ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

DEEPRETAIL2020
2nd WORKSHOP ON DEEP UNDERSTANDING SHOPPER BEHAVIOURS AND INTERACTIONS IN INTELLIGENT RETAIL ENVIRONMENTS
Milan, Italy 11 January 2021

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

Alessandro Bruno[1], Stéphane Lancette[1], Jinglu Zhang[1], Morgan Moore[1], P Ville Ward[2], Jian Chang[1]

[1] National Centre for Computer Animation. Bournemouth University. Poole. UK

[2] Shoppar Ltd, Plexal, 14 East Bay Lane, Stratford, London. E20 3BS. UK

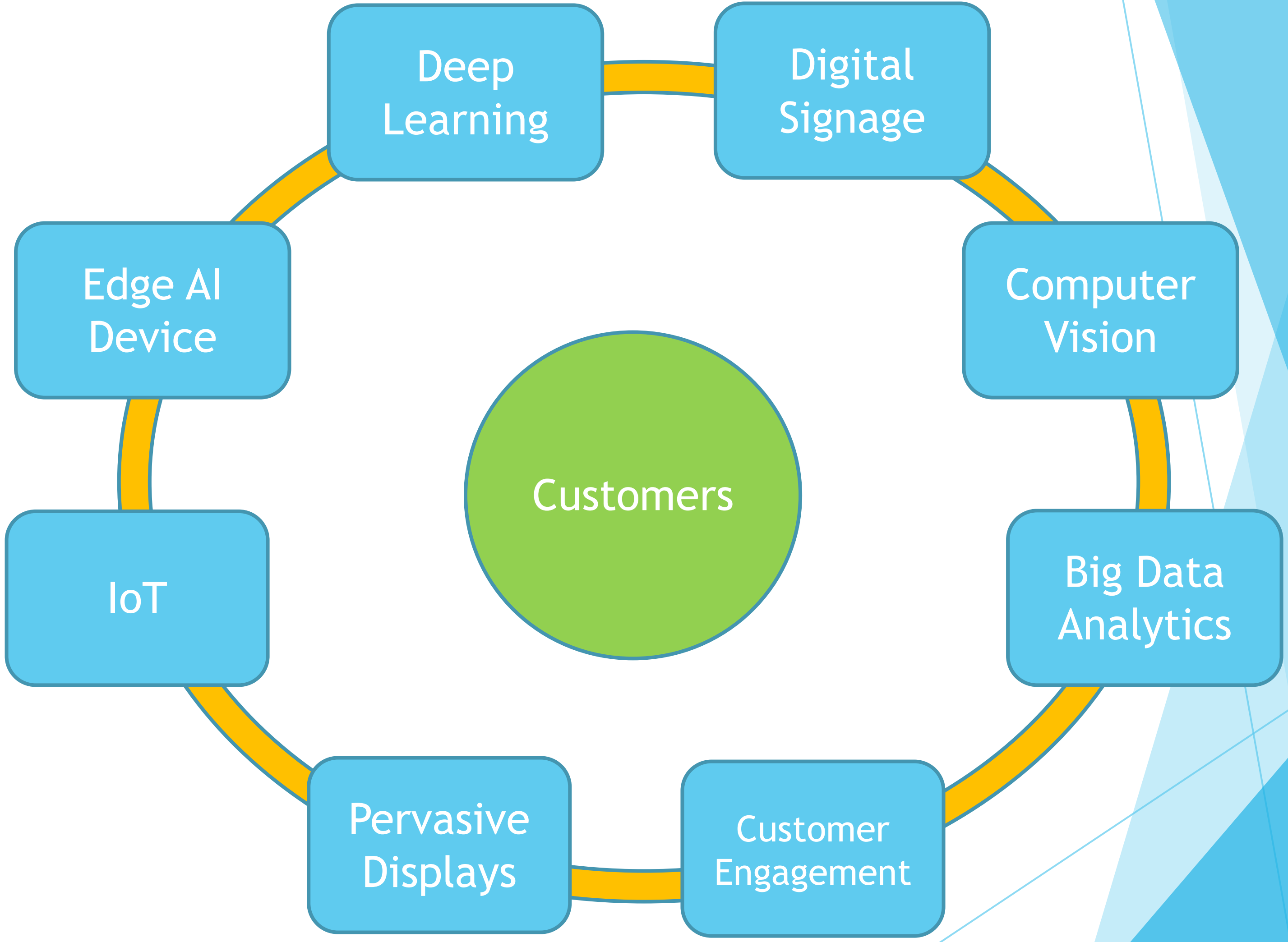
Bournemouth University

Shoppar®

Outline

- Customer Retail Environments
- Visual Saliency and Saliency Maps
- Proposed Method
  - Local and Overall Saliency to all Spatial Layout Permutations
  - Optimisation of Layout Content based on ES (Effectiveness Score)
- Use Case – 2 x 2 Grid-Based Advertisement Layout
- Webcam-based Eye-tracking to validate the proposed method
- Experimental Results
- Outlooks

Customer Retail Environment[1]



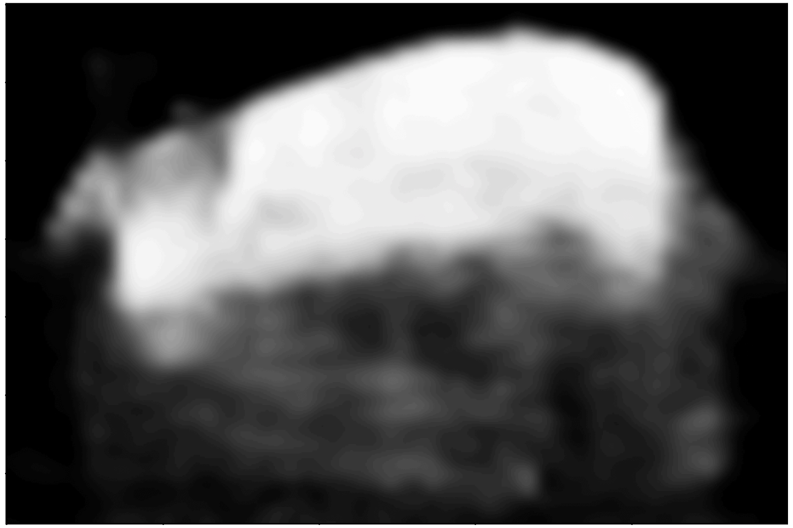
Visual Saliency

- Visual Saliency deals with detecting the most eye-catching regions in images, those regions which naturally stands out of the image. It accounts for bottom-up and top-down visual attention processes over the first few seconds of observation of a given image.

- “Visual saliency computation objective can be described as predicting, locating and mining the salient visual information by simulating the corresponding mechanisms in the human vision system.”[2]

Input Image

Saliency Map



Heatmap

“Given limited computational resources, the human visual system relies on saliency computation to quickly grasp important information from the excessive input from the visual world” [3]

Objectives

1. Set up a New Automatic System to predict the Human Visual System Behaviour of Customers when advertisements pop out;
2. Optimise Content Layout Configurations towards a well-balanced dwell times over each region of the advertising campaign;
3. Assess a direct correlation between the variance of salient local areas and dwell times of the same regions in the image;
4. Make the automatic solution ligthweight enough to be run on common laptops and devices

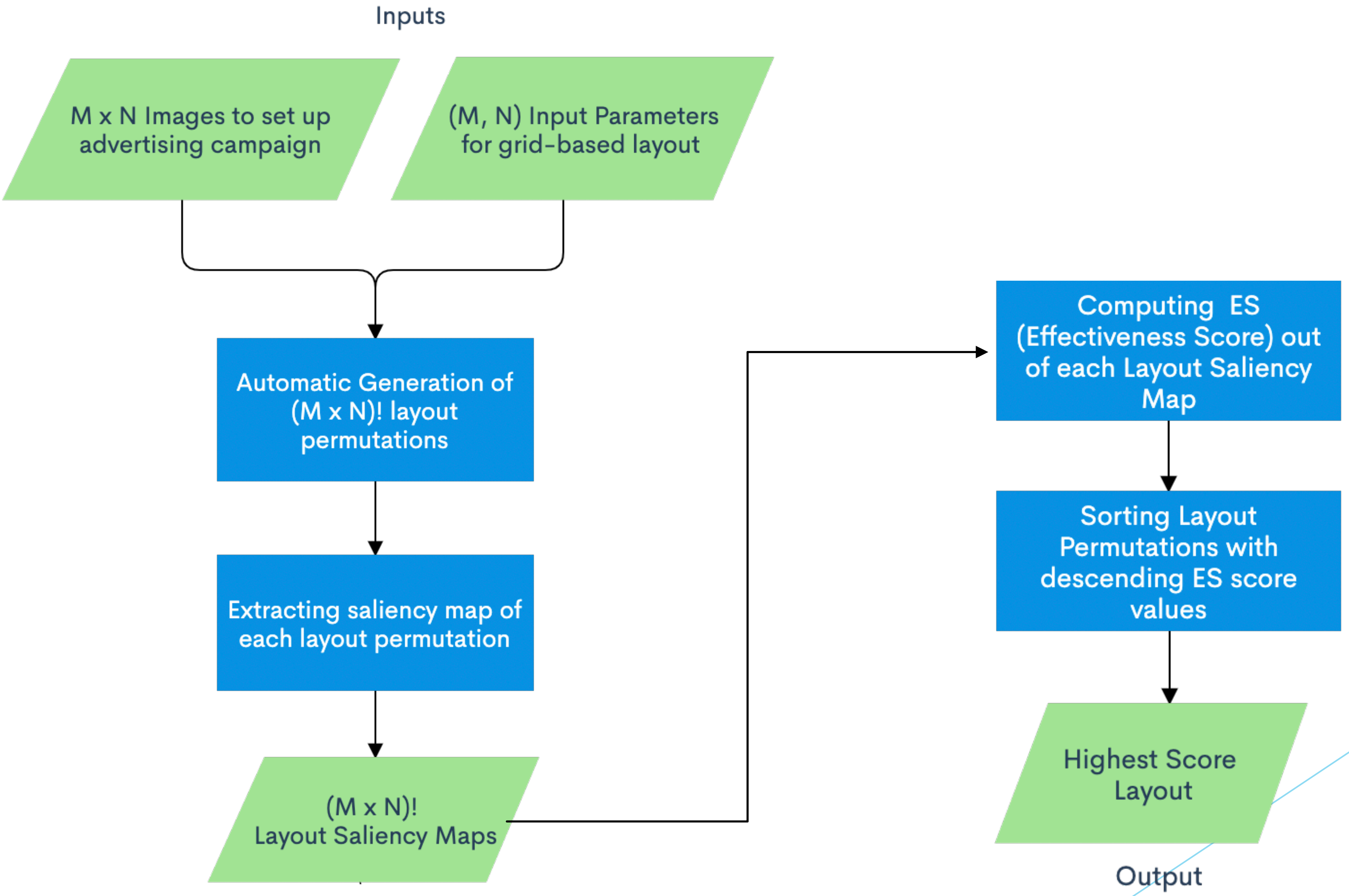
Proposed Method - Premise

- **Premise**: For a given image layout, for example a 2 by 2 grid-based layout, a number of 4! spatial permutations are given (it adds up to 24 spatial permutations).

- The first cue out of some preliminary experiments show different saliency ‘**behaviours**’ of the same regions whose image consists of:

- Some regions, such as the one with a red car, show different local saliency maps across different spatial permutations. Saliency Maps are extracted by using a **deep learning-based solution**[4] trained over an object-oriented image and video dataset called **DAVIS** [5].

Proposed Method – Algorithm (flow-chart)



Proposed Method

For a given layout made up of $M \cdot N$ images, the 'behaviour' of the overall layout saliency is studied by analysing the varying number of salient pixels on each of the $M \cdot N$ images.

In greater detail, the **inverse of the relative variance** of local saliency maps is employed 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_{th} layout content permutation.

$$ES_{(i)} = \frac{\left|\mu(NMSP_k(Layout_{(i)})\right|}{\sigma(NMSP_k(Layout_{(i)}))^2} \qquad (1)$$

$$k = [1, ..., (M \cdot N)] \quad i = \{1, ..., (M \cdot N)!\}$$

Proposed Method

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^{th} image (eq. 2)

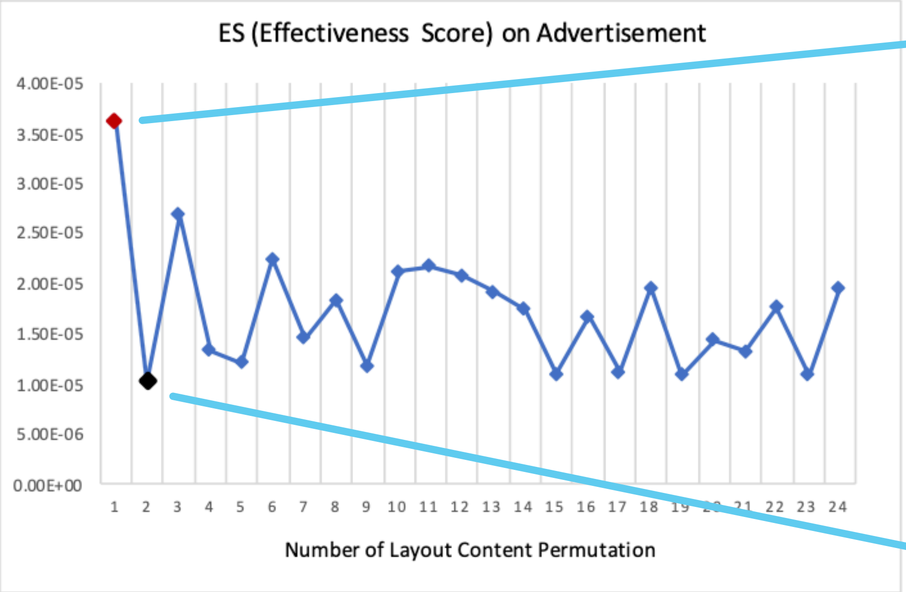
$$NMSP_{(h)} = \sum_{i,j \,\epsilon\, Im} LSM_{(h)}(i,j) \geq th \qquad (2)$$

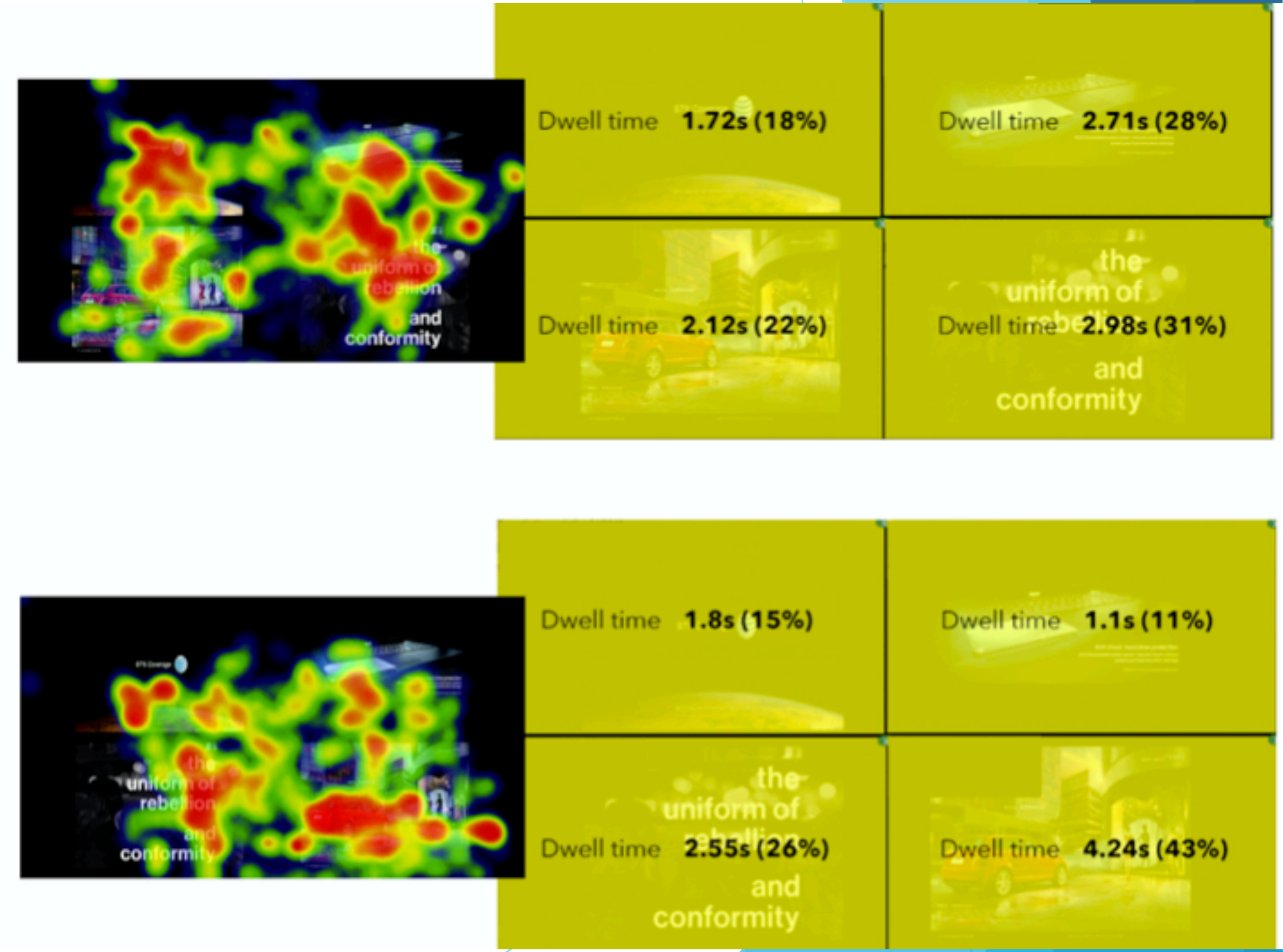
Each Layout content permutation is the union of $M \cdot N$ images $Im_{i'}$ as in equation 3

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

The layout showing the **highest score** is the output of the proposed method.

Experimental Results

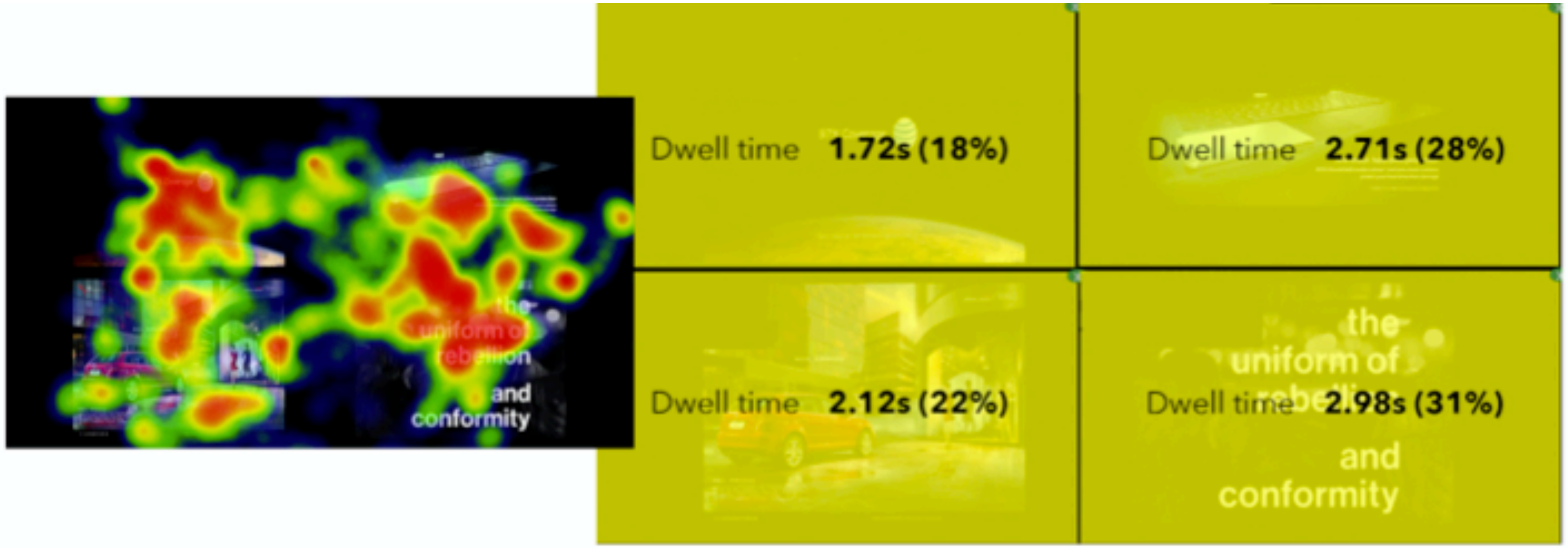




The experiments were extended to 5 advertising campaigns

Validation through eye-tracking sessions

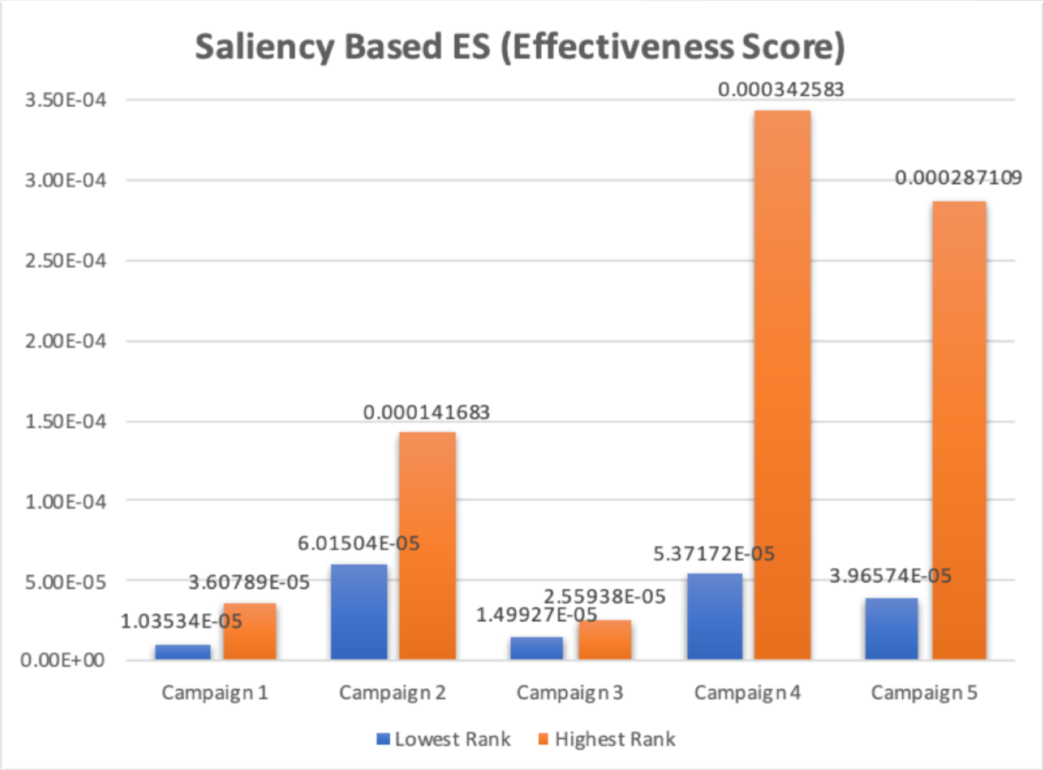
- Gazerecorder[7], a Webtool for webcam-based eye-tracking, was used to carry out the validation of the proposed method;
- 20 participants in the age range [25-40] were shown the layout content permutations with highest and lowest ES value of 5 graphical campaigns with images out of advertisement dataset [6];
- Each Image is shown for 10 seconds;
- Heatmaps and Dwell times are collected as shown below.
- Experiments were conducted to assess consistency between our results and eye-tracking session data which represent a ground-truth.

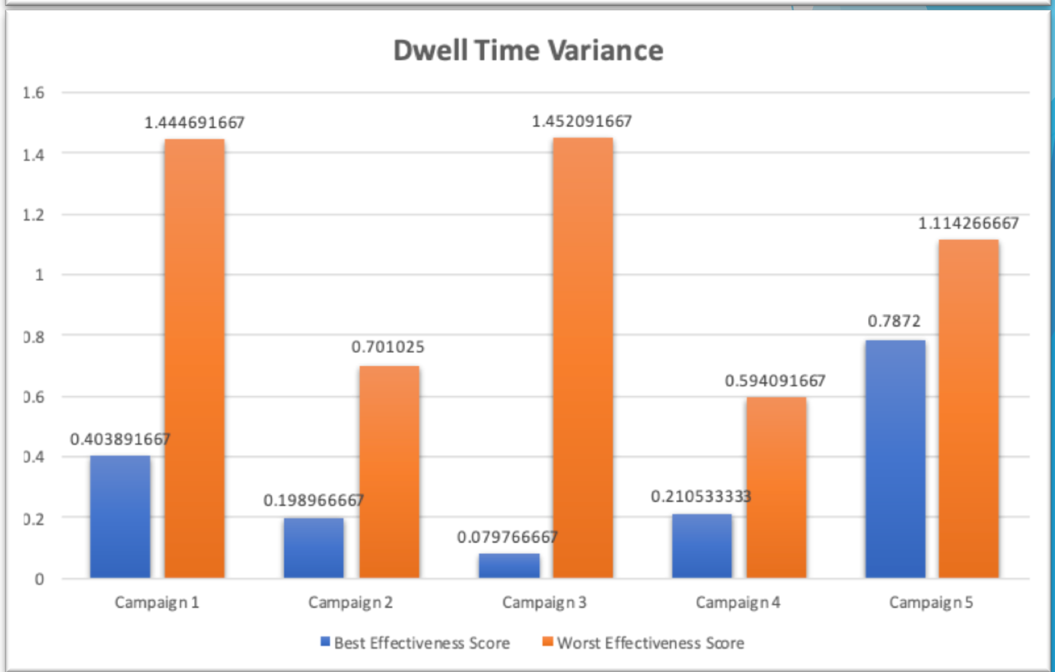


Experimental Results

The graph on the top shows the highest and lowest ES scores for each advertising campaign.

Images scoring best and worst ES values also show more balanced dwell times (graph on the bottom)





Experimental Results (settings)

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;

Average running time on 2-by-2 grid layouts is 40 seconds;

Python 3.8.0

TensorFlow 2.4.0 – Deep Learning Python Framework

Conclusions and Future Works

- In this work a new method for layout advertisement content optimizations is proposed to predict customers' behaviour in intelligent retail environments;
- The method is fully automatic and relies on three main steps:
  - Computation of all spatial permutations of the graphical elements of given advertising campaign;
  - Extraction of saliency maps of each permutation;
  - Computation of the relative variance of salient pixel number of local regions in images;
- As a study case, some experiments were conducted on 5 advertising campaigns and using a 2 by 2 grid based layout;
- Interesting matches are found between best ES scoring spatial configurations and the corresponding dwell times out of eye-tracking sessions with 20 participants.
- Further attention can be focused on the integration of scan-path prediction models to the current solution.
  - That way, both "spatial" and "time" aspects of visual attention will be used to go through advert optimisation.

Credits & References

- Credits

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

- References

[1] Dennis, Charles and Brakus, J Jovsko and Gupta, Suraksha and Alamanos, Eleftherios The effect of digital signage on shoppers' behavior: The role of the evoked experience, Journal of Business research, vol. 67, no.11, pages 2250--2257, (2014), Elsevier

[2] Li, Jia and Gao, Wen, Visual saliency computation: A machine learning perspective, vol. 8408, (2014), Springer

[3] Zhang, Jianming and Malmberg, Filip and Sclaroff Stan: Visual Saliency: From Pixel-Level to Object-Level Analysis. (2019). Springer

[4] Wang, W., Shen, J., Shao, L.: Video salient object detection via fully convolutional networks. IEEE Transactions on Image Processing 27(1), 38–49 (2017)

[5] 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)

[6] Hussain, Z., Zhang, M., Zhang, X., Ye, K., Thomas, C., Agha, Z., Ong, N., Ko- vashka, A.: Automatic understanding of image and video advertisements. In: Pro- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1705–1715 (2017) Url: http://people.cs.pitt.edu/kovashka/ads/

[7] Deja, S.: Gazerecorder. https://api.gazerecorder.com/

Thanks for your attention!

contacts

jchang@bournemouth.ac.uk

peter.ward@shopparapp.com

abruno@bournemouth.ac.uk

zhangj@bournemouth.ac.uk

stephane.lancette@gmail.com

s5113911@bournemouth.ac.uk