An Eye-Tracking Database of Video Advertising

Lucie Lévêque[1] and Hantao Liu[2]

[1]Xi'an Jiaotong Liverpool-University, China; [2]Cardiff University, United Kingdom

Introduction

- **Multimedia** systems have become an integral part of human activity, including entertainment, education, security, and medicine. Humans indeed rely upon visual media to communicate, and it is thus critical to understand **how observers experience visual media**.
- **Eye-tracking**, the process of measuring where people look, has been widely used to reveal multimedia experience. **Saliency maps**, which represent stimulus-driven, bottom-up visual attention, are obtained from the recorded fixations and indicate conspicuousness of scene locations [1].
- The eye-tracking technology can be used in the **commercial sector** to provide evidence of human behaviours. For instance, **video advertisers** need to make sure that potential consumers notice the advertised product while experience the video content (i.e., storytelling).

Eye-tracking experiment

- Our dataset consists of **40 frames** extracted from 40 online **video advertisements** collected on YouTube from diverse content: Animation, Celebrity, Indoor, and Outdoor. Fig. 1 illustrates sample stimuli used.
- The videos provide a wide range of complexity in terms of the **spatial position** of the advertised product (e.g., close of far from the centre).

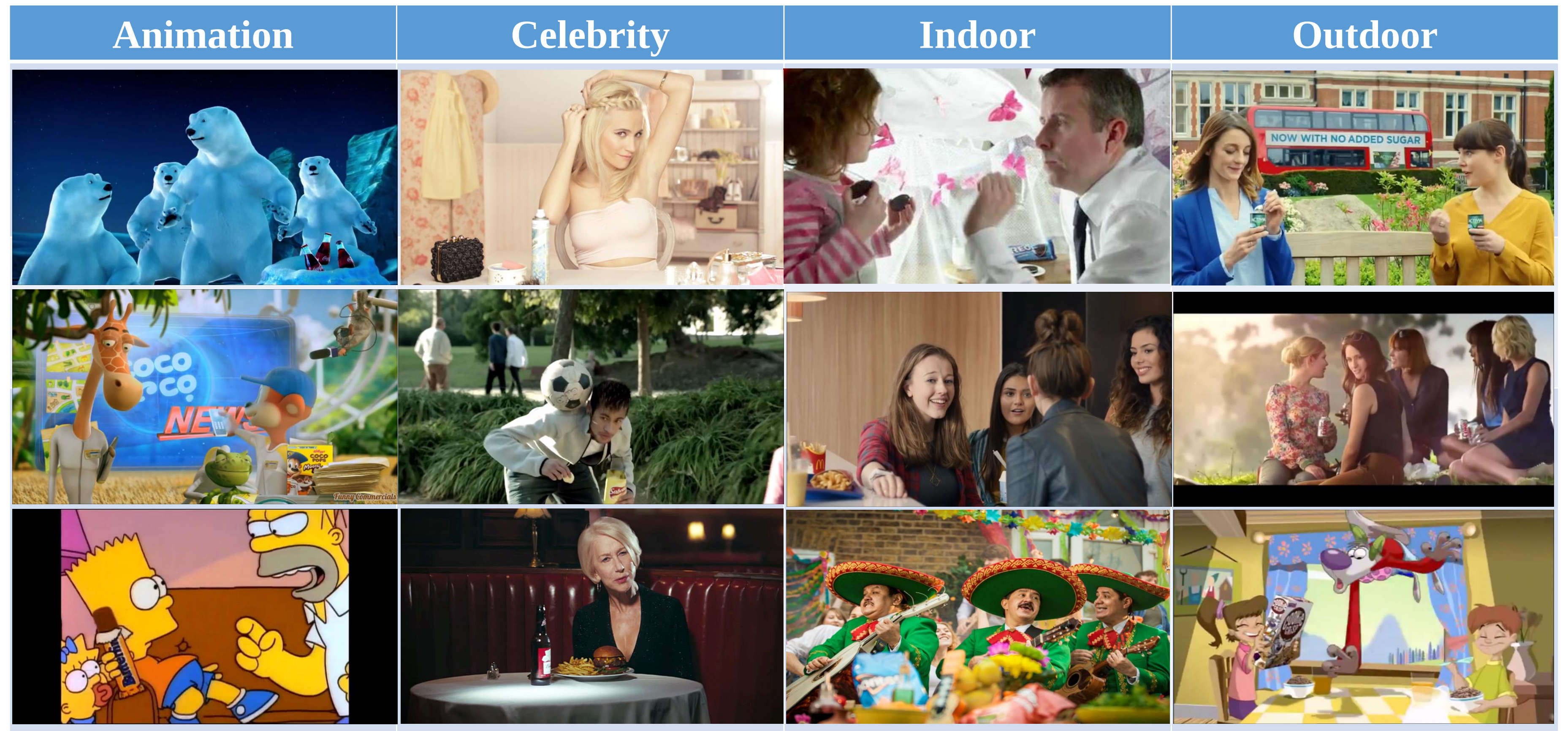

Fig. 1: Illustration of sample stimuli used in our experiment.

- The test stimuli were displayed on a 19-inch LCD monitor with a native resolution of 1080x1920. The eye movements of the observers were recorded using a **SMI Red-m** advanced eye-tracker (sampling rate: 250 Hz, spatial resolution: 0.1 degree, gaze accuracy: 0.5 degree).
- The participants were asked to experience the stimuli in a **natural way** ("view it as you normally would"). Each stimulus was displayed for **one second**, to simulate the reality that viewers always fast-forward through adverts. **28 participants**, including 15 females and 13 males, 18 university students and 10 professionals, participated in the experiment.

Experimental results

- **Fixations** were directly extracting from the raw eye-tracking data using SMI BeGaze Analysis software package. A fixation was rigorously defined using the dispersal and duration based algorithm [2].
- Fixations are accumulated over all 28 subjects to render a **topographic saliency map** for a given stimulus, with each fixation giving rise to a grey-scale patch simulating the foveal vision of the human visual system. The activity of the patch is modelled as a Gaussian distribution of which the standard deviation approximates the size of the fovea. Fig. 2 represents the saliency maps created for two sample test stimuli in our dataset. In general, it can be seen that the salient regions correspond to the **storytelling** (i.e., character). Viewers as well showed a good performance in fixating at the **target product** (e.g., bottle of water).

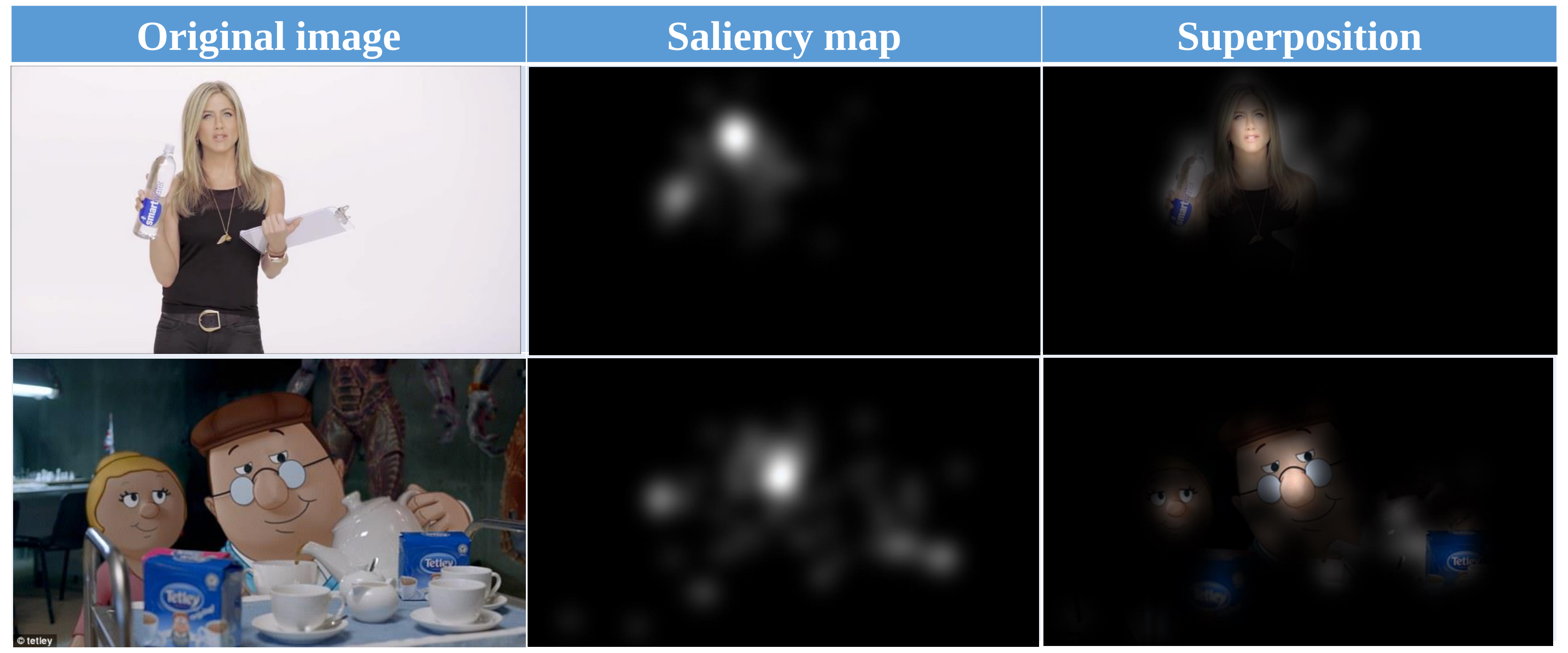

Fig. 2: Illustration of the saliency maps created for two sample stimuli.

Discussion

- Eye-tracking is expensive, cumbersome, and impractical in many circumstances; **computational saliency** is a more realistic way to use visual attention. Saliency models have been developed for different applications [3].
- We carried out an evaluation with five state-of-the-art **saliency models**: AIM, AWS, GBVS, Itti, and RARE2012. Fig. 3 represents the computational saliency maps generated for two sample test stimuli. It can be noticed that the models fail in matching with the eye-tracking data, which was further studied using three **similarity metrics**, CC, NSS, and AUC, as in Fig. 4.

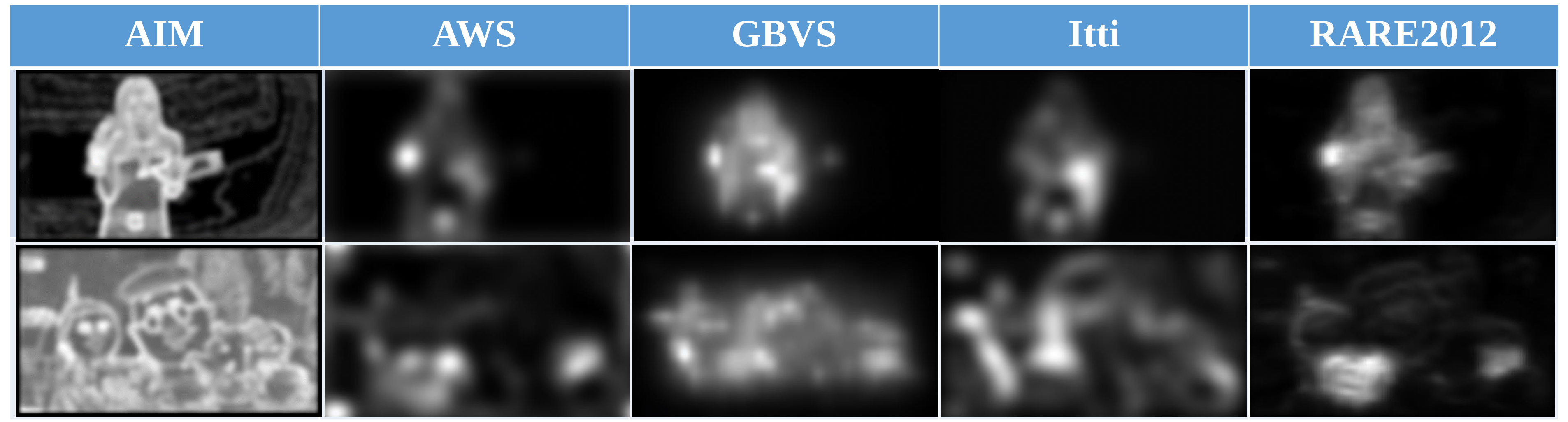

Fig. 3: Illustration of the saliency maps generated by five saliency models for the two stimuli in Fig. 2.

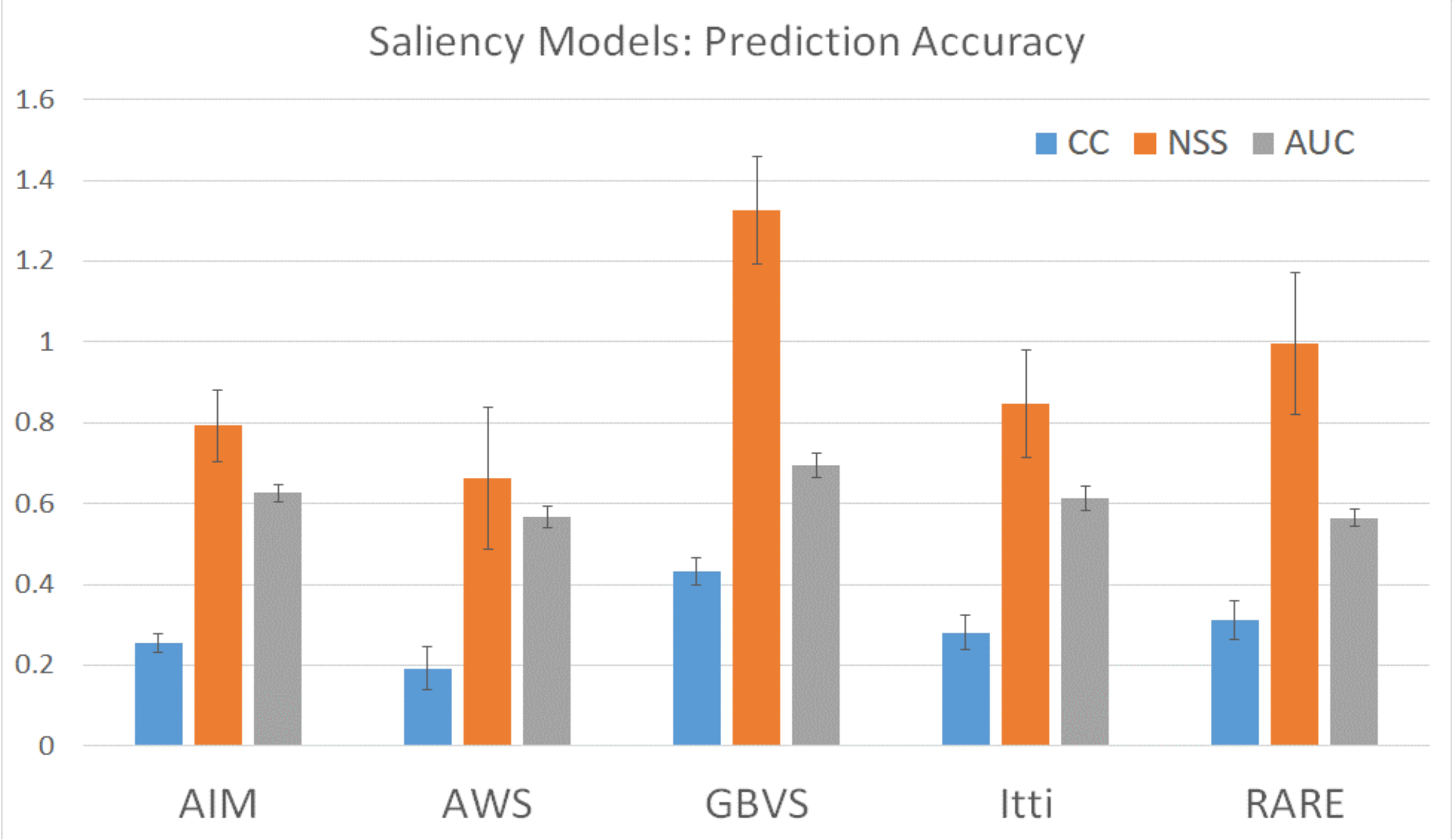

Fig. 4: Illustration of the similarity between human and modelled saliency over the 40 stimuli.

Bibliography

[1] H. Liu, and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data", *IEEE TCSVT*, vol. 21, 2011.

[2] W. Zhang, and H. Liu. "Toward a reliable collection of eye-tracking data for image quality research: challenges, solutions, and applications", *IEEE TIP*, vol. 26, 2017.

[3] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations", *MIT Technical Report*, 2012.