First Workshop on Multimodal AI

Tuesday, 27th June 2023

Social Media: #MultimodalAI23

Venue: The Edge, The Endcliffe Village, 34 Endcliffe Crescent, Sheffield S10 3ED (direction)

Programme

Time	Event
09:30 - 10:00	Registration and Morning Refreshments
10:00 - 10:10	Welcome and Introduction: Haiping Lu, The University of Sheffield
10:10 - 10:50	Keynote 1: Yutian Chen, Google DeepMind **Learning a Generalist Agent on Large Scale Multi-modal Data**
10:50 - 11:30	Morning 3-Minute Pitches
11:30 - 12:00	Break and Poster Session 1
12:00 - 12:40	Keynote 2: Mirella Lapata, The University of Edinburgh **Hierarchical 3D Adapters for Long Video-to-text Summarization**
12:40 - 12:45	Group Photos
12:45 - 14:00	Lunch and Poster Session 2
14:00 - 14:40	Keynote 3: Chew-Yean Yam, Microsoft **Us and AI: Redefining our Relationships with AI**
14:40 - 15:20	Afternoon 3-Minute Pitches
15:20 - 15:25	Final Voting for Best Pitch/Poster and Best Student Pitch/Poster Prizes
15:25 - 16:00	Panel Discussion
16:00 - 16:10	Prize Winner Announcement and Closing Initiative: Haiping Lu
16:10 - 17:00	Tea/Coffee and Networking Forums

Pitch session format: Pitches in each session will be presented in two groups as shaded below. After each group, questions will be taken from the pitch panel and the audience.

Morning Pitch Session (10:50 - 11:30)

Pitch Panel: Martin Callaghan, Sokratia Georgaka, Yukesh Marudhasalam

Name	Title
Xingchi Liu	A Gaussian Process Method for Ground Vehicle Classification Using Acoustic Data
Adam Wynn	BETTER: An automatic feedBack systEm for supporTing emoTional spEech tRaining
Yizhi Li	MERT: Acoustic Music Understanding Model with Large-Scale Self-supervised Training
Rui Zhu	KnowWhereGraph: A Geospatial Knowledge Graph to Support Cross-Domain Knowledge Discovery
Imene Tarakli	Robot as a Schoolmates for Enhanced Adaptive Learning
Sina Tabakhi	From Multimodal Learning to Graph Neural Networks
Nitisha Jain	Semantic Interpretations of Multimodal Embeddings towards Explainable AI

Afternoon Pitch Session (14:40 - 15:20)

Pitch Panel: Nitisha Jain, Olamilekan Shobayo, Fanqi Zeng

Name	Title
Lucia Cipolina-Kun	Diffusion Models for the Restoration of Cultural Heritage
Bohua Peng	Recent Findings of Foundational Models on Multimodal NLU
Thao Do	Social Media Mining and Machine Learning for Understanding Illegal Wildlife Trade in Vietnam
Sam Barnett	Machine Learning in the Multimodal Program
Mohammod Suvon	The Multimodal Lens: Understanding of Radiologists Visual Search Behavior Patterns
Ning Ma	Obstructive Sleep Apnoea Screening with Breathing Sounds and Respiratory Effort: A Multimodal Deep Learning Approach
Felix Krones	Multimodal Data vs Digital Biomarkers in Cardiomegaly

Keynotes

Keynote 1: Yutian Chen, Staff Research Scientist, Google DeepMind, AlphaGo Developer

Title: Learning a Generalist Agent on Large Scale Multi-modal Data

Time: 10:10 - 10:50

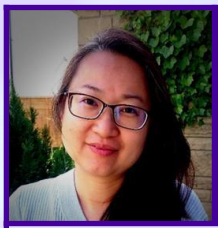
Abstract: The abundant spectrum of multi-modal data provides a significant opportunity for augmenting the training of foundational models beyond mere text. In this talk, I will introduce two lines of work that leverage large-scale models, trained on Internet-scale multi-modal datasets, to achieve good generalization performance. The first work trains an audio-visual model on YouTube datasets of videos and enables automatic video translation and dubbing. The model is able to learn the correspondence between audio and visual features, and use this knowledge to translate videos from one language to another. The second work trains a multi-modal, multi-task, multi-embodiment generalist policy on a massive collection of simulated control tasks, vision, language, and robotics. The model is able to learn to perform a variety of tasks, including controlling a robot arm, playing a game, and translating text. Both lines of work exhibit the potential future trajectory of foundational models, highlighting the transformative power of integrating multi-modal inputs and outputs.

Keynote 2: Mirella Lapata, Professor of Natural Language Processing, University of Edinburgh and UKRI Turing AI World-Leading Researcher Fellow

Title: Hierarchical3D Adapters for Long Video-to-text Summarization

Time: 12:00 - 12:40

Abstract: In this talk I will focus on video-to-text summarization and discuss how to best utilize multimodal information for summarizing long inputs (e.g., an hour-long TV show) into long outputs (e.g., a multi-sentence summary). We extend SummScreen (Chen et al., 2021), a dialogue summarization dataset consisting of transcripts of TV episodes with reference summaries, and create a multimodal variant by collecting corresponding fulllength videos. We incorporate multimodal information into a pretrained textual summarizer efficiently using adapter modules augmented with a hierarchical structure while tuning only 3.8% of model parameters. Our experiments demonstrate that multimodal information offers superior performance over more memory-heavy and fully fine-tuned textual summarization methods.



Keynote 3: Chew-Yean Yam, Principal Data and Applied Scientist, Microsoft

Title: Us and AI: Redefining our Relationships with AI

Time: 14:00 - 14:40

Abstract: The rapid advancement of AI has transformed how we interact with intelligent machines. Unravel the dynamic shifts in human-AI relations across diverse roles that we play in our society. Spark your imagination and seize the power to sculpt this new relationship that is meaningful to you.

Posters

Name	Title
Abdulsalam Alsunaidi	Predicting Actions in Images Using Distributed Lexical Representations
Bohua Peng	Recent Findings of Foundational Models on Multimodal NLU
Chenghao Xiao	Adversarial Length Attack to Vision-Language Models
Chenyang Wang	A Novel Multimodal AI Model for Generating Hospital Discharge Instruction
Christoforos Galazis	High-resolution 3D Maps of Left Atrial Displacements Using an Unsupervised Image Registration Neural Network
Douglas Amoke	Multimodal Data and AI for Downstream Tasks
Jayani Bhatwadiya	Multimodal AI for Cancer Detection and Diagnosis : A Study on the Cancer Imaging Archive (TCIA) Dataset
Jiachen Luo	Cross-Modal Fusion Techniques for Emotion Recognition from Text and Speech
Jingkun Chen	Semi-Supervised Unpaired Medical Image Segmentation Through Task-Affinity Consistency
Lucia Cipolina-Kun	Diffusion Models for the Restoration of Cultural Heritage
Nitisha Jain	Semantic Interpretations of Multimodal Embeddings towards Explainable AI
Prasun Tripathi	Tensor-based Multimodal Learning for Prediction of Pulmonary Arterial Wedge Pressure from Cardiac MRI
Raja Omman Zafar	Digital Twin for Homecare
Sokratia Georgaka	CellPie: A Fast Spatial Transcriptomics Topic Discovery Method via Joint Factorization of Gene Expression and Imaging Data
Wei Xing	Multi-Fidelity Fusion
Yichen He	AI in Evolution and Ecology
Yixuan Zhu	Potential Multimodal AI for Electroencephalogram (EEG) Analysis
Yizhi Li	MERT: Acoustic Music Understanding Model with Large-Scale Self-supervised Training
Yu Hon On	Automatic Aortic Valve Disease Detection from MRI with Spatio-Temporal Attention Maps

Panel Discussion (15:25 - 16:00)

Panel Members	Questions
• Yutian Chen • Arunav Das • Mirella Lapata • Roger Moore • Owen Parsons • Sophie Shang • Marta Varela • Chew-Yean Yam	1. What breakthroughs in multimodal AI do you foresee having the most significant impact in the next five years? 2. How can we navigate and mitigate the ethical concerns associated with advancing multimodal AI technologies? 3. Given the interdisciplinary nature of multimodal AI, how can we better integrate different fields of expertise to accelerate innovation in this area?

Networking Forums (16:10 - 17:00)

Forum 1: Envisioning MultimodalAI'24	**Facilitator:** Emma J Barker, Lei Lu	**Assistant:** Olamilekan Shobayo
	In this forum, attendees can discuss their hopes and expectations for the next year's event, which can guide organisers in their planning.	
Forum 2: Boosting Engagement and Active Participation in Multimodal AI	**Facilitator:** Tingyan Wang, Zhixiang Chen	**Assistant:** Pawel Pukowski
	This forum can provide a space for participants to share strategies and ideas for increasing involvement in multimodal AI, not just for students, but for professionals, hobbyists, and newcomers as well. This could include discussions around projects, collaborative opportunities, competitions, or other hands-on activities that can foster a deeper understanding and appreciation of multimodal AI.	
Forum 3: Cross-Disciplinary Collaboration and Resource Sharing in Multimodal AI	**Facilitator:** Sophie Shang, Arunav Das	**Assistant:** Lawrence Schobs
	In this forum, participants can discuss both the effective strategies for collaboration among various fields of expertise and the challenges and best practices associated with sourcing, creating, and using multimodal AI datasets. This discussion can foster a comprehensive dialogue on fostering innovation and effective resource management in multimodal AI.	
Forum 4: Open-Source Software Development for Multimodal AI	**Facilitator:** Shuo Zhou, Sina Tabakhi	**Assistant:** Rea Nkhumise
	This forum can focus on the importance of open-source software in multimodal AI, discussing the development, collaboration, and best practices in the field.	
Forum 5: Ethical and Responsible Practices in Multimodal AI	**Facilitator:** Luigi A. Moretti, Haolin Wang	**Assistant:** Nitisha Jain
	A forum focused on discussing the ethical implications of multimodal AI, considering responsible AI development and application.	

Parking at the Venue

We have a limited number of free parking permits available for parking at the Edge venue for those who require them. Please email multimodal-ai-enquiry-group@sheffield.ac.uk if you would like to receive a parking permit and have not already informed us of this need. Please note parking is on a first come first serve basis, and unfortunately, we are unable to reserve parking spaces.

Accessing the Internet

Complimentary WiFi is available throughout the venue. Connect to the network WiFiGuest by creating an account, with no password required (see instructions for more details). The Eduroam network is also available (connect to Eduroam).

Catering

Complimentary refreshments and lunch will be provided during the workshop. If you have specific dietary needs or catering questions, please speak to our team at the registration desk for assistance.

Filming and Photography

Please be advised that there will be media coverage, including filming and photography, during the workshop. The images and video taken may be used for promotional purposes on the Centre for Machine Intelligence and Department of Computer Science websites and social media channels. If you do not wish to appear in any video or photography, please inform the organisers via the contact multimodal-ai-enquiry-group@sheffield.ac.uk, or at the registration desk or speak to the photographer/ videographer.

Accessibility and Quiet Room

Please contact the organisers at multimodal-ai-enquiry-group@sheffield.ac.uk if you have any accessibility requirements that you would like to discuss, and we will endeavour to meet your requirements. There will be a quiet, private room available at the venue for delegates to use for various purposes. If you require access to the room, please contact the organisers via multimodal-ai-enquiry-group@sheffield.ac.uk or at the registration desk.

Sponsors and Partners

This workshop is jointly organised by University of Sheffield and University of Oxford under the Turing Network Funding from the Alan Turing Institute, with support from University of Sheffield's Centre for Machine Intelligence and Alan Turing Institute's Interest Group on Meta-learning for Multimodal Data (welcome to sign-up and join).

The Alan Turing Institute

University of Sheffield | Centre for Machine Intelligence

UNIVERSITY OF OXFORD