# A Saliency-based Technique for Advertisement Layout Optimisation to predict Customers' Behaviour

Alessandro Bruno[1][0000−0003−0707−6131], Stéphane Lancettte[1], Jinglu Zhang[1], Morgan Moore[1], Ville P Ward[2], and Jian Chang[1][0000−0003−4118−147X]

[1] National Centre for Computer Animation
Bournemouth University
Poole, United Kingdom BH12 5BB
Email: abruno@bournemouth.ac.uk
abruno,zhangj,slancette,jchang@bournemouth.ac.uk
https://www.bournemouth.ac.uk/
[2] Shoppar Ltd, Plexal, 14 East Bay Lane
Stratford London. E20 3BS. UK
peter.ward@shopparapp.com
https://shopparapp.co.uk/

**Abstract.** Customer retail environments represent an exciting and challenging context to develop and put in place cutting-edge computer vision techniques for more engaging customer experiences. Visual attention is one of the aspects that play such a critical role in the analysis of customers behaviour on advertising campaigns continuously displayed in shops and retail environments. In this paper, we approach the optimisation of advertisement layout content, aiming to grab the audience's visual attention more effectively. We propose a fully automatic method for the delivery of the most effective layout content configuration using saliency maps out of each possible set of images with a given grid layout. Visual Saliency deals with the identification of the most critical regions out of pictures from a perceptual viewpoint. We want to assess the feasibility of saliency maps as a tool for the optimisation of advertisements considering all possible permutations of images which compose the advertising campaign itself. We start by analysing advertising campaigns consisting of a given spatial layout and a certain number of images. We run a deep learning-based saliency model over all permutations. Noticeable differences among global and local saliency maps occur over different layout content out of the same images. The latter aspect suggests that each image gives its contribution to the global visual saliency because of its content and location within the given layout. On top of this consideration, we employ some advertising images to set up a graphical campaign with a given design. We extract relative variance values out the local saliency maps of all permutations. We hypothesise that the inverse of relative variance can be used as an Effectiveness Score (ES) to catch those layout content permutations showing the more balanced spatial distribution of salient pixel. A group of 20 participants have run some

eye-tracking sessions over the same advertising layouts to validate the proposed method.

**Keywords:** Visual Saliency · Retail Environment · Layout Optimisation · Computer Vision · Deep Learning.

## 1 Introduction

Over the last few years, it is observed an increasing demand for software tools for customer retail environments aiming for better understandings of customers' behaviours. As a matter of fact, computer vision-based algorithms have been widely adopted throughout heterogeneous application domains to automatise some context-aware tasks. Some retail companies invest in AI (Artificial Intelligence) and Computer Vision tools to strengthen their rank in a hugely competitive market. The analysis of visual attention processes during the customer experience represents quite a remarkable challenge to set up tasks which could turn out as stepping stones for retail companies. Digital screens are widely adopted in shops and retail environments to grab customers' attention over particular products and services. In this paper, we focus on the study of visual attention and, more particularly, visual saliency as a tool for the assessment and the optimisation of engaging advertisements for customer retail environments. From a computer vision perspective, much of progress has been made on the attempt of imitating and predicting the behaviour of HVS (Human Visual System) over the first seconds of observation of images [20] [6]. Many techniques in the scientific literature approach the analysis of the visual attention focusing on the prediction of eye-movements over the first seconds of observation of a given scene. Visual Saliency is meant to decode spatial prediction of eye-movement returning the so-called saliency maps, that is, grey-scale maps encoding the probability each pixel might grab viewers' attention in the continuous range [0,1]. Most of the visual saliency approaches based on deep neural networks trained over publicly available datasets [7] [8] [16] [4] [3] allow to achieve high accuracy rates in the prediction of eye-movements in different scenarios. Several saliency-based techniques have been proposed in a wide range of contexts such as computer graphics, remote sensing, biomedical imaging [1] [10] [23] . Some applications such as image retargeting, image cropping and image quality assessment [2] [5] employ visual saliency as a perceptually inspired means to accomplish tasks. In our work, we focus our efforts on how saliency maps can be used as a screening tool for the optimisation of the effectiveness of advertisements to predict customers' behaviour. The increasing interest towards computer vision techniques to make customer retail environments more engaging to potential customers reveals new application domains which need to consider visual attention processes as a leverage to improve the effectiveness of their advertisements. The main idea behind the proposed method is based on the relation between peaks of salient blobs in saliency maps and real eye-movements. If a saliency method reaches high benchmarks over different real fixation points datasets, it might occur the

other way around either. We suppose that higher peaks in saliency maps should mostly correspond to the most beaten regions during observation. Furthermore, the sparser the peaks, the more likely viewers will look at areas all over the image. On the other side, the closer the saliency peaks are in a layout, the more likely some locally close regions in the image will grab viewers' attention. In our work, we conduct some experiments on a case study to validate the intuition above behind the proposed method. We rank all spatial permutation of a given advertisement layout with relative variance values of local saliency blobs. We validate the method using real feedback with a web-cam based eye-tracking, collecting eye-movements and fixation points in the first 10 seconds of observation of images. Our contributions are respectively an automatic saliency-based method to set up the most engaging advertisement content for a given layout and images, a validation session conducted with advertising images and a collection of real eye-movement to assess the robustness of the method. The remainder of the paper is organised as follows: Section 2 summarises the state-of-the-art techniques in this topic, section 3 provides a detailed description of the proposed method, section 4 shows the experimental results and section 5 ends the manuscript with conclusions and future works.

## 2 Related Techniques

In this section, we provide the paper with a description of some articles in the scientific literature focused on the improvement of the customer experiences in retail environments. Researchers from different disciplines highlight interesting aspects on shopper behaviours related to visual attention, data collection, 3D mapping and reconstruction, trajectory detection and others which may literally represent a stepping stone for a new concept of customer retail environments.
The authors in [13] conducted some studies with wearable eye-trackers to detect the most important factors which play prominent roles in the final decision of buying a product. Visual attention is a key aspect in the act of looking longer or repeatedly at a product which will be more likely bought. Huddleston et al. [14] reviewed the progress of eye-tracking technology as a research tool in retail and retail marketing.
Khan [17] studied the impact of visual designs on customer perceptions of online assortments highlighting designs with simpler compositions to be the ones more liked by viewers. Paolanti et al. [21] tackled the topic of semantic store mapping using artificial intelligence models and a retail robot. They set up a system to build a 3D map of both the store and product locations. La Porta et al. [18] proposed a method based on a deep learning mobile application able to identify facial expressions and emotions of subjects from an image. On top of it, the lights of a Christmas tree run different special effects because of the emotions of the subjects around the tree.
Gabellini et al. [12] provided the scientific community with a large scale trajectory of shopper movements in stores employing a real-time locating system with Ultra-Wideband (UWB). Vaira et al. [25] put in place a system to avoid


**Fig. 1.** Some layout content permutations and their saliency maps are shown above. Differences among salient blobs of corresponding images across permutations are noticeable.

the so-called OOS (out of stock) problem in stores. The solution is based on two cameras, one with a depth sensor and a very high-resolution webcam recognition tasks.
The authors in [24] paired the utilisation of blue-tooth beacons' signal and sensor fusion approach with RGB-D camera to provide an accurate customer position detection and track shoppers movements. Liciotti et al. [19] proposed a software infrastructure consisting of computer vision algorithms and RGB-D cameras providing information related to the analysis of user-shelf interactions. Fuchs et al. [11] focused their efforts on the employment of edge device equipped with AI tools to help retailers improve the quality of their customers' experience. The next section is a detailed description of all steps of the proposed method.

## 3 Proposed Method

As briefly mentioned in section 1, we propose a saliency-based method for optimising the effectiveness of advertisements in shops and customer retails. We detect salient pixels in images using a deep learning-based solution [26] trained over an object-oriented image and video dataset called DAVIS [22]. As a case of study, we consider a scenario where layout and images are provided to deliver an advertisement.
For the sake of clarity, we refer to the layout as the current combination of images within the given design configuration of images. In figure 1 an example with five permutations of a given $2 \cdot 2$ layout and four input images are shown. Each image represents a local region of the overall advertisement layout. The objective is to convey the most engaging advertisement possible with the given inputs (images and layout). We consider a generic $M$ x $N$ grid layout consisting of $M$ x $N$ images. We notice different spatial permutations of the same layout (see figure 1) showing different saliency maps. All five permutations in figure 1 show blobs whose saliency turns up differently because of the image location within the given layout. Saliency maps encode image pixels in the continuous range [0,1], then the most salient regions in images can be read as regions which may grab viewers' attention most likely. The aforementioned noticeable differences among saliency maps prompt us to further investigate visual saliency as leverage for predicting customers' behaviour in retail environments. A flow-chart in figure 2 describes the main steps of the proposed method. For a given $M$ x $N$ grid-based layout and $M$ x $N$ images composing the advertisement, all $P$=($M$ x $N$)! layout permutations are automatically generated. A saliency map out of each permutation is then extracted, summing up to $P$ saliency maps. The most salient pixels are filtered in using a simple spatial threshold to detect the highest peaks of visual saliency in images.

We want to retrieve effectiveness scores out of each permutation saliency map, aiming for catching the one with the most well-balanced spatial distribution of salient pixels over all the images. If an image of the layout is much more salient than another, viewers will likely dwell on it for longer than the other images. The idea behind our algorithm is to employ scores which allow us to detect the

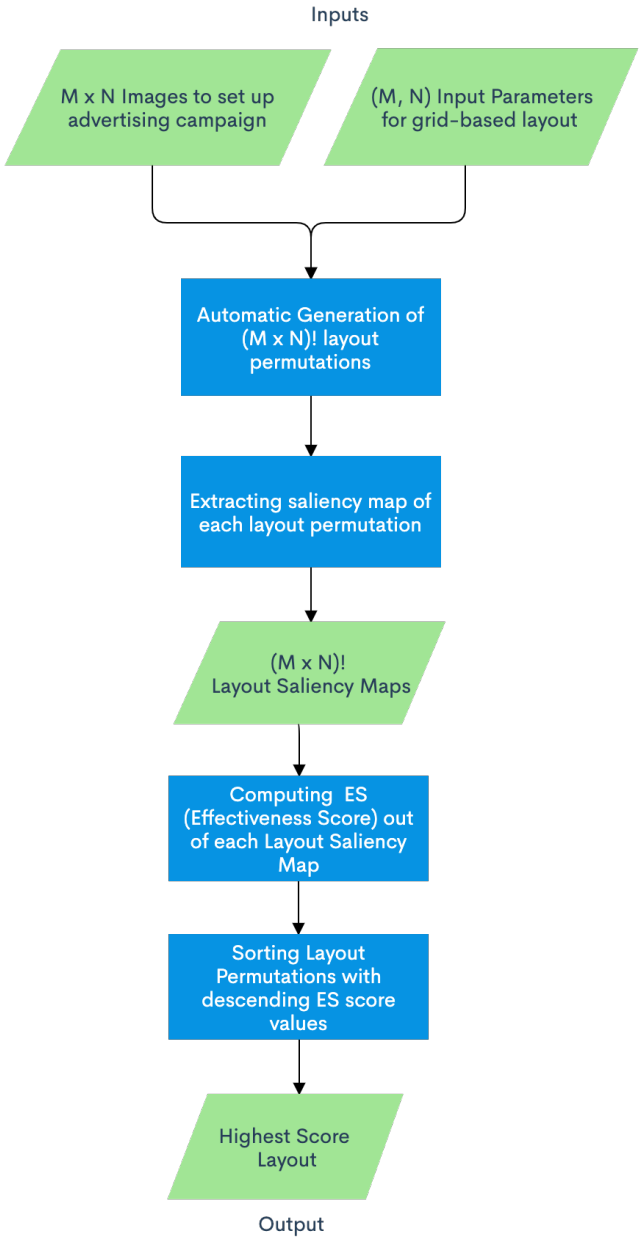


**Fig. 2.** Flow-chart of the proposed method.

layout content permutation that shows lower saliency variance among images. Due to different salient blobs over images in the same layout, we focus our efforts on the analysis of the variance of what we name 'local-saliency'. For the given layout made up of $M \cdot N$ images, we study the 'behaviour' of the overall layout saliency analysing the varying number of salient pixels of each of the image $M \cdot N$ images. In greater detail, we employ the inverse of the relative variance of local saliency maps as $ES$ (Effectiveness Score). In equation 1 $ES$ is the ratio between the absolute mean and variance of $NMSP_k$ with $_k = 1, ..., (M \cdot N)$. $NMSP_k$ stands for Number of Most Salient Pixels of each image in the k$^{\text{th}}$ layout content permutation.

$$ES_{(i)} = \frac{\left|\mu(NMSP_k(Layout_{(i)})\right|}{\sigma(NMSP_k(Layout_{(i)}))^2} \qquad k = [1, ..., (M \cdot N)] \quad i = \{1, ..., (M \cdot N)!\} \tag{1}$$

For a given layout with $M \cdot N$ images, $NMSP_h$ is the number of the most salient pixels in the local saliency map $LSM_{(h)}$ of the h$^{\text{th}}$ image (see equation 2).

$$NMSP_{(h)} = \sum_{i,j \,\epsilon\, Im} LSM_{(h)}(i,j) \geq th \tag{2}$$

Each $Layout$ content permutation is the union of $M \cdot N$ images $Im'_i$ as in the equation 3

$$Layout = \bigcup_{i'=1}^{M \cdot N} Im_{i'} \tag{3}$$

Each layout content permutation is then sorted along with its $ES$ . The layout showing the highest score is the output of our proposed method. To confirm the robustness of our technique, we run through a validation session showing both best and worst score layout permutations to viewers and capture eye-movements over the first 10 seconds of observation of each permutation. To record eye-movements we employ Web Tool for eye-tracking called GazeRecorder [9].

## 4 Experimental Results

In this section, we show experimental sessions on five different advertising campaigns with a $2 \cdot 2$ layout. All images for our experiments are taken both from a publicly available dataset [15] at the link http://people.cs.pitt.edu/ kovashka/ads/ and over the Internet. For each layout, 24 permutations are computed. Their corresponding saliency maps are then extracted, and the 35% most salient pixels are filtered in with equation 2. In figure 3, we show all ES values of each layout content permutation of the four images in the first advertising campaign. ES values of each layout content permutation are quite different. The layout content permutation with the highest ES is highlighted with a red dot, while the

black dot in the graph matches the lowest score layout content permutation. Highest and lowest score layout content permutations are the first and second layouts shown in figure 1. Local saliency maps are quite different between the two configurations. We run experimental sessions on 5 advertising campaigns to

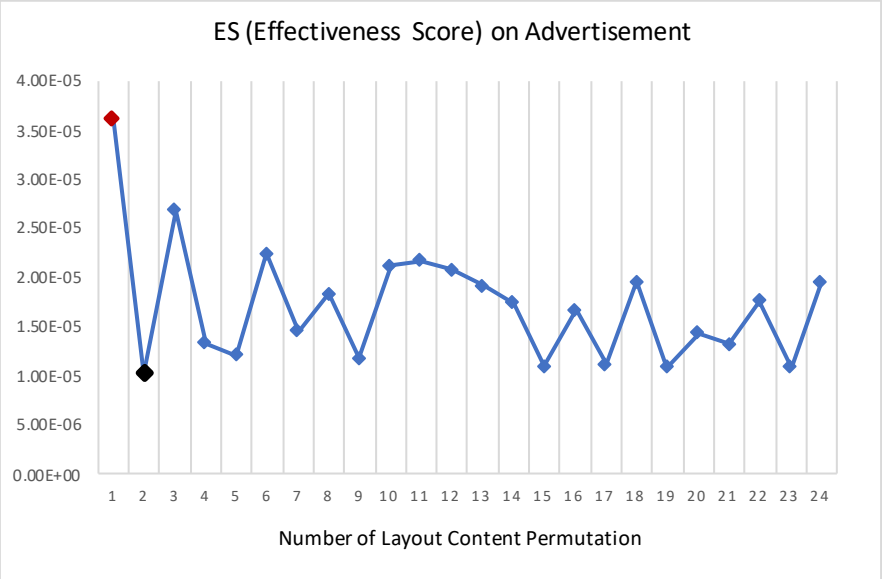


**Fig. 3.** The graph above shows all ES values for each layout content permutations in the advertising campaign 1. The red dot show the highest score while the black one shows the lowest score

pick up the permutations with the highest and lowest ES. In figure 5, ES values are reported across the 5 advertising campaigns. In our case study, we consider 2 by 2 layouts which sum up to 24 possible permutations for each advertising campaign. The total number of layout content permutations is 120. As it is noticeable in the histogram chart in figure 5, gaps between the highest and lowest score over each campaign can be quite different. A plausible explanation is that the content of images which are part of the campaign consists of low, middle and high-level image features having a different impact on the saliency map. The experimental sessions shown so far concern the automatic detection of what we consider to be the most engaging layout content permutation for a given grid-based layout. The remainder of this section is focused on the validation sessions we conduct to assess the robustness of the algorithm behind the proposed method. In greater detail, we pick up all highest and lowest score layout permutations out of 5 advertising campaigns to run through webcam-based eye-tracking sessions. We used an off-the-shelf solution called GazeRecorder based on a web tool integrating a webcam calibration step before eye-tracking sessions.

We employed GazeRecorder to record participants' eye-movements over the first 10 seconds of observation of layout permutations with highest and lowest ES. Twenty subjects took part in the validation sessions giving us the chance to

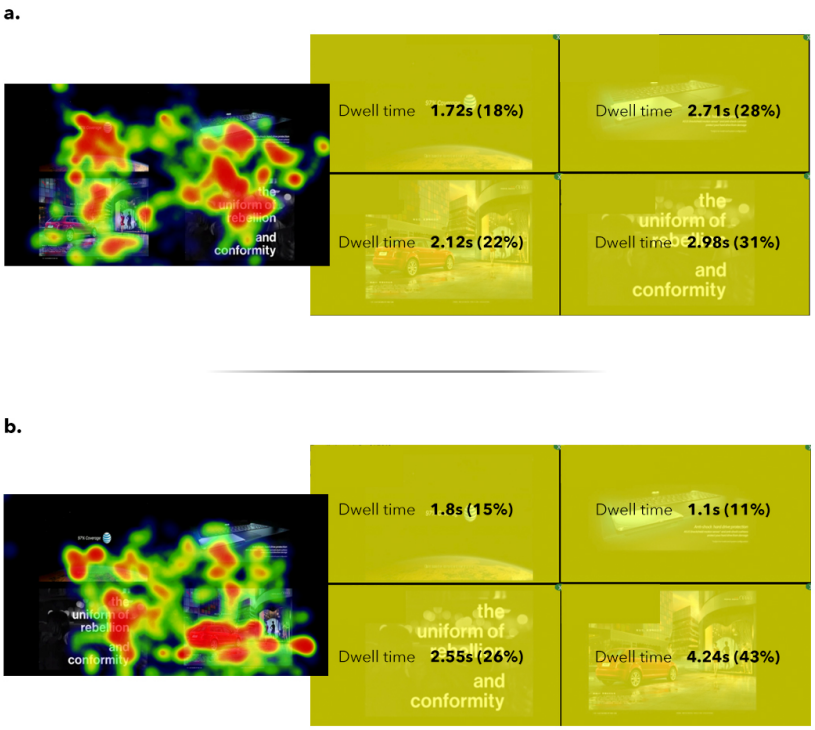


**Fig. 4.** Heat maps of real eye-movements and dwell times over each image in layouts are shown above. Best (a) and worst (b) layout content permutations show different amount of time spent by viewers over matching images between the two layouts.

extract statistically meaningful results out of the experiments. The participants in the eye-tracking sessions were equally distributed by gender. Their age was in the range of 25-40 years old. In our case study, we suppose that a digital screen in a retail environment displays an advertising campaign showing some objects or services the retail company or shop offer to customers. One of the objectives is to grab customers' attention to those objects in the advertisement itself. For this purpose, each of 20 participants is shown the layout content permutations with highest and lowest ES value as per the graph in figure 5. Each image is displayed for a time range of 10 seconds during which on-screen eye-movements of the viewer are recorded. GazeRecorder allows us to set up the validation sessions fine-tuning the time range and the order of layout contents to be shown. Furthermore, coordinates of eye-movements of all sessions are gathered and then

displayed as heatmaps (see figure 4 overlaid with the input image. Other useful pieces of information are retrieved by manually drawing rectangular regions of interest on the layout content permutation. In figure 4, layout content permutations with highest (a) and lowest (b) ES are shown (left-hand side of the figure), the corresponding dwelling times over each image in the layouts are printed out in the right-hand side. Dwell times are meant to provide information on how long viewers spend their time focusing on one out of four pictures in the layout.

As noticeable in figure 4, the layout scoring the highest ES (a) show values of dwell times over each image which are more equally distributed than the ones shown in the layout scoring lowest ES (b). Red-car image observations take 2.12 seconds in the highest ES configuration while more than 4 seconds in the lowest ES configuration. We consider advertisements to be more engaging whether customer's attention gets caught over all products and services displayed with a sort of equalised distribution of dwell times. Due to the consideration above, the lower dwell time variance of the four images in the layout, the more the engagement is spatially distributed. After describing experimental results related to validation session on the first graphical campaign, we show the variance of dwell times of each image for both highest and lowest ES layout permutation over all the 5 advertising campaigns (see figure 6. Highest score layout permutations have dwelling time variance values lower than the lowest score layout permutation. In the graph in figure 6, the lower dwell time variance values, the better layout configuration. The preliminary experimental sessions indicate the highest ES layout content permutations to grab viewers' attention over all the images of a given advertisement with more balanced dwell time distribution. The experiments on the automatic optimisation of the given advertisement layouts and images have been carried out on a 13-inch Mac-book Pro with 16 GB of RAM, 2.4 GHz Quad-Core Intel Core i5, Intel Iris Plus Graphics 655 1536 MB. The average running time of the method on the 2-by-2 grid layout is 40 seconds. The entire project has been developed with python version 3.8.0.

## 5 Conclusions and Future Works

In our work, we show how saliency maps as leverage for automatically optimising advertisement layout contents. We aim to provide retail environments with a lightweight, cost-effective and reliable software solutions to predict customers' eye-movements and dwell times on advertised products in a given graphical layout. Entirely meaningful changes in the level of engagement of an advertisement layout might occur because of changes in positions of its visual elements. We set up a new Effectiveness Score of a layout content as the ratio of absolute mean to the variance of the number of the most salient pixel of each image composing the advertisement itself. The score returns a measure of each image contribution to the overall spatial distribution of saliency in the layout. The final version of an advertisement is the combination of some ingredients such as layout and images. Our technique allows providing the most engaging layout content out of all possible permutations. Our method is validated by some preliminary experiments

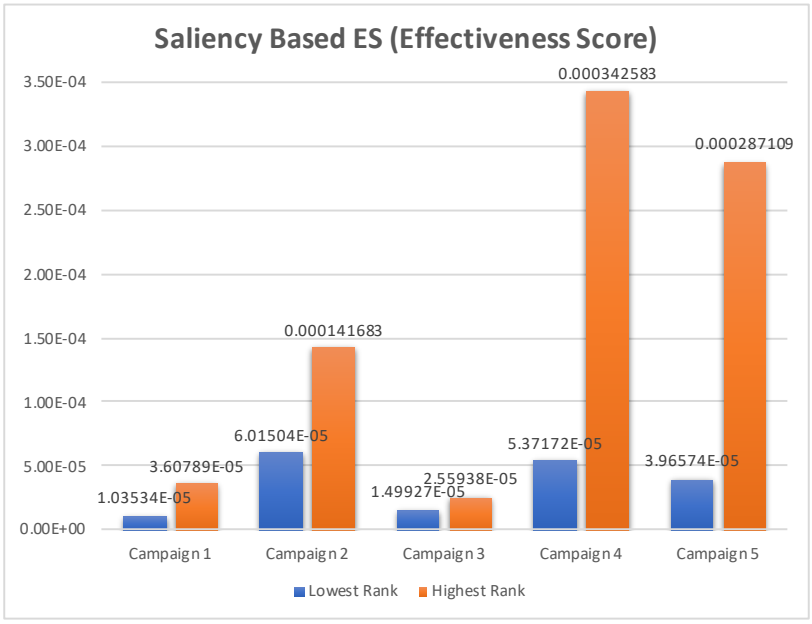


**Fig. 5.** The graph above shows highest and lowest ES values for each advertisement consisting of 4 images. Each advertisement layout is based on a 2 by 2 grid layout, summing up to a total of 24 permutations. Experiments are run over 5 advertisements, meaning 120 possible layout content permutations.

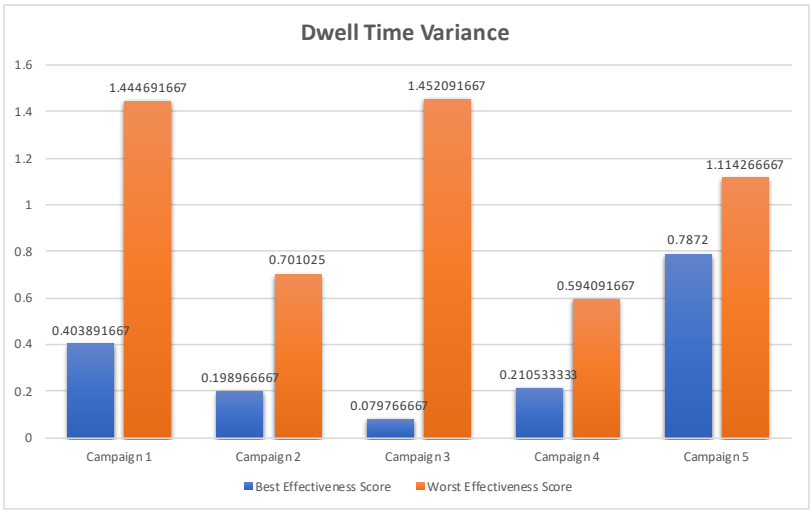


**Fig. 6.** This graph shows variance values of dwell time over each image in layouts with highest and lowest ES for a given advertising campaign. The first two histogram bars in the left-hand side of the graph are the dwell time variances of first campaign (see figure 4 a and b). We run experiments over 5 different campaigns.

with real feedback of 20 subjects who underwent webcam-based eye-tracking sessions. Preliminary results show that the same layout permutations, which are best ranked with ES, are the same ones where visual attention is more equally distributed on all images in the advertisement itself. Our work suggests that visual saliency methods trained with artificial intelligence models can be used as a reliable tool to optimise and predict customers' behaviour when they cast their eyes over advertising layouts. Further investigations are necessary for the effectiveness of the method with finer grid layouts. Adding transformations such as scaling and rotation in the extraction of all possible permutations would probably make the functionalities appealing to a broader audience and application domains. We also aim for extending our work to other scenarios taking into account graphic elements with different priority scales.

## Acknowledgment

This research was supported by Innovate UK. Smart Grants (39012) - Shoppar: Dynamically Optimised Digital Content.

## References

1. Abouelaziz, I., Chetouani, A., El Hassouni, M., Latecki, L.J., Cherifi, H.: 3d visual saliency and convolutional neural network for blind mesh quality assessment. Neural Computing and Applications pp. 1–15 (2019)
2. Ardizzone, E., Bruno, A.: Image quality assessment by saliency maps. In: VISAPP (1). pp. 479–483 (2012)
3. Borji, A., Itti, L.: Cat2000: A large scale fixation dataset for boosting saliency research. CVPR 2015 workshop on "Future of Datasets" (2015), arXiv preprint arXiv:1505.03581
4. Borji, A., Sihite, D.N., Itti, L.: Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. Image Processing, IEEE Transactions on **22**(1), 55–69 (2013)
5. Bruno, A., Gugliuzza, F., Ardizzone, E., Giunta, C.C., Pirrone, R.: Image content enhancement through salient regions segmentation for people with color vision deficiencies. i-Perception **10**(3), 2041669519841073 (2019)
6. Bruno, A., Gugliuzza, F., Pirrone, R., Ardizzone, E.: A multi-scale colour and keypoint density-based approach for visual saliency detection. IEEE Access **8**, 121330–121343 (2020)
7. Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., Torralba, A.: Mit saliency benchmark. http://saliency.mit.edu/
8. Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., Durand, F.: What do different evaluation metrics tell us about saliency models? arXiv preprint arXiv:1604.03605 (2016)
9. Deja, S.: Gazerecorder. https://api.gazerecorder.com/
10. Diao, W., Sun, X., Zheng, X., Dou, F., Wang, H., Fu, K.: Efficient saliency-based object detection in remote sensing images using deep belief networks. IEEE Geoscience and Remote Sensing Letters **13**(2), 137–141 (2016)

11. Fuchs, K., Grundmann, T., Fleisch, E.: Towards identification of packaged products via computer vision: Convolutional neural networks for object detection and image classification in retail environments. In: Proceedings of the 9th International Conference on the Internet of Things. pp. 1–8 (2019)
12. Gabellini, P., D'Aloisio, M., Fabiani, M., Placidi, V.: A large scale trajectory dataset for shopper behaviour understanding. In: International Conference on Image Analysis and Processing. pp. 285–295. Springer (2019)
13. Gidlöf, K., Anikin, A., Lingonblad, M., Wallin, A.: Looking is buying. how visual attention and choice are affected by consumer preferences and properties of the supermarket shelf. Appetite **116**, 29–38 (2017)
14. Huddleston, P.T., Behe, B.K., Driesener, C., Minahan, S.: Inside-outside: Using eye-tracking to investigate search-choice processes in the retail environment. Journal of Retailing and Consumer Services **43**, 85–93 (2018)
15. Hussain, Z., Zhang, M., Zhang, X., Ye, K., Thomas, C., Agha, Z., Ong, N., Kovashka, A.: Automatic understanding of image and video advertisements. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1705–1715 (2017)
16. Judd, T., Durand, F., Torralba, A.: A benchmark of computational models of saliency to predict human fixations. In: MIT Technical Report (2012)
17. Kahn, B.E.: Using visual design to improve customer perceptions of online assortments. Journal of Retailing **93**(1), 29–42 (2017)
18. La Porta, S., Marconi, F., Lazzini, I.: Collecting retail data using a deep learning identification experience. In: International Conference on Image Analysis and Processing. pp. 275–284. Springer (2019)
19. Liciotti, D., Frontoni, E., Mancini, A., Zingaretti, P.: Pervasive system for consumer behaviour analysis in retail environments. In: Video Analytics. Face and Facial Expression Recognition and Audience Measurement, pp. 12–23. Springer (2016)
20. Nguyen, T.V., Zhao, Q., Yan, S.: Attentive systems: A survey. International Journal of Computer Vision **126**(1), 86–110 (2018)
21. Paolanti, M., Pierdicca, R., Martini, M., Di Stefano, F., Morbidoni, C., Mancini, A., Malinverni, E.S., Frontoni, E., Zingaretti, P.: Semantic 3d object maps for everyday robotic retail inspection. In: International Conference on Image Analysis and Processing. pp. 263–274. Springer (2019)
22. Perazzi, F., Pont-Tuset, J., McWilliams, B., Van Gool, L., Gross, M., Sorkine-Hornung, A.: A benchmark dataset and evaluation methodology for video object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 724–732 (2016)
23. Sran, P.K., Gupta, S., Singh, S.: Segmentation based image compression of brain magnetic resonance images using visual saliency. Biomedical Signal Processing and Control **62**, 102089 (2020)
24. Sturari, M., Liciotti, D., Pierdicca, R., Frontoni, E., Mancini, A., Contigiani, M., Zingaretti, P.: Robust and affordable retail customer profiling by vision and radio beacon sensor fusion. Pattern Recognition Letters **81**, 30–40 (2016)
25. Vaira, R., Pietrini, R., Pierdicca, R., Zingaretti, P., Mancini, A., Frontoni, E.: An iot edge-fog-cloud architecture for vision based pallet integrity. In: International Conference on Image Analysis and Processing. pp. 296–306. Springer (2019)
26. Wang, W., Shen, J., Shao, L.: Video salient object detection via fully convolutional networks. IEEE Transactions on Image Processing **27**(1), 38–49 (2017)