FAQ

For more questions, please feel free to contact salient360@univ-nantes.fr

Q: How do I get access to the tools and dataset

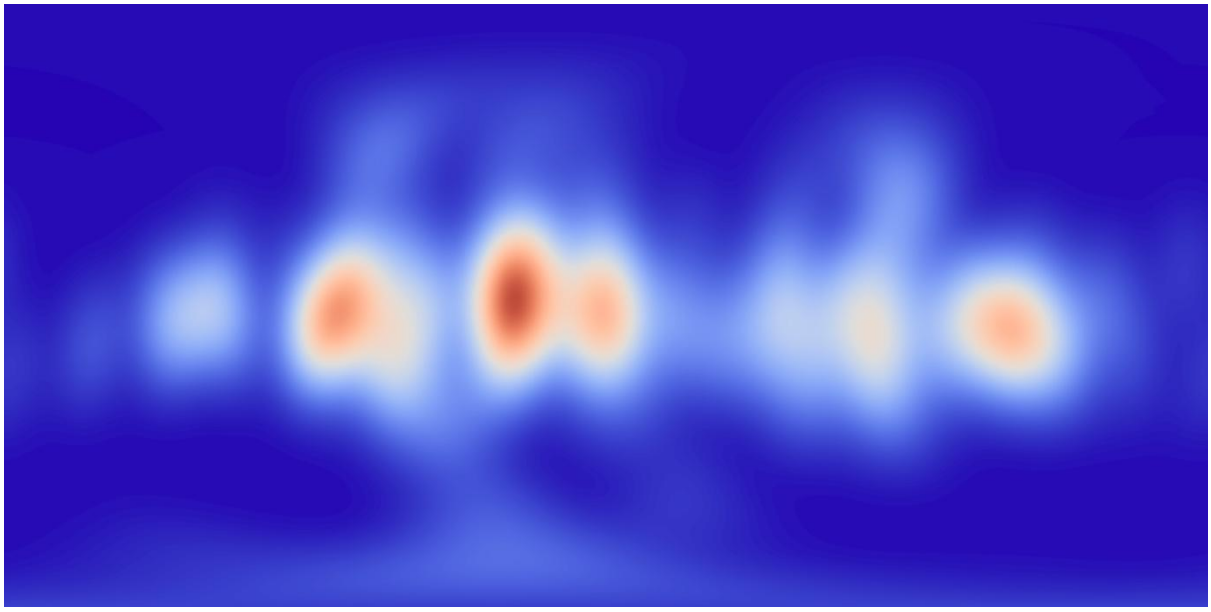
All data and tools are on a repository accessible through FTP site (Connection details will be sent to interested Parties/researchers to participate).

Interested Parties/researchers should express their interest sending an email to salient360@univ-nantes.fr

Q: What type of model can be submitted?

Participants are free to submit computational models in the following category (Each of the submissions will be evaluated with the respective collected ground-truth by the organizers):

1. Head motion based saliency model (Model type 1): these models are expected to predict Ground Truth Heat Map (GTHM) derived from the "movement of the head" only. The output of the submitted model must be a heat map in the equirectangular space (see section C of Annex). The GTHM are generated from head motion tracking data. For each sample head position, we consider the center of the view port. In the rectilinear viewport domain, a two degree Gaussian foveation filter centered in the center of the view-port is applied. For details on the Methods to evaluate the performance (see section B of Annex). A sample of such heat map in equirectangular is shown in Figure 1.



1. ***Sample head based saliency map SH1.jpg in the equirectangular space that illustrates the various salienct areas in the image***

2. (head+eye)-motion based saliency model (Model type 2): these models are expected to predict Ground Truth Heat Map (GTHM) derived from the "movement of the head" as well as from the "movement of the eye within the viewport". The output of the submitted model must be a heat map in the equirectangular space (see section C of Annex). For each sampled eye position, a two degree gaussian foveation filter centered at the fixation position (in the rectilinear viewport domain) is applied. A sample of such heat map in equirectangular is shown in Figure 2. For details on the Methods to evaluate the performance see (see section B of Annex).

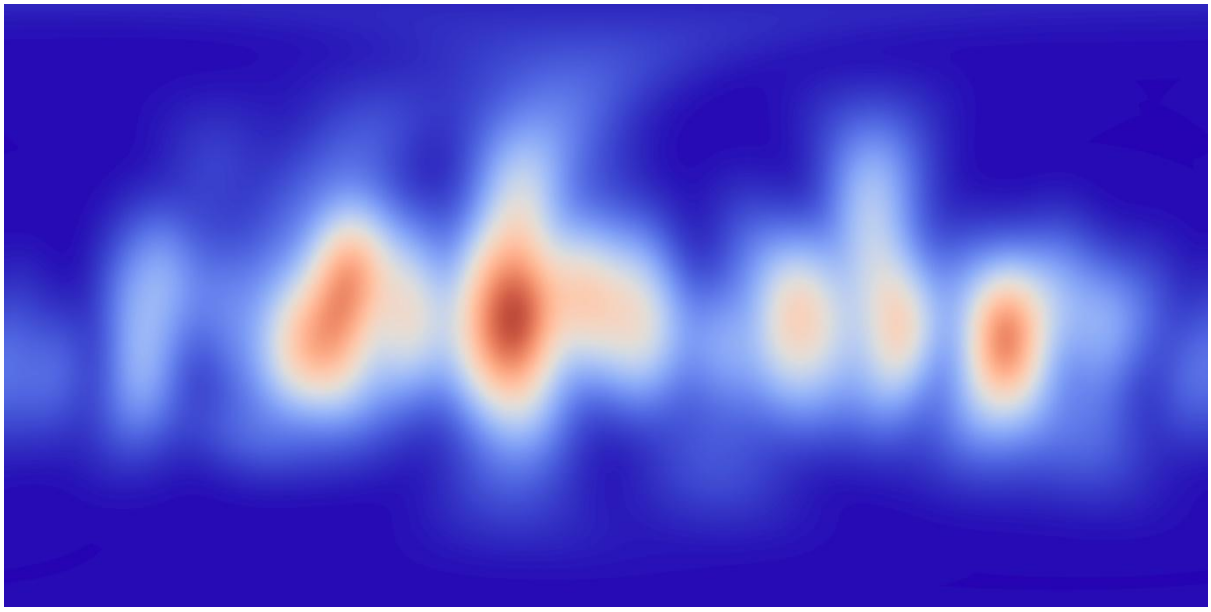


Figure 2: Sample head+eye based saliency map SHE1.jpg that illustrates the various saliency maps in the image

3. Scan-paths of observers in the entire 360 panorama (Model type 3): the models are expected to predict the Groundtruth scan-path (GTSP) that are obtained from the head and eye-movement data from several observers. Each 360 stimuli, therefore has as many scan-path vectors as the number of observers N who have watched the image, the length of each scan-path being 25 seconds, which is the amount of time the observers were shown the stimuli. The models from the participants are therefore also expected to output N scan-paths, and these sets of N predicted scan-paths will be compared with the set of N scan-paths from the ground truth (see section E of Annex).

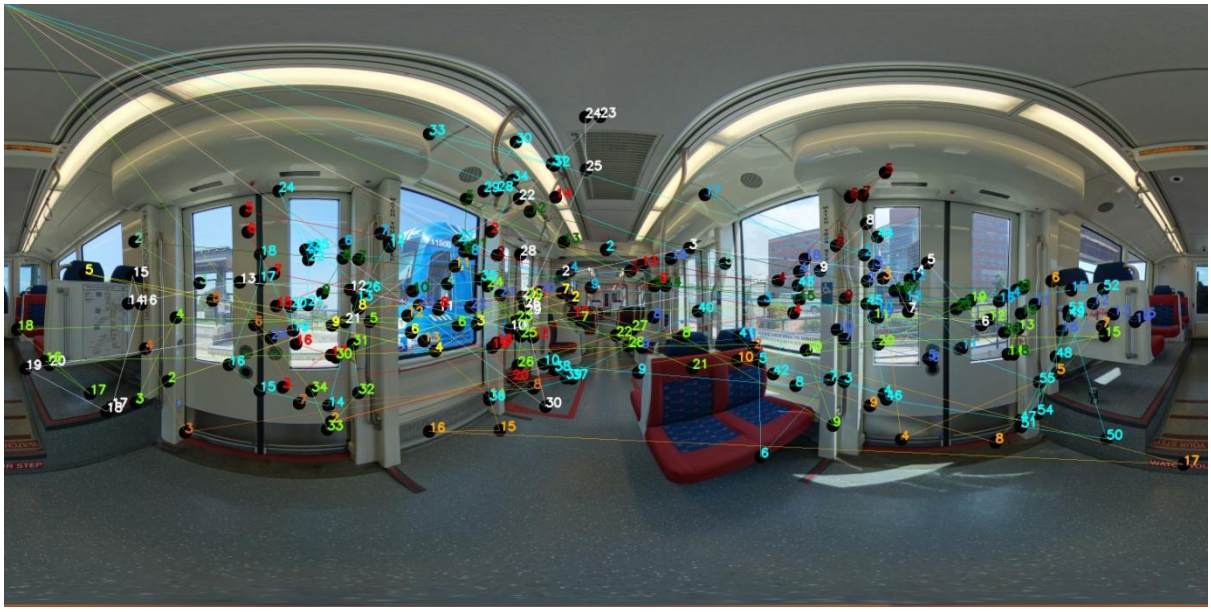


Figure 3: Sample scan-path file SP1.jpg that illustrates the scan-paths of several observers(each observer is in a different color) mapped back onto the equirectanglar image

Q: How can I start to test my model?

The organizers are providing several tools:

1. a Test Dataset: for each type of model (1, 2 and 3), the organizers disclose some original images, test conditions as ground truth data in corresponding output space for each type of model ;
2. ground truth data and parsing tools: Ground truth data for all the 3 types of models and the tools to parse them will be provided to parties who have expressed their interest to participate to the grand challenge. More details can be found in Section A of Annex ;
3. evaluation methods and tools: Evaluation tools for all the 3 types of models will be provided to parties who have expressed their interest to participate to the grand challenge. More details can be found in Section B of Annex.

Q: What is the Test Dataset?

The test dataset is composed of sub dataset (one per Model type) disclosed by the organizer so participants can develop/test their models.

The Dataset have been obtained in the following conditions:

- The head mounted display(HMD) Oculus-DK2 was used for this test. It has a frame refresh rate of 75Hz, resolution of 960x1080 per eye and a total viewing angle of 100 degree in both horizontal and vertical directions. The gyroscopic sensors within the device are able to transmit the orientation data at a rate equal to the device frame refresh rate. A small eye-tracking camera from Sensomotoric Instruments(SMI) was integrated into the device and was able to transmit eye-tracking data binocularly at 60Hz. The software had a feature to check for calibration accuracy every two minutes and re-calibrated each time if necessary.
- To maintain a natural (free-viewing like) gaze pattern, subjects were made to view the scene normally without the need to provide explicit quantitative measurements in HMD . They were instructed to watch the scene as normally as possible with a combination of head and eye-movement. As naive Observers may have unnatural behavior when exposed for the first time to omni directional visualization in HMD, a training session was run beforehand in order to familiarize them with such experience.
- Each 360 images were shown for 25 seconds and there was a 5 second gray screen between two images. The observers were themselves seated comfortably in a turn-chair and were free to rotate the full 360 degrees and also move the chair within the

room if necessary. **The position of each 360 image was reset to the equirectangular image center each time a new 360 image appeared** (irrespective of their position). This was done to ensure that all observers start at the same starting position in the panorama.

where to find it	number of images	Model
ftp provided to participants	20	Head motion based saliency (Model 1)
ftp provided to participants	40	Head+Eye motion based saliency
ftp provided to participants		Scanpath models

Q: How will my model will be evaluated?

In addition to the test dataset, an evaluation dataset composed of 20 images with ground truth data for each type of model remain undisclosed till submission of the model.

The evaluation is performed with the same tools as the one distributed to the participant. Each participant will receive all performance indicator corresponding to the submitted model.

Participant can submit their model till the deadline for submission.

Results will be released one week after the model submission deadline for all models, irrespective of their respective time of submission.

Participants are invited to submit a paper by the paper submission deadline. It is advised to prepare a draft of the paper before getting

the results from the organizer. Note that while it is recommended to submit a paper, it is not mandatory to participate.

Q: What and how should I submit?

For the final submission, we request the participants to submit the binaries of their models (so that we may evaluate it on our 20 image verification dataset). For more details on the procedure of submission, please refer to section C of Annex.

For More detailed info see Annex

Metrics	Ground Truth	Model
Sec D	Sec A	Head motion based saliency (Model Type 1)
Sec D	Sec B	Head+Eye motion based saliency (Model Type 2)
Sec E	Sec C	Scanpath models (Model Type 3)

ANNEX

Section A. Presentation of ground truth data and parsing tools

The groundtruth for all the three types of models are provided on the FTP link: One can find find 40 saliency maps to evaluate model 2 and 40 scanpaths to evaluate model 3, as well as 20 saliency maps to evaluate model 1.

The tools to read each of the three types of data(saliency maps and scanpath text files) can also found in the FTP.

- The ground-truth data for model type 1, is organised into a binary file " SHE_<ImageNumber>.bin" containing float values (4 bytes), depicting the Heat Map in the equirectangular space. The Heat Map data is organised row-wise across the image. The minimum value of saliency is 0 and the sum of all image pixel saliencies equals to one.
- The ground-truth data for model type 1, is organised into a binary file " SH_<ImageNumber>.bin" containing float values (4 bytes), depicting the Heat Map in the equirectangular space. The Heat Map data is organised row-wise across the image. The minimum value of saliency is 0 and the sum of all image pixel saliencies equals to one.
- The ground-truth data for model type 3, is organised in a text file named "*SP<ImageNumber>.txt*". Each line contains a quadruple vector that indicates the Fixation Number, Fixation-Time, X-Position (Equirectangular) and Y-Position (Equirectangular) respectively. The fixation number increments serially for a particular observer and resets to 1 when we reach the next observer, after all of the fixations of the given observer are completed. The fixation time is indicated in seconds and X and Y positions are indicated in pixels (of the respective Equi-Rectangular image).
- Parsing tools for reading all of this data is also included in the FTP link mentioned

Section B. Performance evaluation tools

For evaluating the performance of each model against the

groundtruth data collected from the experiment, special methods and tools are provided. More specifically,

- The evaluation tool for model types 1 and 2, requires two inputs: the predicted saliency maps in equirectangular format generated from the participants model, and the respective ground-truth saliency maps that we have provided to the participants. Because, an equi-ectangular image is stretched at the poles, we cannot compare the saliency of two equi-rectangular images directly. For this purpose, we first perform a uniform sampling over the viewing sphere and then find the corresponding points on the equirectangular image to ultimately sniff out the saliency value at these points. After choosing appropriate sampling points where the saliency values will be actually compared, we then obtain the respective values of saliency from these equirectangular images using the metrics described in Section D;
- The evaluation tool for model 3 (scan-path models), requires 40(M) sets of N scanpaths where M is the total number of images and N is the total number of observers who have viewed each image. The scan-paths will then be compared in accordance to the metrics described in section E.
- All tools for performing this evaluation is also included in the mentioned FTP.

Section C. Submission formats

For model types 1 and 2, the binaries must be capable of taking in a 360 degree image in equirectangular RGB format as an input, and produce a binary file with the same equirectangular dimension as the

source image. This output binary file must contain the saliency values of the 360 panorama in equirectangular format, organised rowwise. The sum must be equal to 1 and the minimum value must be equal to 0.

Matlab: matOut=YourFunc(imgIn)

imgIn is the input equirectangular image organised in an RGB, the size(imgIn) being [Height,Width,3].

matOut is the output “double” matrix having the saliency values. Its size is [Height,Width]

C++: void YourFunc(imgIn,imgOut)

imgIn is the input equirectangular image as an RGB, the size(imgIn) being 3xWidthxHeight and it is organised first by color (order RGB), then by row and then by column

imgOut is the output equirectangular image a float array, the size(imgOut) being WidthxHeight and it is organised first by row and then by column

For model type 3, the submitted binary should be capable of taking an an input equirectangular image in RGB format and output the scan-paths.

Matlab: scanPaths=YourFunc(imgIn)

imgIn is the input equirectangular image organised in an RGB, the size(imgIn) being [Height,Width,3].

scanPaths is the output [4,N] matrix containing the (x,y,Tst,Tend) positions of the fixations

C++: void YourFunc(imgIn,imgOut)

imgIn is the input equirectangular image as an RGB, the size(imgIn) being 3xWidthxHeight and it is organised first by color (order RGB), then by row and then by column

imgOut is the output equirectangular image a float array, the size(imgOut) being WidthxHeight and it is organised first by row and then by column

Section D. Metrics to evaluate saliency maps

We reuse several earlier metrics that are used for comparing 2D saliency maps. More specifically, we use the:

- Area under the curve (AUC) (Judd et al)
- Normalised Scanpath saliency (NSS)
- KL Divergence
- Earth Mover Distance (EMD)
- Correlation Coefficient (CC)

Matlab codes for computing these metrics will also be released.

Section E. Metrics to evaluate scan-path models

For comparing scanpaths we use the scan-path similarity metric used in earlier articles by Le Meur et al [Refer to 1]. Here , each of the **m** individual saccades in the model generated scan-path is compared to that of the **n** saccades in the ground truth scanpath. Three separate aspects of similarity are considered: spatial proximity of the saccade starting points i.e 2D euclidean separation between their starting points on the sphere, Difference in direction and magnitude as indicated by the vector difference of the saccades and third, the temporal proximity of the two saccade midpoints in which the saccade initiation latency of 219ms is modelled as a rectangular function. Each of the three measures are normalized by the

maximum possible value, i.e. the screen diagonal for the first two measures and are then averaged together to produce an overall similarity score ranging from 0 to 1. We then compare the 48 saccades generated by the model with the 48 observers presented in the experiment. An average is then made after finding the best-matches among all of them.

[1]http://www.lutin-userlab.fr/site/accueil/documents/10.3758_s13428-012-0226-9.pdf