AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

Theme Development Workshop: Trusted AI - The future of creating ethical and responsible AI systems

Breakout Session: **AI explainability for vision tasks**

Session organizer: Vasileios Mezaris (CERTH)

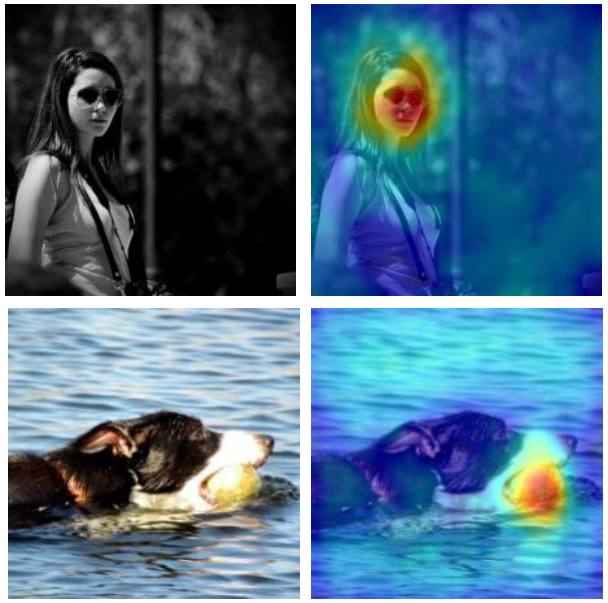
13 Sept. 2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

Breakout Session goals

- Discuss:
  - The current capabilities of AI explainability methods for visual data classifiers;
  - How explanations can be presented to the data scientist / end-user;
  - What we can expect to understand from the provided explanations; and,
  - The next steps towards advanced AI explainability

A few questions to start with…



M. Ntrougkas, N. Gkalelis, V. Mezaris, "TAME: Attention Mechanism Based Feature Fusion for Generating Explanation Maps of Convolutional Neural Networks", Proc. IEEE ISM 2022. DOI:10.1109/ISM55400.2022.00014; https://github.com/bmezaris/TAME.

I. Gkartzonika, N. Gkalelis, V. Mezaris, "Learning Visual Explanations for DCNN-Based Image Classifiers Using an Attention Mechanism", Proc. ECCVW 2022. DOI:10.1007/978-3-031-25085-9_23; https://github.com/bmezaris/L-CAM.

- For image classification tasks, explanations often come in the form of heat maps: this is useful, but is it the best we can do?
- How about other visual tasks (e.g. video understanding, video summarization) – what explanations can we / need to generate?
- Has our community converged to established protocols for evaluating the goodness of explanations?
- What are the most demanding / promising usages of such explanations? Who needs them? (just AI engineers?)
- Are our explanations good enough for these applications?

Our Invited Experts

- Prof. Jenny Benois-Pineau, Université de Bordeaux, France
- Dr. Gabriele Ciravegna, INRIA, Université Côte D'Azur, France

AI4media

Follow us

@ai4mediaproject

www.ai4media.eu

info@ai4media.eu

Our Consortium

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911