Discussion and recommendations from the 2023 "Gaze Meets ML" workshop breakout session

Alexandros Karargyris[*1]   AKARARGYRIS@GMAIL.COM
Shuaib Ahmed[†2]   SHUAIB.AHMED@MERCEDES-BENZ.COM
Ryan Anthony de Belen[†3]   R.DEBELEN@UNSW.EDU.AU
Bonny Banerjee[†4]   BBNERJEE@MEMPHIS.EDU
Timur Ibrayev[†5]   TIBRAYEV@PURDUE.EDU
Satyananda Kashyap[*6]   SATYANANDA.KASHYAP@IBM.COM
Elizabeth Krupinski[*7]   ELIZABETH.ANNE.KRUPINSKI@EMORY.EDU
Chenyi Kuang[†8]   KUANGC2@RPI.EDU
Paul Madu[†9]   PMADU@MY.APSU.EDU
Amarachi Blessing Madu[*10]   BMAMARACHI@VT.EDU
Silvia Makowski[†11]   SMAKOWSK@UNI-POTSDAM.DE
Athul M. Mathew[†12]   ATHULMATHEW.AM@GMAIL.COM
Tim Rolff[†13]   TIM.ROLFF@UNI-HAMBURG.DE
Bert Shi[†14]   EEBERT@UST.HK
Joy Wu[*6,15]   JOYTYWU@GMAIL.COM
Dario Zanca[*16]   DARIOZANCA@GMAIL.COM

[1]*MLCommons, USA*

[2]*Mercedes-Benz R&D, India*

[3]*University of New South Wales, Australia*

[4]*University of Memphis, USA*

[5]*Purdue University, USA*

[6]*IBM, USA*

[7]*Emory University, USA*

[8]*Rensselaer Polytechnic Institute, USA*

[9]*Austin Peay State University, USA*

[10]*Virginia Tech, USA*

[11]*Universität Potsdam, Germany*

[12]*Elm Company, Saudi Arabia*

[13]*University of Hamburg, Germany*

[14]*Hong Kong University of Science and Technology, Hong Kong*

[15]*Stanford University, USA*

[16]*FAU Erlangen-Nuremberg, Germany*

Abstract

The Gaze Meets ML (GMML) workshop at NeurIPS aims to bring together diverse machine learning communities to foster research that leverages eye gaze (visual attention) to fulfill synergy between human attention/cognition and machine learning model development and

---

* conceptualized & moderated

† contributed

evaluation. Towards this mission, the 2023 GMML workshop ran a breakout session to foster the research community by discussing open challenges. Three focus breakout session areas were identified through a selection process: Datasets, Community, and Vision and Actions for the Future. The findings and discussion points from this session were collected during the meeting and further organized and expanded after the meeting for efficient presentation here. The following sections detail each topic.

Keywords: gaze, visual attention, eye tracking, machine learning, reinforcement learning

1. Topic: Datasets

1.1. Background and motivation

Publicly available eye-tracking datasets are an integral part for fostering research and development. These datasets can be utilized for a wide range of applications and tasks such as gaze estimation, visual attention modeling, language understanding, robotic control, to domain specific tasks like autonomous driving, advertising and medical image diagnosis, etc. To better understand the state of the art the group first surveyed publicly available datasets and then summarized them in an open-source table available at https://gazemeetsml.org/. The specifications of the table (i.e., collected information) were discussed during the meeting, prioritizing analysis of limitations in existing gaze datasets to help with developing ideas for best practices in dataset creation. It is important to mention that this table was not an exhaustive list of available datasets but it rather aimed to provide an initial view of the current status of dataset resources used for eye gaze related research and development. As such, we welcome the research community to help us maintain and contribute to this table.

1.2. Current limitations and recommendations

During the summarization of over 30 publicly available eye gaze datasets, the GMML dataset discussion work group was able to reach a consensus: Improvement through standardization of dataset structure and documentation would be key to fostering advancement in this domain. Standardizing a common eye gaze dataset format has been a challenge both because researchers build datasets according to their requirements as well as due to the lack of standardized output gaze data from different eye tracking machines. Another important finding related to the acquired eye gaze data which may range from raw to post-processed gaze data, with or without the preprocessing code. This hinders both research reproducibility and the ability to maximally leverage the potential problems that the released datasets can address.

The GMML dataset discussion work group recommended the creation of a Common Data Model (CDM) following popular standardized dataset formats such as COCO, which tackles many machine learning tasks relevant for GMML. Recommendations include but are not limited to:

1. A metadata schema that integrates diverse information related not only to eye gaze but also to machine learning. Examples include license type, contributors, intended applications, eye tracker type and settings, gaze data collection protocol, participants' demographics, citations, code repository URLs, preprocessing workflow, machine learning model weight URLs, other documentation etc.

2. Inclusion of raw (unprocessed) gaze data with reformatting to a common data structure.

3. A standard way to index input images or other non-gaze data for both static and dynamic (i.e. video) data.

4. Inclusion of annotation data (e.g. bounding boxes, points, masks, etc) in a COCO-like format to facilitate re-use of resources between research communities.

No dataset is perfect. However, given the high monetary cost and time inefficiency often required to build large machine learning datasets, the GMML community believes that the community needs to develop a common standardized data structure.

2. Topic: Community

2.1. Background and motivation

Building a community in eye-tracking and human attention modeling is crucial for fostering collaboration, and promoting the development of standardized practices and shared datasets. A community plays a central role by facilitating the dissemination of findings and encouraging the adoption of innovative technologies across various fields.

2.2. Community tools and guidelines

In the context of advancing eye-tracking and visual attention modeling research in AI/ML, a targeted community-building strategy involves the following steps and tools: First, define the vision and goals of the community, outlining objectives such as sharing research, fostering collaboration, and facilitating learning in the field. Second, identify core members and stakeholders for governance, emphasizing those actively involved in eye-tracking-related publications, patents, or product development. Specify governance structures, roles, and policies to ensure effective community management. Third, select communication platforms tailored to the domain, including forums, mailing lists, and social media groups specifically focused on eye-tracking technologies. Fourth, organize events such as workshops and webinars centered around visual attention research, utilizing event management tools to streamline coordination. Fifth, promote knowledge sharing by showcasing members' expertise, publications, and datasets related to eye-tracking on the community website. Leverage content management tools for efficient collaboration and open peer review. Sixth, support professional development by sharing job opportunities and funding specifically related to eye-tracking research. Seventh, advertise the community through tutorial videos, presentations, and guest speaker talks focused on eye-tracking methodologies. Finally, actively solicit feedback from members using survey tools to continually enhance the community's relevance and impact in the realm of eye-tracking and visual attention modeling.

2.3. Challenges

The field of eye-tracking and human visual attention research is multifaceted, with several distinct communities contributing to its advancement. One prominent community is

researchers in cognitive psychology, focusing on understanding the underlying cognitive processes associated with gaze behavior and attention. Another community comprises computer vision and machine learning specialists who leverage eye-tracking data to develop algorithms and applications, such as gaze-based human-computer interaction. Additionally, there are communities within neuroscience exploring the neural mechanisms underlying visual attention. Each community often employs distinct terminology and methodologies, making collaboration and communication challenging.

To bridge the gaps and enhance collaboration among these communities, a potential solution is the establishment of interdisciplinary conferences and workshops that bring together researchers from different backgrounds. These events could facilitate knowledge exchange, foster a shared understanding of terminology and methodologies, and encourage collaborative projects. Additionally, the creation of standardized datasets that incorporate diverse aspects of eye-tracking and visual attention could provide a common ground for researchers to benchmark their methods and findings. This approach would not only promote a more cohesive understanding of gaze behavior across disciplines but also pave the way for more effective interdisciplinary collaborations in advancing the field.

2.4. Community sustainability

Collaboration with technology companies invested in augmented reality, virtual reality, or human-computer interaction could lead to mutually beneficial partnerships. These sponsors could benefit from insights into cutting-edge research, product testing opportunities, and access to a talent pool well-versed in eye-tracking technology. By diversifying revenue streams and establishing partnerships aligned with the community's goals, sustainable growth and development can be achieved. Sustainability strategies for an eye-tracking community extend beyond traditional sponsorship avenues. To foster financial stability, the community could explore membership models, where researchers, institutions, or industry professionals pay a fee for exclusive access to resources, events, and collaborative opportunities. Offering training programs or certification courses in eye-tracking methodologies could attract participants willing to pay for specialized education. In terms of sponsorship, companies in the eye-tracking hardware and software industry would find value in supporting the community and gaining exposure to a network of potential users and researchers.

3. Topic: Vision and Actions

3.1. Background and motivation

Gaze patterns provide valuable insights into cognition, attention, and user intentions, offering a natural and intuitive means of interaction with machines without input through gaze-assisted interfaces. Gaze tracking aids in personalization efforts, improving experiences by delivering tailored content and interactions. It enhances accessibility, particularly for individuals with mobility limitations, and enhances augmented and virtual reality (AR/VR) experiences by simulating natural eye movements and interactions, increasing realism. Integrating gaze and AI is driving significant innovations, with adaptive interfaces and immersive environments to offer superior user experiences.

3.2. Discussion

Our eyes and brains are amazingly good at what they do, but are often easily fooled (e.g., visual illusions). Medical imaging relies on human visual perception and cognition to render accurate and efficient diagnoses. Unfortunately, errors occur (false negatives and positives) that impact patient care. Errors occur for many reasons (e.g., poor image quality, experience, cognitive biases), leading to decision variability.

For over 30 years, researchers have explored the potential of computers and image analysis techniques to aid provider decisions, with artificial intelligence (AI) emerging as the most recent platform. No AI schemes developed to date are 100% accurate and error free. Thus, we must improve our understanding of human-computer interfaces and how to optimize the synergy between human decision makers and computer-based decision aids.

Since medical imaging is a visual modality, we can capitalize on the capabilities of the human visual system. Can we use gaze for AI evaluation? Can gaze assess the impact of AI on clinical decision making? Can it determine the best way to present AI information to users?

To assess the impact of AI on clinical decision-making, one key question is what type of AI output will best convey information. Should we use heatmaps, probability graphs, arrows pointing to abnormalities, a bounding box or outline? Efficiency is also critical – if it takes more time to use AI without gains in accuracy it will not be used.

For example, how do we best present AI to radiologists reviewing chest images for COVID and its severity level? Clinically, radiologists first search images without AI and make their initial assessment. They can then "turn on" the AI. The form of AI output (heatmap in Figure 1 vs heatmap plus a probability graph in Figure 1) impacts decisions and review time. Even radiologists with different levels of training may be impacted differently, which could be important for how we train residents and how much we expose them to AI to reduce the impact of potential deskilling.

If eye gaze technology can be used unobtrusively during clinical practice it could enable synergies between users and AI. We could track the gaze when evaluating a brain MRI glioblastoma changes (Figure 3) and analyze it in real time for automatic annotation for an AI system to analyze areas where the radiologist spent significant time viewing. Using this paradigm, it may be possible to train a convolutional neural network to use both image and human gaze data to create outputs.

3.3. Recommendations

During the session the GMML discussion work group was able to propose the following recommendations for the broad community:

1. Develop a database dedicated to tracking state-of-the-art methods in gaze-assisted machine learning and related domains to serve as a repository for the latest advancements, techniques, and approaches (e.g., research papers, algorithms, software). This will facilitate collaboration and dissemination in the research community and promote innovation and advancement in the field.

2. Create a taxonomy for gaze-assisted machine learning to foster understanding and clarity and enhance terminology consistency across publications and meetings.
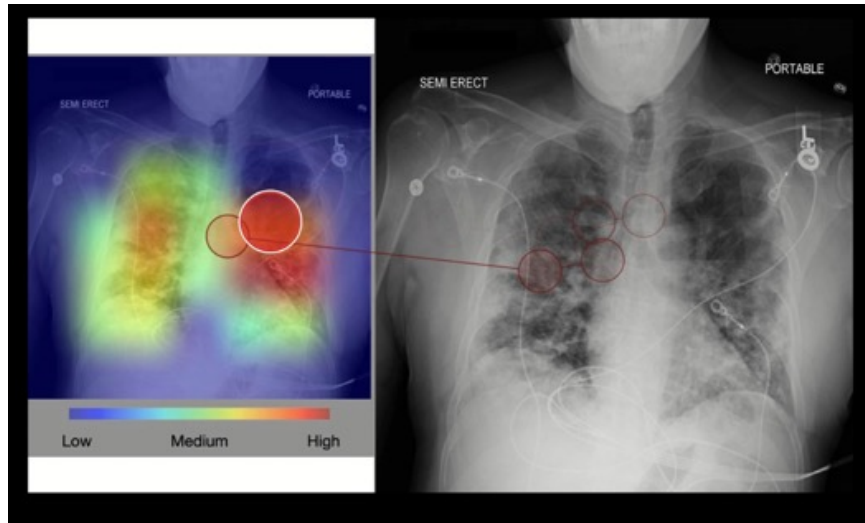
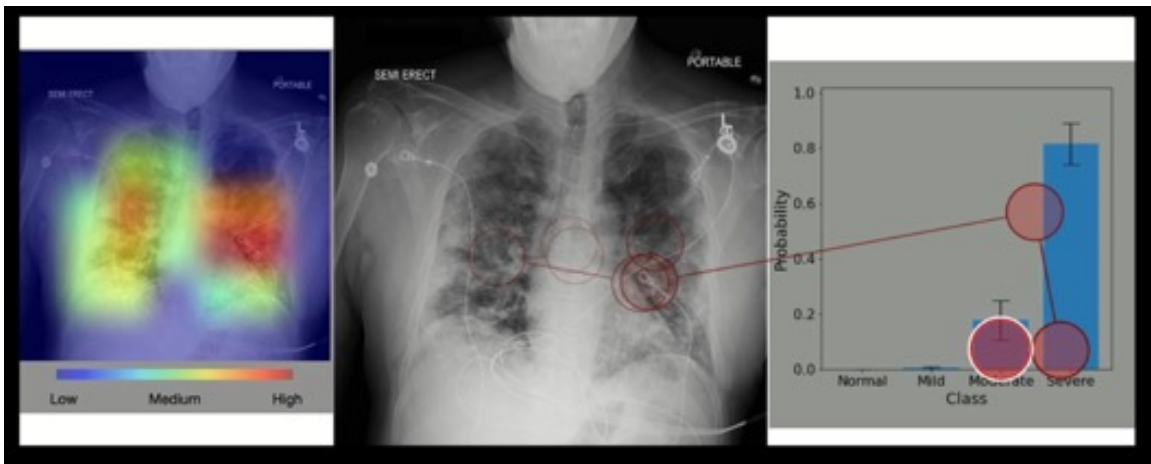Figure 1: Radiologist's gaze searching chest x-ray for COVID with an AI heatmap.

Figure 2: Radiologist searching chest x-ray for COVID with AI heatmap + probability graph.
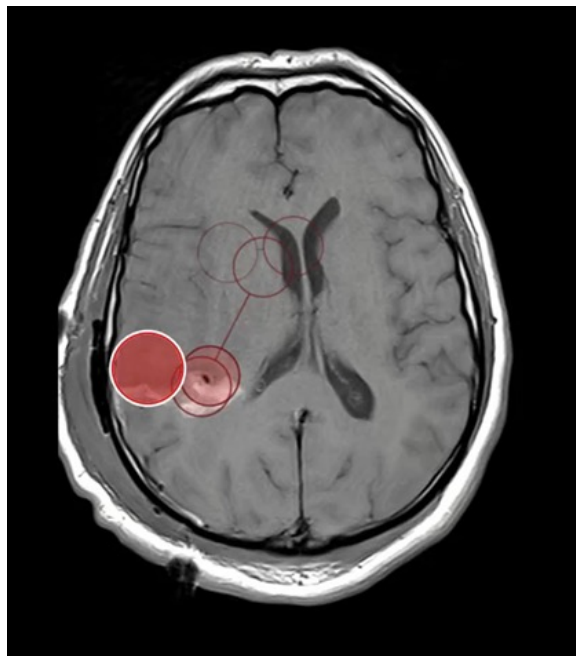
Figure 3:   Eye-gaze used to annotate images and train a convolutional neural network.

4. Conclusions and recommendations

The breakout session series at the GMML workshops has proven to be an excellent venue for experts from various backgrounds to participate in open discussions, brainstorming, and effective ideas exchange related to eye gaze. Through an open topic selection process the researchers have the unique opportunity to discuss topics that impact their community and cross pollinate ideas. This year's topics were not such an exception. In contrast to last year the groups were encouraged to provide recommendations for improvement of the current state of the art in eye gaze space. While each section above provided details of the discussions the main recommendations have been further refined and are summarized below:

1. The GMML community believes that ***there is a need to improve the current state of eye gaze datasets to help drive research and development effectively***. Specifically, the GMML community supports an open discussion within the broad eye gaze community (academia and industry) on proposing solutions for eye gaze dataset interoperability and standardization. Solicitation for feedback and discussion can be achieved through an iterative process like Delphi to reach a community consensus with specific requirements and deliverables in a transparent and neutral way. This can add gravity and credibility to the proposed specifications as well as adoption which is a prerequisite of success.e.

2. The GMML community underscores the importance of an open community of experts to help research and development in eye gaze space. ***Towards this the recommen-***

dation of the group is the establishment of a community portal (e.g. wiki) that aims to foster the community. Specific recommendations for this portal include, but not limited to, a list of related datasets and state-of-the-art methods continuously updated, webinar series hosting, visual attention terminologies, related job announcements and a taxonomy of knowledge related to eye gaze (i.e. visual attention, psychology, AI/ML). The group believes that such an open portal can help sustain participation and engagement of the community with direct outcomes (i.e. knowledge dissemination).