiency improvements. This included reducing the staining time from weeks to minutes and enabling multiple live cell time-point studies of a single plate instead of multiple plates. This research led to an oral presentation and paper at the International Symposium of Bioimaging 2023 [6]. Conducted large-scale analysis into the generalisability of virtual stain models identifying novel findings around how to utilise the diversity of bioimaging data to better train generalisable models resulting in considerable reductions in the number of models required to be trained for large-scale industry applications. [7]. Final paper currently under review has four components. I explore the posterior distribution of the biological features extracted from virtual staining samples extracted from generative models. I propose a new method Bright2BioNet that can directly predict a posterior distribution over biological feature values directly from unstained samples. I devise novel methods for evaluating this posterior and early works into the use of simulation with Bayes theorem to truly evaluated a predicted posterior distribution. Submission planned for August Bioimaging Journal AI Special Addition.

Massachusetts, USA. June 2024 - Present

London, UK.

Feb 2024 - Present

Philadelphia, USA.

University of Birmingham

Teaching Associate

- **Responsibilities**: Supporting undergraduate and postgraduate students on Data Structures & Algorithms, Artificial Intelligence, Computer Vision & Imaging, Neural Computation, Visualisation and Software Engineering & Professional Practises modules.
- Achievements: Designed and delivered multiple slurm and python based exercise workshops for different modules.

The Alan Turing Institute

Enrichment Student

- **Responsibilities**: Contribute ideas and code for improving Scivision functionality as well as further building the community of developers and users to improve the adoption of Scivision. Research and collaborate with other Turing Fellows and enrichment students to improve my Ph.D research.
- Achievements: Pitched Scivision at The Alan Turing Institute Artificial Intelligence UK (AIUK) conference 2023 PitchFest competition [8]. Planned and facilitated an all-day workshop on Uncertainty Quantification for Generative Models at The Alan Turing Institute inviting world-leading guest speakers from Google DeepMind, Microsoft and The Alan Turing Institute to facilitate collaboration and discussion of the latest developments within this field.

Haleon

Data Scientist

London, UK

London, UK

London, UK.

Jan 2022 - Jan 2023

Jul. 2018 - Oct. 2020

Recruited into GSK Consumer Healthcare (now Haleon) Market Analytics group for Northern Europe to implement ML-based forecasting methodologies, utilise automation to improve data gathering processes and provide commercial insights for the leadership team.

- **Responsibilities**: Research and implement suitable ML-based forecasting methodologies to improve forecasting accuracy. Improve operational efficiencies through automation of data extraction, modeling and visualisation. Collaborate with cross-functional teams of Senior Brand Managers, Finance Directors, and Tech to drive commercial decision-making.
- Achievements: Implemented E2E pipeline for data gathering, data processing, a deep learning based approach to forecasting (LSTM) and visualisation of results. This led to an increase in market forecast accuracy for brand teams leading to data-centric decision making and ultimately successful brand launches.

GSK Tech

Technical Associate

Jan. 2017 - May. 2018 Joined the Business & Technology GSK Future Leaders Program working on the Global ERP system role out across GSK.

- SAP Testing Framework: Lead the design and delivery of SAP Cyber Security Penetration Testing framework across GSK Global ERP system.
- **Robotic Process Automation (RPA)**: Part of the global team that established and ran the Center of Excellence (CoE) for GSK's enterprise-wide RPA strategy.

Doctor Care Anywhere

Analyst

London, UK Jan. 2016 - Dec. 2016

Birmingham, UK.

London, UK.

- Business Operations: Data exporting, manipulation and conversion in order to support a range of industry client set-ups.
- Business Analytics: Delivery of Data analysis and reporting (e.g. ad-hoc requests) using a combination of Advanced Excel, FetchXML, SQL and Power BI.

EDUCATION

Ph.D. Computer Science

University of Birmingham

• Focus: Exploring the application of ML-based image translation to bioimaging tasks.

• Advisors: Prof. Iain Styles [9] Dr. Alexander Krull [10] & Dr. Minh Doan [11]

M.Sc Business Analytics; Distinction

Imperial College Business School

• Focus: Deep learning for image analysis, ML, Linear Algebra, Statistics and Optimisation & Decision Models

BA Business & Management; 1st Class Honors

University of Portsmouth

Sep. 2011 – Aug. 2014 Portsmouth, UK.

Sep. 2018 – Aug. 2020

Nov. 2020 – Present (Oct. 2024)

Birmingham, UK.

Sep 2021 - June 2024

Skills

• Python, Pytorch, Tensorflow, Tensorboard, CUDA, Scikit-learn, Numpy, Pandas Matplotlib, Inkscape and Slurm.

Awards, Workshops & Invited Talks

- Data Science in Production with Python [12]: 2-day course from Cambridge Spark focused on productionising ML pipelines.
- Uncertainty Quantification for Generative Models [13]: Independently planned and facilitated an all-day workshop at The Alan Turing Institute inviting 4 guest speakers (Google DeepMind, Microsoft, The Alan Turing Institute) and a round table discussion of the latest developments within the field.
- "Evaluation of virtual stain models for high-throughput screening" [14]: talk given at International Symposium for Biological Imaging (ISBI) 2023

Selected Publications

- 1. Samuel Tonks, Chih Hsu, Steve Hood, Ryan Musso, Ceriden Hopely, Minh Doan, Erin Edwards, Alexander Krull, and Iain Styles. "Evaluation of virtual staining for high-throughput screenings". In 20th IEEE International Symposium on Biomedical Imaging. IEEE, 2023. [15]
- Alexander Krull, Hector Basevi, Benjamin Salmon, Andre Zeug, Franziska Müller, Samuel Tonks, Leela Muppala, and Aleš Leonardis. "Image denoising and the generative accumulation of photons." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1528-1537. 2024.[16]
- Samuel Tonks, Cuong Quoc Nguyen, Steve Hood, Ryan Musso, Ceridwen Hopely, Steve Titus, Minh Doan, Iain Styles, Alexander Krull "Can virtual staining for high-throughput screening generalize?" 2024.[17]