expoze

Our platform explained

March 2020

Introduction: **Predictive eye-tracking**

Using state-of-the-art advances in computer vision and artificial intelligence, we simulate eye-tracking without the need for a labor-intensive study with participants. Our software predicts eye fixations on natural images by highlighting areas in the image that have high *salience*. Salience is defined as those areas in an image or video that receive relatively more visual attention than other areas, as can be seen in Figure 1. It can be modeled by accumulating fixated eye gaze locations from many people.

Based on these insights from cognitive science, we make use of artificial neural networks to find patterns in visual data that naturally draw the eyes towards them. Different to regular eye-tracking studies that are performed with a limited number of people, our neural nets are trained on the visual responses of thousands of people. Using this many different people to train the network has the advantage that the predictions are less prone a personal bias that would be introduced when using a limited number of people, such as is often the case in real eye-tracking studies.

Furthermore, because we make use of state-of-the-art technology, we can process images in a matter of seconds, and videos in a matter of minutes. Results are ready while you wait. Our platform is available in an online environment, where you simply upload your image or video and receive its saliency predictions in a matter of minutes.

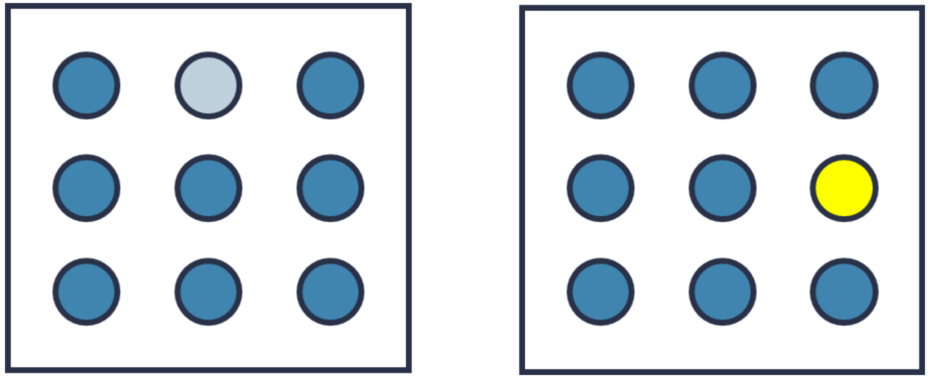


Figure 1: Visual saliency refers to the idea that certain parts of a scene are pre-attentively distinctive and create some form of immediate significant visual arousal. In the images above the yellow dot on the right has higher visual salience than the divergent blue dot on the left.

expoze

The technology behind **Expoze.io**

Our technology is based on solid computer vision research, inspired by cognitive science [1, 2, 3]. Example results (given for instance in Figure 6), show that our predictions are very comparable to the results from a real eye-tracking study.

The first step in our process is to gather saliency data from a large set of visual stimuli, such as advertisement images or video frames. This data is then exposed to a diverse group of real people in an online environment. By simulating eye-tracking, there is no need for this group to participate in a study with real eye-trackers.

After gathering enough training data, the results are fed to an artificial neural network (Figure 2), using a specialized training scheme. After optimizing the neural net, it can simulate the saliency results gathered in the first step. Beyond that, the neural net can make accurate saliency predictions of visual material it has never seen before.

As our network can process data at a very rapid rate, we are not only able to make saliency predictions of large numbers of images, but we are even able to perform these predictions for entire videos.

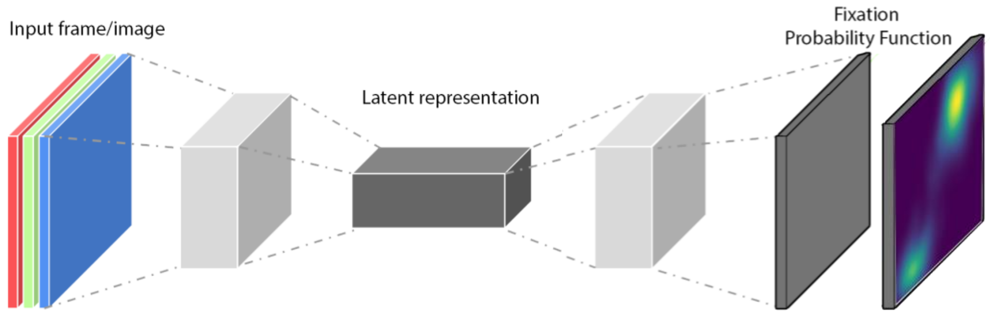


Figure 2: An artificial neural network models the latent visual representation of high salience areas of the input frame/image. The output of the neural network is a fixation probability function, that can be visualized as a heatmap.

1: *MIT Saliency Benchmark*, Bylinskii et al. http://saliency.mit.edu/
2: *A Benchmark of Computational Models of Saliency to Predict Human Fixations*, Judd et al. (2012)
3: *SALICON: Saliency in Context*, Ming et al. (2015)

Validation of our results

We have validated the results of our artificial neural networks using state of the art measurement algorithms, like the [MIT Saliency Benchmark](). We use two metrics to show that are predictions are comparable to real eye tracking studies.

Similarity scoring:
By viewing the prediction results as a histogram of probability distributions, we can compare our results to a histogram of probability distributions based on real eye fixations. Using a histogram intersection measurement, we can compare the performance of the predictions to real eye fixations. The more the two distributions overlap, the more the prediction is in line with real eye tracking results, as can be seen in Figure 3.

Area Under the Curve (AUC) of Precision and Recall:
By considering salient pixels above a certain threshold as fixated, we can measure the precision and recall of a salience map. We can compare fixated pixels to the fixation results from real eye-tracking as follows:

- True Positive (TP) fixations: pixels that **should** be fixated and **are**
- True Negative (TN) fixations: pixels that **should not** be fixated and **aren't**
- False Positive (FP) fixations: pixels that **should not** be fixated, but **are**
- False Negative (FN) fixations: pixels that **should** be fixated, but **aren't**

Precision is then defined as: $\frac{TP}{TP+FP}$, giving the rate of accurately marked fixations. Recall is defined as: $\frac{TP}{FN+TP}$, giving the rate of relevantly marked fixations. In Figure 4 a visual representation of precision and recall is given.

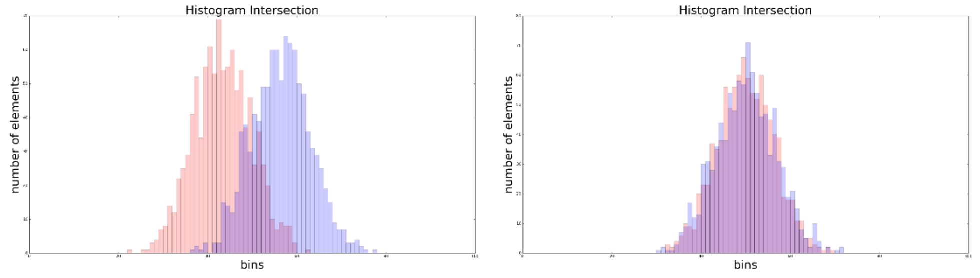
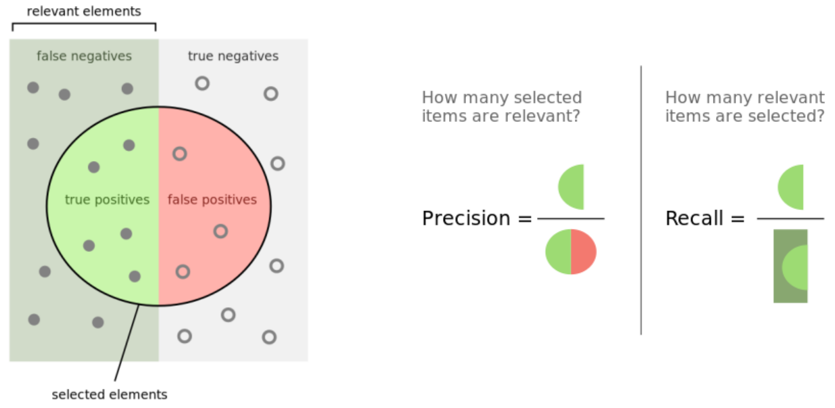


Figure 3 (above): Probability distributions with different overlaps.
Figure 4 (below): Precision and recall measurement.



Validation of our results

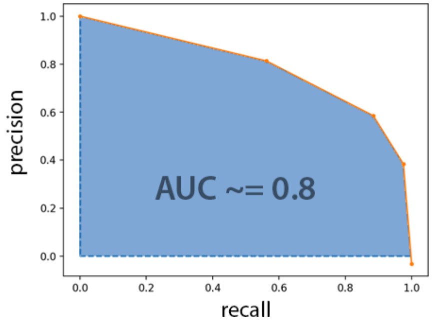
By varying the fixation threshold on the saliency map, a graph can be created that shows the precision and recall (Figure 5). By measuring the area under the curve of the precision/recall graph, a single measure can be given for the performance of the predictions.

A low fixation threshold will retrieve a lot of fixations (high recall), at the expense of the accuracy of these predictions (low accuracy). Conversely, a high fixation threshold will only retrieve correct predictions (high accuracy), at the expense of retrieving few fixations (low recall).

Therefore, an AUC value closer to 1.0 indicates that number of fixations that were accurately predicted, is high.

Performance comparisons for the two different metrics are given in the table to the right. We show that our neural network predictions (Expoze.IO) achieve competitive results, when compared to the average results of near infinite human eye-tracking predictions (Baseline, ~∞ human), and outperforms a single human (Baseline, 1 human). This is understandable given individual variation in saliency of humans (bias), while our models predict the saliency of the average human. Finally, we show that we outperform control groups and competitive salience models.

A qualitative saliency visualization (Figure 6) shows the appeal of our models. The best heatmaps (images 3 and 4) show the acuity of our predictions, compared to real eye-tracking results (image 2). The heatmaps accurately and clearly show areas with high salience.



	AUC	SIM
Baseline (~∞ human)	0.92	1.00
Baseline (1 human)	0.80	0.38
Expoze.IO	**0.87**	**0.60**
Baseline (center)	0.78	0.45
SaliencyToolbox*	0.75	0.44
Baseline (chance)	0.50	0.33

Figure 5 (above): A Precision/Recall curve with an AUC of about 0.8.
Figure 6 (below): Qualitative comparison of results. From left to right: original, ground truth, best prediction, second best prediction and worst prediction.

expoze