DETERMINATION OF THE OPTIMAL POSITION AND CONTENT FOR AN ONLINE ADVERTISEMENT USING SALIENCY DETECTION ALGORITHMS AND EYE-TRACKING TECHNIQUES

Aaishwarya Bansal; Jannat Kaur; Kavita Bhavsar; Meet Rawal
Department of Electrical and Computer Engineering
University of Waterloo

ABSTRACT

This paper focuses on different aspects of static display ads and how they can be made more effective. Various saliency detection and eye tracking algorithms have been used to analyze which features and positions can make an ad noticeable. Multiple ad templates and web pages were designed. A comparison algorithm was developed that compares the original image with the heat maps generated from saliency detection algorithms to detect the most efficient technique. The outcomes from the optimal algorithm were then compared with the outcomes from the eye tracking i.e. ground truth to subjectively analyse whether the results align.
Keywords - Saliency Detection, Eye-Tracking, Display advertisements

1. INTRODUCTION

Digital advertisements are a form of internet marketing that is used to create a brand value and promote the products/services of a firm. The end goal of advertising is to increase the sales and the revenue of the firm. Online advertisement is an effective way for a business to grow and expand its size. Hence, the firm can strategically devise methods to engage the potential customers and turn them into a loyal ones.

Digital ads can be of various types such as display ads, SEM ads, SMM ads, pay per click ads, video ads etc. Our main focus is on the display ads and its various types as they count to 47% percent of the total spending on advertising. [12] Display advertisements are the visual ads that appear on the sides of the screen or as pop-ups on any third party websites. These are the most basic form of online advertising and are usually very affordable.

During the early days, billboards and flyers were used as modes of marketing but it did not ensure the effectiveness of the method. With the advent of social media and various other online platforms, nearly 90% of the market and its audience has shifted online. Therefore, advertisers try to make sure that the ads are noticeable. There are many parameters which can be used to make sure that the people are actually looking at the ads. These parameters are the basis of our research which will help us understand and analyze the factors that can increase the chances of a firm to be recognised over the others that are running in parallel to it.

Type of display ads paired with the kind of content being advertised can greatly affect the performance and the impact of ad on the potential clients. [13] On an average, over 47% percent of the market of any firm is derived from display ads and nearly 50% of the people end up ignoring those ads because the content is not eye catching. The main problem faced by marketers is to target the right kind of audience whom they shall put their money on such that the revenue generated is greater than the investment. Most of the time, people end up ignoring the ads which results in lower utility than the expected utility at the beginning of the campaign. One way to resolve this problem is to track the movement of the eyes of various people and analyze what kind of ads are most effective and what part of the ad captures the attention of the end user.

2. METHODOLOGY

In this paper, we are focusing on making static display ads more effective as they are cheaper to bid on and require less memory for storage. The main parameters we are using for measuring the effectiveness of ads are displayed in table 1.

PARAMETER	CONTENT
Position	Top, Margins and merged with content
Size	Leaderboard, full banner, buttons, medium-large rectangles, skyscrapers, non-standard
Colors	Warm colors, cool colors
Content Type	Text,Image

Table 1: Parameters used for measuring the effectiveness of an ad[14].

We have created a dataset taking into consideration all the above parameters. Given the parameters, we created a total of 144 different templates that has covered all the possible combinations. These templates are then tested with 3 different industrially accepted saliency detection algorithms to detect the most prominent features of these ads. Once detected, then the eye tracking algorithm captures the gaze of human eyes to find which position is most effective to place the ad on. The outcome from the eye tracking is compared with the saliency detection algorithms to see if the heat maps align with each other i.e. to check the validity of saliency algorithms with actual eye tracking. Figure 1 explains the flow of methodology.

2.2. Saliency Detection

[1] Saliency Detection is an algorithm that uses image processing and computer vision techniques to locate the most prominent features of an image. It uses deep learning to find the salient region by extracting information about

relevant parts of an image. Whenever we look at a picture, there are certain things which are more prominent and we tend to pay more attention to those details. Saliency detection narrows down the important parts of an image and determines how quickly the user focuses on those relevant aspects. Advertisers are generally aware that people don't have long attention spans and therefore, they try to catch the user's eye with a single glance. Saliency detection techniques are used to design ads that have more noticeable content. Therefore, we define three different saliency detection algorithms that we used to compute what particular features in an advertisement make it more observable.

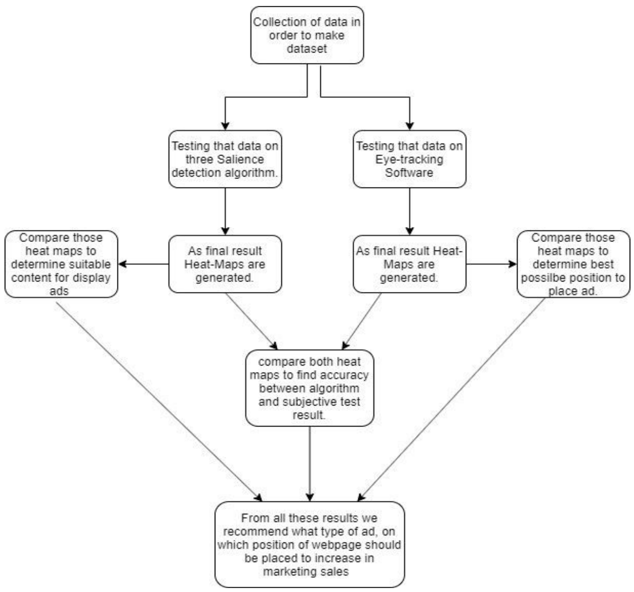


Figure 1: Flow of methodology

2.2.1. SalNet

[3] SalNet is a saliency detection microserve to integrate saliency detection in the image/video via a serverless API. For fuzzier images, the saliency detection algorithm uses an unsharp mask to sharpen the edges of the image, which increases its ability to extract the most prominent features in an image. [4] The SalNet's architecture involves shallow and deep layers. [5] The U-Net (SalNet's framework) is a convolutional neural network(CNN) that segments images more precisely and quickly. [6] U-Net architecture uses feature maps to convert an image into a vector and then uses the same feature maps to expand a vector to a segmented image. This technique preserves the structural integrity of an image and reduces distortion.

2.2.2. Deep Gaze 2

[7] Deep Gaze 2 uses VGG-19 deep neural network model to spot regions of interest in an image. The saliency is predicted by training a few read out layers on VGG features without additional fine-tuning. [11]It uses the technique in which the features are directly transferred from the deep neural network for the recognition of the object. According to [7], deep gaze 2 holds a benchmark under the MIT300 curve metrics. Also, after cross validating, deep gaze 2 has described nearly 87% of gain in patterns of fixations. Therefore, deep gaze 2 is a strong test of transfer learning.

2.2.3. Salicon

Salicon, short for Saliency in context, focuses on predicting and understanding visual attention using human intelligence and computer vision algorithms. [11] Salicon uses deep neural networks to prognosticate the probability distribution over the image of the position of gaze fixation and uses color as a feature to compute saliency accordingly to boolean maps that are generated by randomly choosing a threshold value for the feature maps.

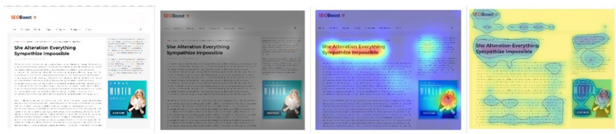


Figure 2 : (From left to right): Original image, Heat Map generated by Salicon, SalNet and DeepGaze 2 superimposed on the original image.

2.2.4 Heat Maps

The results of saliency detection algorithms are obtained as heat maps as shown in figure 2. [2] The main purpose of heat maps is to visualise the volume of locations in a dataset and assist viewers by directing them towards data visualizations that matter most. In the heat maps generated from the above three models are gray-scaled and transformed such that the computation on the later stages could be comparable.

2.3. Eye Tracking

To compute the optimal position of an ad on a webpage, we have analysed two algorithms that track the eye movement.

2.3.1. RealEye

RealEye uses the widely accepted theory of cross cultural universality of emotions to implement emotion detection. First, the process begins with calibration. Here, the system identifies how a tester's eyes move when they look at different parts of the screen. Then, the software creates an individual's mean face shape based on the calibration and uses that personalised definition of a neutral face as a base from which the relevant variations are measured. According to [9], RealEye has an accuracy of approximately 100px (~1.5cm) and an average error of nearly 4.17 deg on visual angle. RealEye is free, affordable and easy to use tool to see where you look in an image. The final results are in the form of heat maps.

2.3.2. GazeRecorder

GazeRecorder is a webcam based eye-tracking software which is used by various website developers, filming or advertising agencies to ensure the effectiveness and obtain analytics of their end product [15]. In this software, eye tracking starts with calibration, from which the accuracy of the eye tracking will be defined. Eye movements are recorded by the GazeRecorder and the final results are generated in two types of heat maps: one is a static heat map and the other is a dynamic heat map with adaptive time window.[15][8]On comparison it was found that the accuracy was greater than expected when the head was allowed to move freely. The measurements obtained from the above calibration were comparable to the ones generated using the eye tracking device with infrared light (SMI Red 250) [16].

3. EXPERIMENTAL SETUP

3.1. Dataset

We have designed a **novel** dataset for the execution of the project. It contains a total of 168 images that cover all the parameters that will be used for the evaluation and comparison.

For saliency detection, we have created 144 samples for ads which are detected for prominent features. The parameters for this part have been broadly categorised into three types i.e. size, colour and content. The sizes have been broadly defined into six categories: leaderboard (728X90 pixels), full banner (468X60 pixels), buttons (120X60 pixels), medium/large rectangles (330X228 pixels), skyscrapers (120X1600 pixels) and non standard dimension. The colors have been split into 2 major types i.e. cool and warm. Cool colors can be defined as light, pastel colours whereas warm colors are the ones which bright or difficult to look at. For the warm category, we considered red, orange and yellow and for cooler colours, we considered sky blue, light green and pastel purple. And the content of the ad is defined as the textual, image/animation data on the ad. Variations of size, colors and type of content are used for each webpage to generate the dataset. We have included every factor twice and therefore, we were able to generate 6*6*2*2 i.e. 144 test samples.

For eye tracking, we have created web pages based on position. The positions have been categorised as the top of the webpage, the sides of the webpage (both left and right) and the placement of the ads within the content of the page. Now, to capture the eye movement of the user and to determine the optimal position for placing the ad, we have created 24 unique web pages and then taken screenshots of the same which will help us determine the perfect position for the placement of the ads. Each of the templates contains three ads per page placed at three different positions. All of these ads are similar to each other in terms of the content but will be placed at different positions.

3.2. Evaluation Metrics

The metrics that we used for comparison of heat maps are Structural Similarity Index (SSIM) and Mean Squared Error (MSE).[10] MSE gives an average squared difference between the estimated value and the actual value. The SSIM is a method to measure the similarity between two images. It is a perceptual metric that quantifies image quality degradation caused by data compression, transmission, etc. SSIM puts everything in a scale of -1 to 1. A score of 1 means that the given images are similar and a score of -1 means they are different.

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2 \qquad \text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

4. RESULTS

4.1. Performance comparison of saliency detection algorithms

We test the three algorithms i.e. SalNet, DeepGaze 2 and Salicon on our dataset to generate heat maps.. Then we compute the MSE and SSIM by comparing the original image with the one generated with heat map. The one with least MSE and more SSIM is considered as the optimal algorithm for saliency detection of the most prominent features. Table 2 represents the average MSE and SSIM values for the three algorithms when tested on 144 samples.

ALGORITHMS	EVALUATION METRICS	
	MSE	**SSIM**
SALNET	13133.93	0.75
DEEP GAZE 2	11788.38	0.778
SALICON	13350.968	0.764
GAZE RECORDER	12066.12	0.78

Table 2: Average of the results generated by saliency detection algorithms and Gaze Recorder when tested on 144 samples.

4.2. Determination of the optimal position of advertisement in a webpage

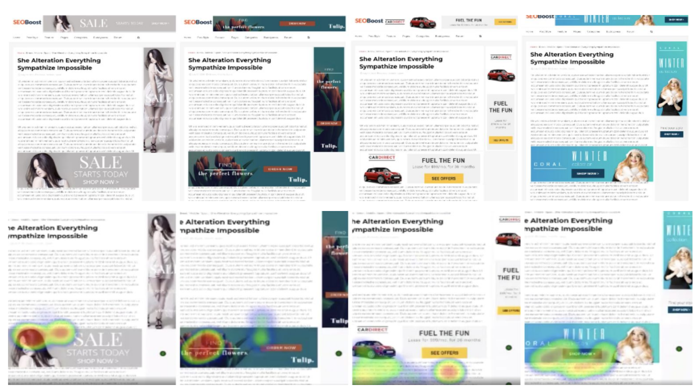


Figure 3: Four different web pages from the dataset for ad position comparison and heat maps generated by Gaze Recorder.

We considered two algorithms for eye-tracking i.e. RealEye and GazeRecorder. While conducting the eye tracking experiment, GazeRecorder resulted to be a better eye tracker than RealEye when tested. It generated a more accurate heat map of the region that we looked at. So, we compared the results of GazeRecorder to find the optimal position for ad placement.

For position comparison, we tested our dataset on 15 individuals for 3 seconds. We considered a 3 second period to generate the heat map of where the user looked at first or what region was seen quickest by the user. 13 out of the 15 generated heat maps showed that most people looked at ads that were placed around the content of the web page or at bottom. Figure 3 represents a few of the results.

5. ANALYSIS

5.1. Comparison with the State of the Art

The heat map that is generated when a real human sees a web page with an ad, highlights the region that was seen quickest by the user. The average values of MSE and SSIM for this eye tracking experiment are the ground truth with which we have compared the saliency detection algorithms.

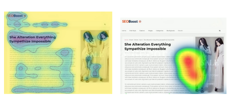
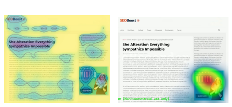
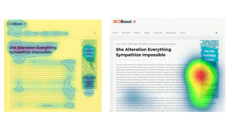
	1. Deep Gaze 2		**2.Gaze Recorder**	
Heat maps of 1 and 2:	**MSE**	**SSIM**	**MSE**	**SSIM**
	12423.42	0.77	13461.32	0.64
	11435.99	0.78	9452.22	0.81
	11178.49	0.79	10763.11	0.80
	12168.68	0.78	11432.91	0.78

Table 3: Results of Deep Gaze 2 and Gaze Recorder on specific web pages from the data set along with the heat maps.

Here, we compare the heat map generated by Gaze Recorder with the ones obtained from the three saliency detection algorithms to analyse if of the saliency detection algorithms align with the results of eye-tracking algorithm.

We tested the 144 ad content samples on the Gaze Recorder algorithm for heat map generation. We performed the test for 3 seconds on the same 15 humans, where each person was made to look at around 9-10 ad samples. Then we compute the MSE and SSIM by comparing the heat map and the original image and finally we take the average of the results for comparison. The result can be seen in table 2. Some unique cases have been compared in table 3 where MSE and SSIM values of SalNet and GazeRecoder for specific ads are displayed.

It is evident from table 2 that the average MSE and SSIM values generated by Gaze Recorder were close to the ones generated by SalNet, Deep Gaze 2 and salicon. There were variations in some cases but the average came out to be similar. Therefore, it can be said that the human eye tends to see the most salient features in an ad. So saliency detection is a good technique for the advertisers to implement and make their ads stand out and noticeable.

However, the variation in some cases show that sometimes even the algorithms fail to foresee what the user will see first in an image. Since the volume of available data for the prediction of saliency is scant, the intricacy of the deep architectures are difficult to perform better than the current state of the art.

5.2. Limitations and Future Scope

The limitations in this research are that we have considered some assumptions for our research. We are considering that the background of the webpage is either white or a light color (as defined for the majority of websites with bright colors for the content). We have conducted the experiment using non-blocking static ads and we are assuming that the ads are relevant to the content. The ground truth is the result obtained by executing eye-tracking on 15 people and our dataset contains screenshots of the web pages that we created for this experiment. So, further research can be done in this field by removing these constraints.

6. CONCLUSION

After analyzing all the above experiments and results, the following can be concluded: The saliency detection algorithms suggest that ads with warm colors, skyscraper sized and ones with images more than the text are more prominent and eye catching. The DeepGaze2 algorithm has proved to be most efficient saliency detection algorithm due to least MSE and its closeness to 1 for SSIM. The eye tracking algorithm suggests that the best position for placing the ad is it being placed within the content. The optimal position to place an advertisement is at the bottom of the web page. User tends to see an ad quickly when it was placed near the bottom, around the content of the website. Further analysis suggests that only the ads with prominent features were able to hold the attention of the user for more than 3 seconds. It can also be advised to place the ad within a frame such that it appears first hand on the screen over it appearing when it is scrolled down.

REFERENCES

[1] B. Kurtulmus, "Introduction to Image Saliency Detection," Algorithmia Blog, 24 January 2017. [Online]. Available:https://algorithmia.com/blog/introduction-image-saliency-detection. [Accessed 27 November 2019].

[2] W. contributors, "Heat map," Wikipedia, The Free Encyclopedia., 18 September 2019. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Heat_map&oldid=916275374. [Accessed 27 November 2019].

[3] S. Kim, "Improve Saliency Detection," Algorithmia Developer Center, 12 September 2019. [Online]. Available: https://algorithmia.com/developers/tutorials/recipes/sharpen_saliency. [Accessed 27 November 2019].

[4] Han L., Li X., Dong Y. (2018) SalNet: Edge Constraint Based End-to-End Model for Salient Object Detection. In: Lai JH. et al. (eds) Pattern Recognition and Computer Vision. PRCV 2018. Lecture Notes in Computer Science, vol 11259. Springer, Cham

[5] Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham

[6] H. Sankesara, "UNet - Introducing Symmetry in Segmentation," Medium, 23 January 2019. [Online]. Available: https://towardsdatascience.com/u-net-b229b32b4a71. [Accessed 27 November 2019].

[7] M. K. a. T. S. A. W. a. M. Bethge, "DeepGaze II: Reading fixations from deep features trained on object recognition," *ArXiv,* vol. abs/1610.01563, no. 1610.01563, 2016.

[8] GazeRecorder, "Track your eye movement and create video recordings of it by relying on this handy application that uses your webcam as an input source," GazeRecorder, 11 January 2019. [Online]. Available: https://gazerecorder.com/gazerecorder/. [Accessed 27 November 2019].

[9] A. Cellary, "Why Eye Tracking Software Is Right For You,"RealEye,[Online].Available:https://www.realeye.io/blog/why-eye-tracking/. [Accessed 27 November 2019].

[10] M. Binieli, "Machine learning: an introduction to mean squared error and regression lines," freeCodeCamp, 16 October2018.[Online].Available:https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/. [Accessed 27 November 2019].

[11] X. Huang, C. Shen, X. Boix and Q. Zhao, "SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 262-270.

[12] J. Brooke, "7 Types of Online Advertising," adSkills, 18 May 2019. [Online]. Available: https://www.adskills.com/7-types-of-online-advertising/. [Accessed 27 November 2019].

[13] Adnow, "Native advertising, the new marketing workhorse," Marketing Land, 15 November 2016. [Online]. Available: https://marketingland.com/native-advertising-new-marketing-workhorse-197856. [Accessed 27 November 2019].

[14] M. c. MarketingSherpa, "Sherpa 101: Online Display Ads - Types, Sizes, Rates & Metrics," MarketingSherpa, 28 November 2007. [Online]. Available: https://www.marketingsherpa.com/article/how-to/online-display-ads-types-sizes. [Accessed 27 November 2019].

[15] S. Deja, "Comparison of accuracy and precision of eye tracking: GazeFlow vs. SMI RED 250.," 11 August 2013. [Online]. Available: http://gazerecorder.com/Raport/. [Accessed 27 November 2019].

[16] GazeFlow, "GazeFlow Webcam Eye Tracking engine," GazeRecorder, 15 May 2019. [Online]. Available:https://gazerecorder.com/gazeflow/. [Accessed 27 November 2019].