Optimisation of Saliency-driven Image Content Ranking Parameters

Matthew Kenely

Supervisor: Dr. Dylan Seychell
Co-Supervisor: Prof. Ing. Carl James Debono

June 2024

Submitted in partial fulfilment of the requirements for the degree of Science in Information Technology (Honours) (Artificial Intelligence).

L-Università ta' Malta
Faculty of Information & Communication Technology

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris sed ipsum risus. Nulla aliquet quis quam sed eleifend. Donec rutrum, dolor id vulputate pharetra, nulla tortor laoreet nisl, pellentesque dapibus velit dolor suscipit purus. Phasellus vitae eleifend sem. Integer ultricies ex in neque pellentesque, vitae facilisis orci aliquam. In pellentesque mollis turpis, eu tristique lacus eleifend nec. Vestibulum orci neque, rhoncus vitae convallis eu, suscipit quis dui. Nulla libero elit, porta sit amet sagittis vel, placerat sit amet tortor. Aliquam hendrerit dolor sit amet sollicitudin ornare. Aliquam placerat sodales est, in vestibulum nisl efficitur in. Nulla venenatis aliquam sem, at volutpat nisl pellentesque eleifend. Praesent vitae euismod nulla, eget vehicula turpis. Duis quis tellus vitae nisi tempus tincidunt.

Nam quis aliquet nisi, non pharetra ligula. Phasellus pulvinar mattis neque, nec interdum justo condimentum hendrerit. Mauris fermentum venenatis faucibus. Pellentesque egestas eleifend libero, quis placerat ante fermentum et. Suspendisse accumsan gravida rhoncus. Vestibulum auctor sodales vehicula. Pellentesque a urna et elit placerat laoreet a quis turpis. Morbi ut sem at nunc posuere malesuada vel nec libero. Nam et urna suscipit, bibendum diam sed, aliquet leo. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut pretium mauris et nulla malesuada, mattis tristique lectus molestie. Aenean accumsan iaculis quam, eget varius libero placerat eu. Fusce mauris justo, vulputate a sollicitudin a, malesuada at est. Sed ac augue elit. Maecenas massa lorem, tincidunt vitae neque et, maximus dictum tellus.

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris sed ipsum risus. Nulla aliquet quis quam sed eleifend. Donec rutrum, dolor id vulputate pharetra, nulla tortor laoreet nisl, pellentesque dapibus velit dolor suscipit purus. Phasellus vitae eleifend sem. Integer ultricies ex in neque pellentesque, vitae facilisis orci aliquam. In pellentesque mollis turpis, eu tristique lacus eleifend nec. Vestibulum orci neque, rhoncus vitae convallis eu, suscipit quis dui. Nulla libero elit, porta sit amet sagittis vel, placerat sit amet tortor. Aliquam hendrerit dolor sit amet sollicitudin ornare. Aliquam placerat sodales est, in vestibulum nisl efficitur in. Nulla venenatis aliquam sem, at volutpat nisl pellentesque eleifend. Praesent vitae euismod nulla, eget vehicula turpis. Duis quis tellus vitae nisi tempus tincidunt.

Nam quis aliquet nisi, non pharetra ligula. Phasellus pulvinar mattis neque, nec interdum justo condimentum hendrerit. Mauris fermentum venenatis faucibus. Pellentesque egestas eleifend libero, quis placerat ante fermentum et. Suspendisse accumsan gravida rhoncus. Vestibulum auctor sodales vehicula. Pellentesque a urna et elit placerat laoreet a quis turpis. Morbi ut sem at nunc posuere malesuada vel nec libero. Nam et urna suscipit, bibendum diam sed, aliquet leo. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut pretium mauris et nulla malesuada, mattis tristique lectus molestie. Aenean accumsan iaculis quam, eget varius libero placerat eu. Fusce mauris justo, vulputate a sollicitudin a, malesuada at est. Sed ac augue elit. Maecenas massa lorem, tincidunt vitae neque et, maximus dictum tellus.

Contents

1 Introduction

In 2018, Seychell *et al.* developed **SaRa** [1], an approach which allows for the automatic ranking of the saliency of objects within images without the prior need of a trained model. This approach splits images into a grid and processes each segment individually. The aim of this project is to analyse how varying the size of this grid, as well as the resolution of the input images, affects the generated content ranking results, and optimise the parameters of **SaRa** accordingly.

All images presented in this project are taken from the **COTS** Dataset [2] or taken by myself.

2 Background

2.1 Saliency

This project is primarily concerned with saliency, a computer vision problem where the goal is to automatically characterise parts of an image which stand out relative to neighbouring regions [3] and attract visual human attention.

Saliency detection is usually carried out in one of four approaches: eye-fixation, traditional, information theory and deep learning. Convolutional Neural Network (CNN) architectures are commonly used and are trained using reference images paired with manually annotated corresponding saliency maps (an n-bit grayscale image where pixel values denote the level of saliency of a particular region, with 0 implying no saliency, and n implying maximum saliency).

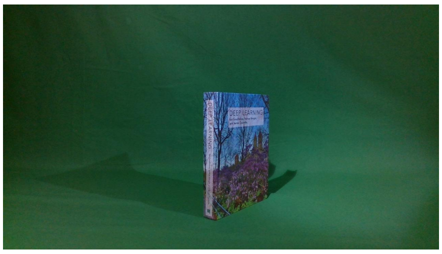

Figure 2.1 Reference image of a book (subject) against a green background

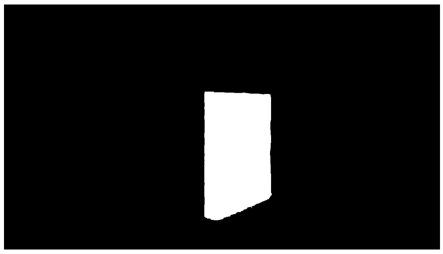

Figure 2.2 Corresponding saliency mask for the image in Figure 2.1

2.2 Entropy

Generally, entropy (denoted by H) is a physical property associated with a state of disorder, randomness, or uncertainty. When applied to the context of saliency, entropy refers to the measure of uncertainty or randomness in a saliency map, i.e. the extent to which pixel values are uniformally distributed in the map [4].

Shown in Equation 2.1 is Shannon's entropy equation.

$$H(X) = -\sum_{i=1}^{|t|} P\left(x_i\right) \log_2\left(P\left(x_i\right)\right) \tag{2.1}$$

In [1], rather than applying Shannon's entropy to the entirety of a saliency map, it is used to calculate a score for each of the individual grid segments and rank them accordingly.

2.3 Literature Review

2.3.1 Ranking regions of visual saliency in RGB-D content

In **Ranking regions of visual saliency in RGB-D content** [1], Seychell *et al.* propose a technique (**SaRa**) which identifies salient objects in images without the use of a pre-trained model and which outperforms other saliency algorithms with the aim of facilitating the creation of more immersive 3D environments with low real-time computational costs.

SaRa works in two stages:

1. Texture and depth images are inputted, as well as the desired size of the grid T to be overlayed on the original image. A saliency map is generated following Itti's system [5] and the implementation in [6]. This saliency map, as well as the depth map are processed according to T.

2. Each segment t in T is processed and a score is calculated, comprising proximity to the image center (to factor in center-bias [7]), entropy calculated using Equation 2.1 and corresponding values from the depth map. The segments are ranked in descending order of this score.

2.3.2 Information Theory

Bruce *et al.* [4] propose an approach to saliency measure automation based on Shannon's self-information measure and modelled after the human visual cortex.

3 Methodology

I used the Python code developed by Seychell *et al.* as a basis for this project, restructured it, removed redundant variables and one-line functions, added function dosctrings and reformatted it to conform to the PEP-8 style guide. The source code for this project is available **here**.

3.1 Python scripts

3.1.1 main.py

```python
def main():
    n = 25

    path = '../COTS Dataset/Part 1 - Single Objects/objects/'
    name = 'cmt_mug_colour.jpeg'
    im = path + name

if __name__ == "__main__":
    main()
```

Listing 3.1 Python example

4 Experimentation

4.1 Experiment 1 – Average Entropy Maximisation

In this experiment, I implemented a Python script which runs the **SaRa** algorithm on different images at different grid sizes (from 5×5 up to 14×14). The average entropy of the top 25% highest-ranked grid segments at each grid size configuration is calculated in order to determine which grid size gives the largest proportion of entropy for a given image.

$$\begin{aligned} G &: \text{Grid sizes } 5 \times 5 \text{ to } 14 \times 14 \\ H &: \text{Entropy} \\ I_g &: \text{Image with } \textbf{SaRa} \text{ grid applied} \\ S &: \text{Highest ranked grid segments} \end{aligned}$$

$$\operatorname*{argmax}_{g \in G} \sum_{i=1}^{\frac{1}{4} \times g^2} \mathrm{H}(I_g\{S[i]\})$$

Below is a comparison of the results of this approach being applied to two different images. I chose these images specifically to demonstrate that this approach is able to adapt grid size according to how large the salient objects in the scene are in proportion to the size of the image – the tagine in `tajin_colour.jpeg` is larger than the mug in `cmt_mug_colour.jpeg`.

Figure 4.1 tajin_colour.jpeg

Figure 4.2 cmt_mug_colour.jpeg

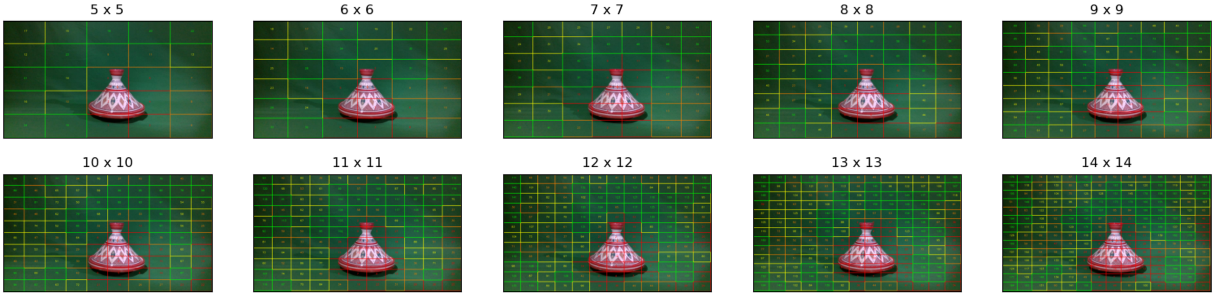

Figure 4.3 SaRa implemented on **tajin_colour.jpeg** using grid sizes from 5×5 up to 14×14

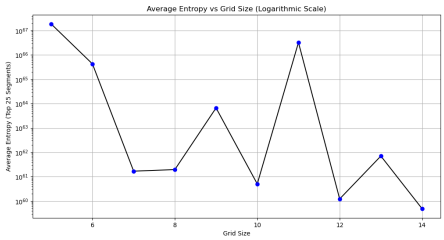

Figure 4.4 Average entropy vs Grid Size in **tajin_colour.jpeg**

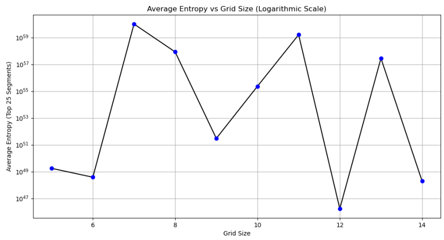

Figure 4.5 Average entropy vs Grid Size in **cmt_mug_colour.jpeg**

There are peaks in the average amount of entropy generated when the grids align to the salient objects in the images in such a way that the entropy within each segment is maximised.

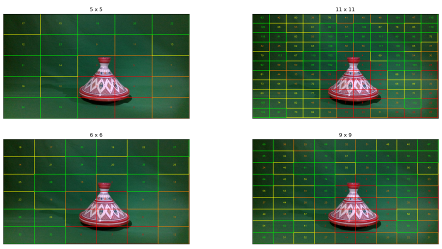

Figure 4.6 Top 4 grid sizes which result in the largest average amount of entropy in **tajin_colour.jpeg**

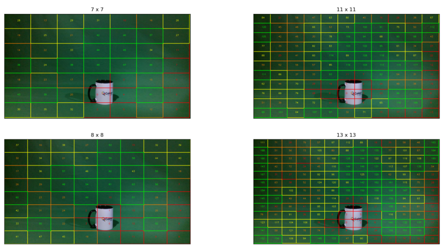

Figure 4.7 Top 4 grid sizes which result in the largest average amount of entropy in **cmt_mug_colour.jpeg**

Shown in Figure 4.8 and Figure 4.9 are the optimal grid sizes for maximising the amount of entropy generated in each of the grid segments. It is notable that the tagine (the larger object) has a smaller grid size, i.e. each segment is larger.

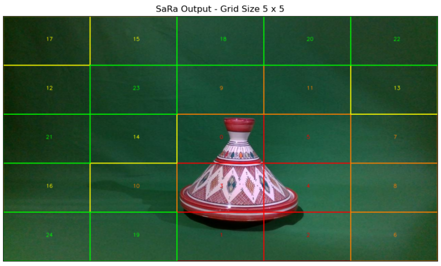

Figure 4.8 tajin_colour.jpeg with maximised entropy grid size (5×5)

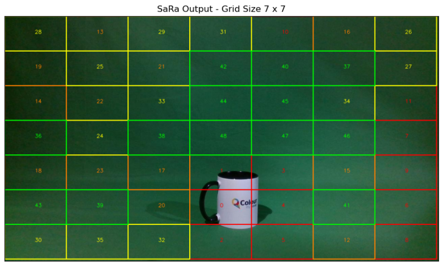

Figure 4.9 cmt_mug_colour.jpeg with maximised entropy grid size (7×7)

4.2 Experiment 2 – Resolution Variation

4.2.1 Fixed Grid Size

4.2.2 Average Entropy Maximisation

4.3 Experiment 3 – Flower Sepal Abnormalities

Figure 4.10 Botrytis cinerea (gray mold) on a flower

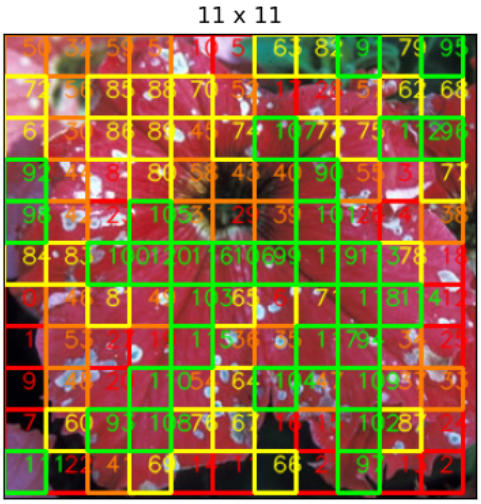

Figure 4.11 Corresponding **SaRa** grid (increased entropy throughout the image, especially at lower segment sizes)

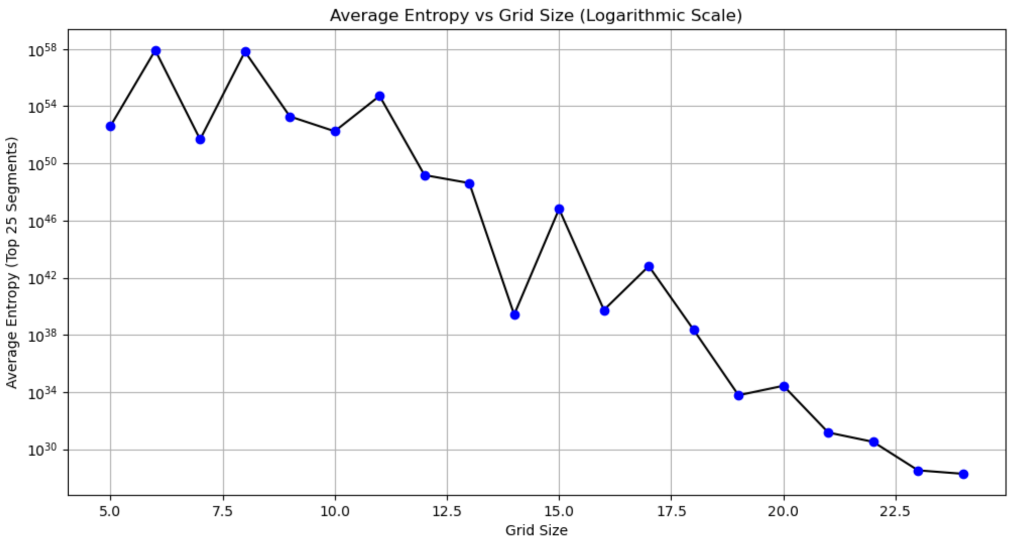

Figure 4.12 Average entropy vs Grid Size in the unhealthy flower

Figure 4.13 Healthy flower

Figure 4.14 Corresponding **SaRa** grid (uniform lack of saliency except in the middle where the pistil is present, i.e. low entropy)

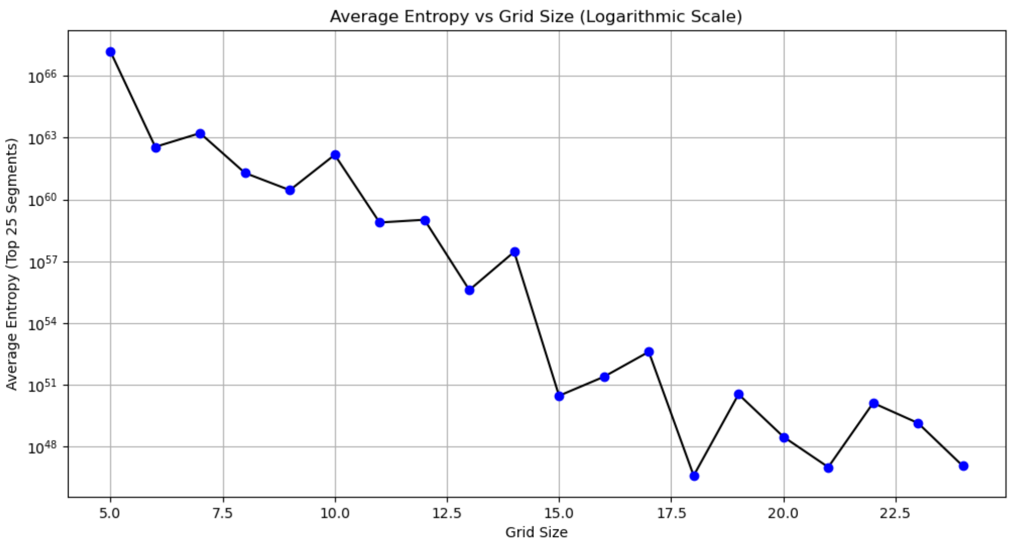

Figure 4.15 Average entropy vs Grid Size in the healthy flower

5 Evaluation

References

[1] D. Seychell and C. J. Debono, “Ranking regions of visual saliency in RGB-D content,” en, 2018, Accepted: 2022-03-01T17:18:37Z Publisher: IEEE. DOI: 10.1109/IC3D.2018.8657902. [Online]. Available: https://www.um.edu.mt/library/oar/handle/123456789/90087 (visited on 07/11/2023).

[2] D. Seychell, C. J. Debono, M. Bugeja, J. Borg, and M. Sacco, “COTS: A Multipurpose RGB-D Dataset for Saliency and Image Manipulation Applications,” *IEEE Access*, vol. 9, pp. 21 481–21 497, 2021, Conference Name: IEEE Access, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3055647.

[3] A. Borji and L. Itti, “State-of-the-Art in Visual Attention Modeling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, Jan. 2013, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, ISSN: 1939-3539. DOI: 10.1109/TPAMI.2012.89.

[4] N. Bruce and J. Tsotsos, “Saliency Based on Information Maximization,” in *Advances in Neural Information Processing Systems*, vol. 18, MIT Press, 2005. [Online]. Available: https://proceedings.neurips.cc/paper/2005/hash/0738069b244a1c43c83112b73514 Abstract.html (visited on 07/13/2023).

[5] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998, ISSN: 1939-3539. DOI: 10.1109/34.730558.

[6] A. Kimura, *pySaliencyMap*, May 2023. (visited on 07/13/2023).

[7] T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to predict where humans look,” in *2009 IEEE 12th International Conference on Computer Vision*, Sep. 2009, pp. 2106–2113. DOI: 10.1109/ICCV.2009.5459462.

Appendix A Sample A

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris sed ipsum risus. Nulla aliquet quis quam sed eleifend. Donec rutrum, dolor id vulputate pharetra, nulla tortor laoreet nisl, pellentesque dapibus velit dolor suscipit purus. Phasellus vitae eleifend sem. Integer ultricies ex in neque pellentesque, vitae facilisis orci aliquam. In pellentesque mollis turpis, eu tristique lacus eleifend nec. Vestibulum orci neque, rhoncus vitae convallis eu, suscipit quis dui. Nulla libero elit, porta sit amet sagittis vel, placerat sit amet tortor. Aliquam hendrerit dolor sit amet sollicitudin ornare. Aliquam placerat sodales est, in vestibulum nisl efficitur in. Nulla venenatis aliquam sem, at volutpat nisl pellentesque eleifend. Praesent vitae euismod nulla, eget vehicula turpis. Duis quis tellus vitae nisi tempus tincidunt.

Nam quis aliquet nisi, non pharetra ligula. Phasellus pulvinar mattis neque, nec interdum justo condimentum hendrerit. Mauris fermentum venenatis faucibus. Pellentesque egestas eleifend libero, quis placerat ante fermentum et. Suspendisse accumsan gravida rhoncus. Vestibulum auctor sodales vehicula. Pellentesque a urna et elit placerat laoreet a quis turpis. Morbi ut sem at nunc posuere malesuada vel nec libero. Nam et urna suscipit, bibendum diam sed, aliquet leo. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut pretium mauris et nulla malesuada, mattis tristique lectus molestie. Aenean accumsan iaculis quam, eget varius libero placerat eu. Fusce mauris justo, vulputate a sollicitudin a, malesuada at est. Sed ac augue elit. Maecenas massa lorem, tincidunt vitae neque et, maximus dictum tellus.

Appendix B Sample B

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris sed ipsum risus. Nulla aliquet quis quam sed eleifend. Donec rutrum, dolor id vulputate pharetra, nulla tortor laoreet nisl, pellentesque dapibus velit dolor suscipit purus. Phasellus vitae eleifend sem. Integer ultricies ex in neque pellentesque, vitae facilisis orci aliquam. In pellentesque mollis turpis, eu tristique lacus eleifend nec. Vestibulum orci neque, rhoncus vitae convallis eu, suscipit quis dui. Nulla libero elit, porta sit amet sagittis vel, placerat sit amet tortor. Aliquam hendrerit dolor sit amet sollicitudin ornare. Aliquam placerat sodales est, in vestibulum nisl efficitur in. Nulla venenatis aliquam sem, at volutpat nisl pellentesque eleifend. Praesent vitae euismod nulla, eget vehicula turpis. Duis quis tellus vitae nisi tempus tincidunt.

Nam quis aliquet nisi, non pharetra ligula. Phasellus pulvinar mattis neque, nec interdum justo condimentum hendrerit. Mauris fermentum venenatis faucibus. Pellentesque egestas eleifend libero, quis placerat ante fermentum et. Suspendisse accumsan gravida rhoncus. Vestibulum auctor sodales vehicula. Pellentesque a urna et elit placerat laoreet a quis turpis. Morbi ut sem at nunc posuere malesuada vel nec libero. Nam et urna suscipit, bibendum diam sed, aliquet leo. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut pretium mauris et nulla malesuada, mattis tristique lectus molestie. Aenean accumsan iaculis quam, eget varius libero placerat eu. Fusce mauris justo, vulputate a sollicitudin a, malesuada at est. Sed ac augue elit. Maecenas massa lorem, tincidunt vitae neque et, maximus dictum tellus.