Can AI mimic human visual attention to assess e-commerce landing page engagement?

Lisa Colombo[1] and Alessandro Bruno[1]

[1] *IULM University, Via Carlo Bo 1, 20143 Milan, Italy*

Abstract

This paper presents a study on artificial intelligence (AI) models applied to extract visual saliency from images. In particular, the research assesses the accuracy of AI in replicating human-like attention mechanisms by comparing AI-generated saliency maps with eye movement data captured through eye-tracking technology. A case study is conducted to evaluate landing page engagement with viewers. Saliency maps of banners from 4 e-commerce landing pages are extracted with TranSalNet, an AI-based Visual Saliency model and compared to eye movements recorded with a webcam-based eye-tracking platform. Normalised Scanpath Saliency (NSS), Kullback-Leibler Divergence (KL-Div), and Area Under the Curve (AUC) metrics reveal AI models performing well in central regions of visual stimuli while exhibiting some false positives and false negatives in peripheral areas. The study offers insights into visual attention and e-commerce landing page assessment from a computational viewpoint.

Keywords

Artificial Intelligence, Visual Saliency, Eye Movements, e-commerce, landing page.

1. Introduction

In recent years, artificial intelligence (AI) [1] has made significant advances in replicating human cognitive functions, particularly in visual attention. Visual saliency [2], a critical aspect of human visual perception, refers to the ability of the human visual system to selectively focus on specific regions within a scene based on their distinct features, such as colour, brightness, or contrast. This ability allows humans to efficiently navigate complex visual environments by directing attention to the most relevant or noticeable objects. Understanding and predicting visual saliency has compelling applications, from image recognition to enhancing user experiences in digital interfaces, including websites and mobile apps.

Visual saliency prediction models have evolved from early biologically inspired approaches, such as Itti et al.'s saliency map model [3], to more sophisticated AI-based models that leverage deep learning to simulate human attention. These models attempt to predict the areas of an image or visual scene that likely capture human attention by analysing low-level features (such as edges, textures, and colours) and, in some cases, integrating higher-level cognitive elements, such as prior knowledge or task relevance.

In e-commerce, understanding what captures users' attention is crucial for optimising website design and improving user experience [4]. With the growing complexity of digital interfaces, knowing how users interact with visual elements like product images, banners, and call-to-action buttons can significantly impact a platform's effectiveness in driving user engagement and conversions. Accurate visual attention prediction can help e-commerce sites optimise the placement of key elements, ensuring that users quickly find what they are looking for and are more likely to engage with the platform.

[1] lisa.colombo29@studenti.iulm.it (L. Colombo); alessandro.bruno@iulm.it (A. Bruno)
0000-0003-0707-6131

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

This paper investigates AI's capabilities in predicting human visual attention on e-commerce websites through a study that compares AI-generated saliency maps with human attention data captured via eye-tracking technology. Using TranSalNet [5], a state-of-the-art saliency prediction model, we evaluated how well the model replicates human attention across various e-commerce platforms, including Amazon, eBay, Shein, and Vinted. The study focuses on identifying key areas of user interest and examines the accuracy of the AI model in predicting attention on visually prominent and peripheral elements.

This work aims to contribute to the growing field of AI-driven user experience optimisation by exploring the alignment between AI predictions and human attention patterns. The findings highlight the strengths and limitations of current AI models in saliency prediction and suggest areas for further refinement to capture the complexity of user interactions in digital environments.

2. Related Techniques

Visual saliency modelling techniques are typically grouped into bottom-up and top-down approaches, crucial in predicting where humans look in a visual scene. Bottom-up techniques rely on the inherent properties of the image itself, such as colour, contrast, and intensity, to determine which areas are more likely to attract attention. These stimulus-driven methods do not account for the observer's goals or prior knowledge. In contrast, top-down techniques incorporate higher-level cognitive processes, such as task relevance and user intent, influencing attention based on expectations and goals rather than purely visual cues.

In early bottom-up models, such as the one proposed by Itti et al. [3], saliency is determined by combining different low-level features through centre-surround mechanisms. The model computes the Difference of Gaussian (DoG) for intensity, colour, and orientation to detect regions of local contrast. This model assumes that areas of high contrast across these features are more likely to attract attention. The (DoG) operation is represented as in equation 1:

$$DoG(x,y) = G_{\sigma_1}(x,y) - G_{\sigma_2}(x,y) \tag{1}$$

$G_{\sigma_1}(x,y)$ and $G_{\sigma_2}(x,y)$ are Gaussian functions with standard deviations σ_1 and σ_2, respectively. This difference between the Gaussian functions enables the model to highlight high-contrast regions, typically areas that draw human attention.

As research advanced, top-down approaches [6] were introduced to account for the observer's intent, task relevance, and context. These models incorporate feedback mechanisms that adjust saliency predictions based on high-level cognitive factors. For example, a user searching for a specific product in an online store will focus on elements such as product images, search bars, and filters, even if these elements are not the most visually salient according to a bottom-up approach. By integrating the observer's goals, top-down methods complement bottom-up processes and offer a more comprehensive understanding of attention.

In addition to these foundational models, Hou and Zhang [7] introduced a novel frequency-based approach to saliency that operates in the frequency domain rather than the spatial domain. Their model suggests that the spectral residual, which captures unpredictable or irregular aspects of an image, is crucial in determining visual saliency. This approach uses the Fourier transform to separate an image's amplitude and phase components, focusing on the spectral residual to highlight areas that differ from the surrounding content. The saliency map is then generated based on this residual information (see equation 2):

$$S\ (x, y) = F^{-1}(log\ log\ \left(A(\omega)\right) - log\ log\ (\underline{A}(\omega)))\ \ \cdot e^{i\phi(\omega)} \qquad (2)$$

In equation 2 F^{-1} denotes the inverse Fourier transform, $A(\omega)$ the amplitude spectrum of the image, and $\phi(\omega)$ the phase spectrum. The spectral residual, $log\ log\ \left(A(\omega)\right) - log\ log\ (\underline{A}(\omega))$, captures the irregularities in the image that contribute to its saliency.

The frequency-based approach can highlight unpredictable elements in an image, making it highly effective in identifying salient regions that traditional spatial-based methods may miss.

As visual saliency research evolved, more sophisticated methods emerged, combining bottom-up and top-down processes. The Graph-Based Visual Saliency (GBVS) model by Harel et al. [8] takes a global approach, representing an image as a fully connected graph where nodes represent different regions and edges are weighted by visual similarity. A random walker algorithm is applied to identify globally unique areas, offering a more holistic understanding of saliency that considers the image's overall structure.

With the advent of deep learning, Convolutional Neural Networks (CNNs) [9] have dramatically improved saliency prediction by learning low-level and high-level features directly from data. These models combine the strengths of both bottom-up and top-down approaches. They use the stimulus-driven nature of bottom-up processes to detect contrasts and edges while leveraging high-level information such as object categories, context, and user goals to fine-tune saliency predictions. CNNs have proven effective in dynamic environments such as e-commerce platforms, where understanding user attention is critical for optimising design and enhancing user interaction.

Tliba et al. [10] proposed SatSal (Self-Attention Saliency), an encoder-decoder deep learning model that leverages skip connections during decoding to account for high and low-level features from images. SAtSal also relies on convolutional self-attention modules connecting the encoder to the decoder branches.

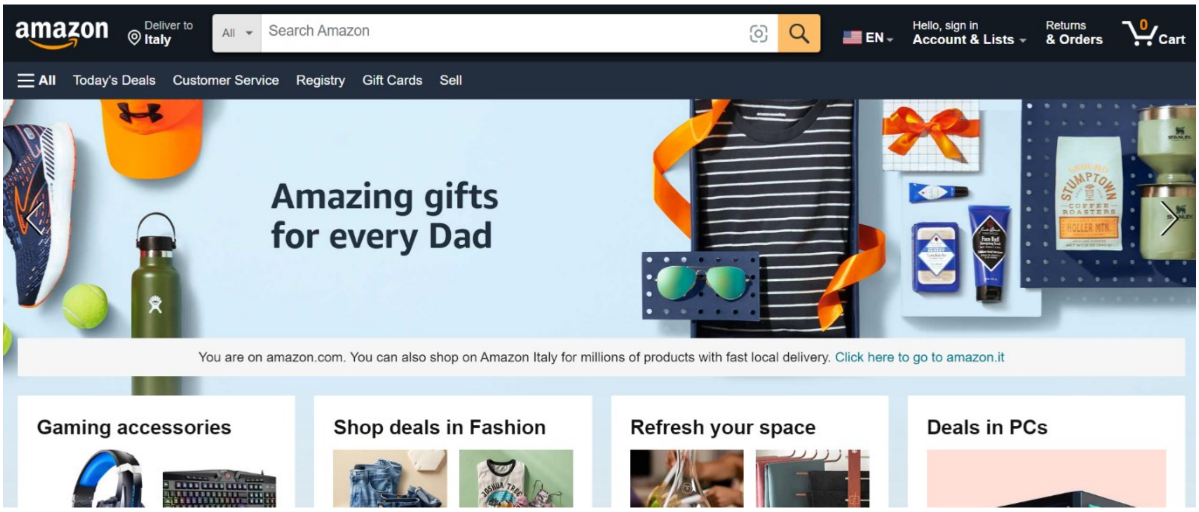
TranSalNet leverages dense connections and residual networks, which helps the model maintain spatial detail while reducing computational complexity. This approach allows the model to capture high-level semantic features (such as object categories) and low-level features (like colour and contrast), making it well-suited for complex, dynamic environments such as e-commerce websites.

In addition to the visual attention domain, saliency models can be used in many applications, such as image segmentation [11], video summarisation [12], image content enhancement [13], and automatic image cropping [14].

3. Method and Materials

In this study, we aimed to evaluate the performance of an AI model, TranSalNet [5], in predicting visual saliency on various e-commerce platforms by comparing its output with human eye-tracking data. The experiment involved 97 participants aged between 20 and 35, all with prior experience in online shopping. Each participant was asked to use Realeye.io eye tracking platform, that will present participants the chosen homepage images for a maximum of 4 seconds each. Each participant begins the experiment with a calibration process to ensure the eye tracker accurately records their gaze. The calibration involves having the participant look at a series of points on the screen to establish a baseline for their eye movements. Once calibrated, participants are shown a series of e-commerce homepage images for a maximum of 4 seconds each using the realeye.io eye tracking platform. This brief exposure period is designed to simulate the real-world scenario where users quickly scan a webpage to form an initial impression.

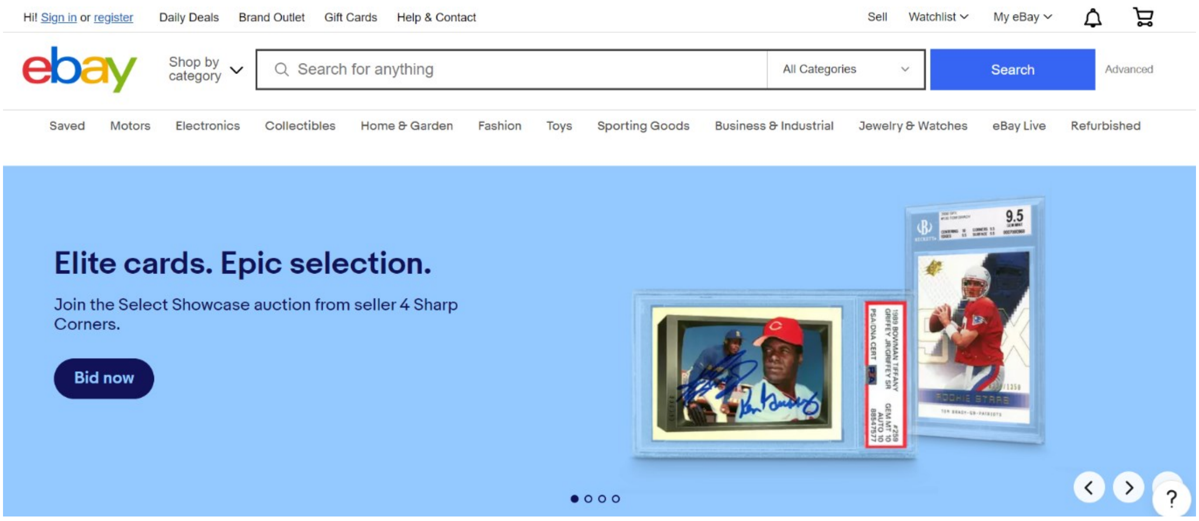
These are the images shown to the participants:



Amazing gifts for every Dad

You are on amazon.com. You can also shop on Amazon Italy for millions of products with fast local delivery. Click here to go to amazon.it

Figure 1: *Amazon's homepage*



Elite cards. Epic selection.

Join the Select Showcase auction from seller 4 Sharp Corners.

Figure 2: *eBay's homepage*

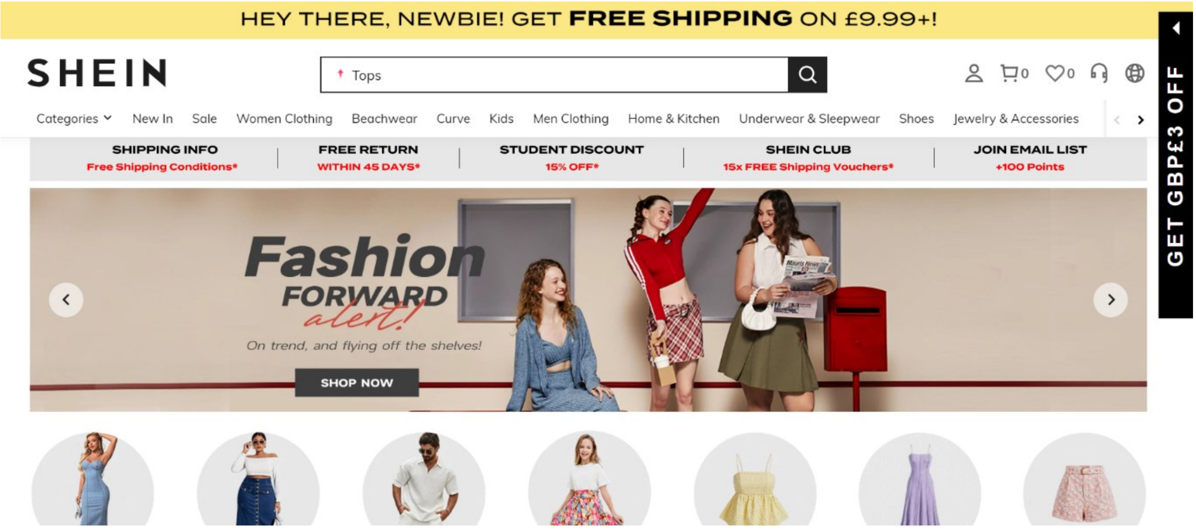


Figure 3: *Shein's homepage*

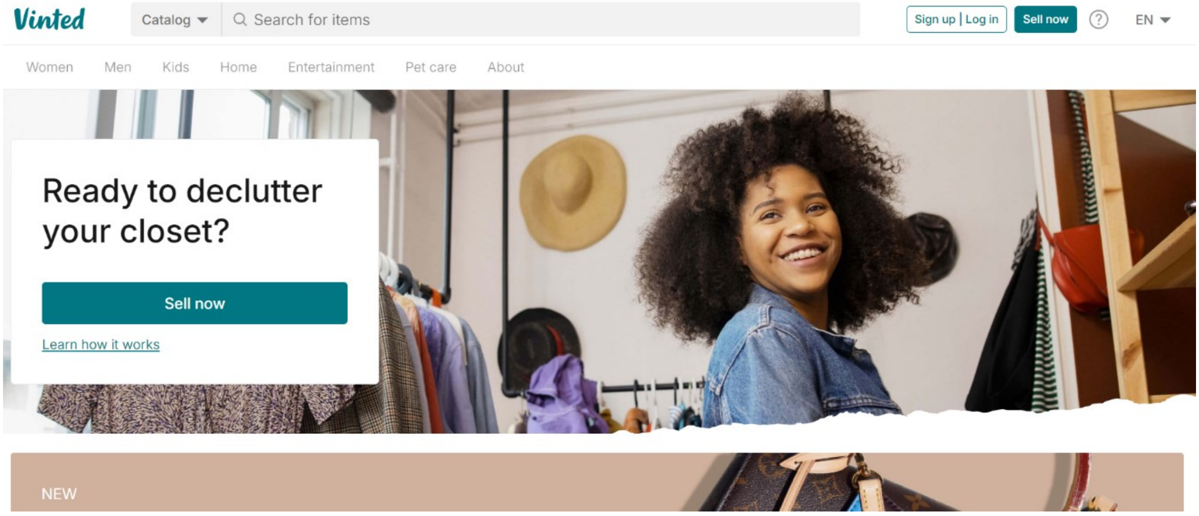


Figure 4: *Vinted's homepage*

As shown in the images, the stimuli used in this experiment consisted of static images and dynamic content from each website's homepage, incorporating visual elements such as product listings, banners, call-to-action buttons, and navigation menus. These elements provided diverse visual content to test the model's ability to predict user focus in a real-world context. The eye-tracking data was transformed into heat maps, representing areas of high user attention. The AI model's predicted saliency maps were then compared with these heatmaps to assess the model's accuracy.

The TranSalNet model processes images to predict visual saliency by following a carefully structured pipeline. First, input images are preprocessed using padding and resizing to fit the required input size of 384x288 pixels. This step ensures consistency across different image inputs.

The model architecture, which can be TranSalNet_Dense or TranSalNet_Res based on the defined flag, leverages dense connections and residual networks to extract hierarchical features critical for accurate saliency prediction. These features are then processed through the network, and the predicted saliency map is generated.

After generating the saliency maps, they are compared to ensure a one-to-one analysis with eye-tracking data. The pipeline used in this study is illustrated in Figure 1, showcasing the steps from image preprocessing to final visualisation.

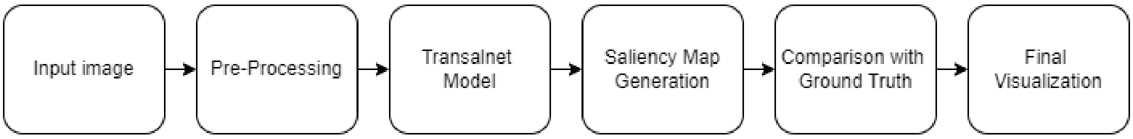


Figure 5: *TransalNet's pipeline*

To evaluate the model's performance, we employed five key metrics: Normalized Scanpath Saliency (NSS), Kullback-Leibler Divergence (KL-Div), Area Under the Curve (AUC), the Correlation Coefficient (CC) and the Similarity Index Measure (SIM).

The Normalized Scanpath Saliency (NSS) metric was particularly important in assessing the correspondence between predicted saliency maps and the actual fixations recorded in the eye-tracking data. Mathematically, NSS is calculated as:

$$NSS = \frac{1}{N}\sum_{i=1}^{N}\frac{S(x_i y_i)-\mu_S}{\sigma_S} \tag{3}$$

where $S(x_i y_i)$ represents the saliency score at the fixation point $(x_i y_i)$, μ_S is the mean saliency score across the image, and σ_S is the standard deviation of the saliency scores. This metric essentially measures how much the predicted saliency values at fixated locations deviate from the mean saliency value, thus providing a direct way to evaluate the model's accuracy in predicting human attention.

In addition, the Kullback-Leibler Divergence (KL-Div) was used to assess the similarity between the predicted saliency distribution and the distribution of fixations recorded during the eye-tracking sessions. The formula for KL-Div is:

$$D_{KL}(P||Q) = \sum_i P(i)\, log\frac{P(i)}{Q(i)} \tag{4}$$

where $P(i)$ is the probability distribution of human fixations and $Q(i)$ is the predicted saliency distribution. KL-Div measures how much the two distributions diverge from each other, with a smaller value indicating a better match between the model's predictions and the human data.

The Area Under the Curve (AUC) was employed as a general performance metric, evaluating the model's ability to discriminate between fixated and non-fixated areas of the image. A high AUC score suggests that the model accurately identifies the regions of interest within the image, closely matching the human visual attention patterns observed through the eye-tracking data.

Moreover, the Correlation Coefficient (CC) measures the linear correlation between the predicted saliency map and the observed eye-tracking data. A value of 1 indicates a perfect correlation, while a value of 0 indicates no correlation. The formula for CC is:

$$CC = \frac{\sum(P(i)-\bar{P})(Q(i)-\bar{Q})}{\sqrt{\sum(P(i)-\bar{P})^2\sum(Q(i)-\bar{Q})^2}} \tag{5}$$

where $P(i)$ and $Q(i)$ represent the predicted and observed saliency values, respectively, and $\bar{P}$ and $\bar{Q}$ are the mean saliency values of the two distributions.

The Similarity Index Measure (SIM) evaluates the similarity between the predicted and observed saliency maps. It compares the two distributions of saliency values, producing a score between 0 and 1, where 1 indicates a perfect match. The SIM is calculated using the following formula:

$$SIM = \sum \min\left(P(i), Q(i)\right) \tag{6}$$

where $P(i)$ and $Q(i)$ represent the predicted and observed saliency distributions, respectively.

4. Experimental Results

In this experiment, we aimed to explore how effectively TranSalNet, an AI model designed for saliency prediction, could replicate human visual attention patterns when applied to e-commerce websites. The central focus is to check how well the model performed in predicting areas of interest, such as product images, banners, and call-to-action buttons, which are critical elements in an online shopping experience. Using eye-tracking data as the ground truth, we compared the model's predicted saliency maps to actual user fixations, analysing performance across multiple platforms. Each e-commerce site presented different visual layouts, offering a variety of challenges for the model in identifying primary and secondary areas of user interest.

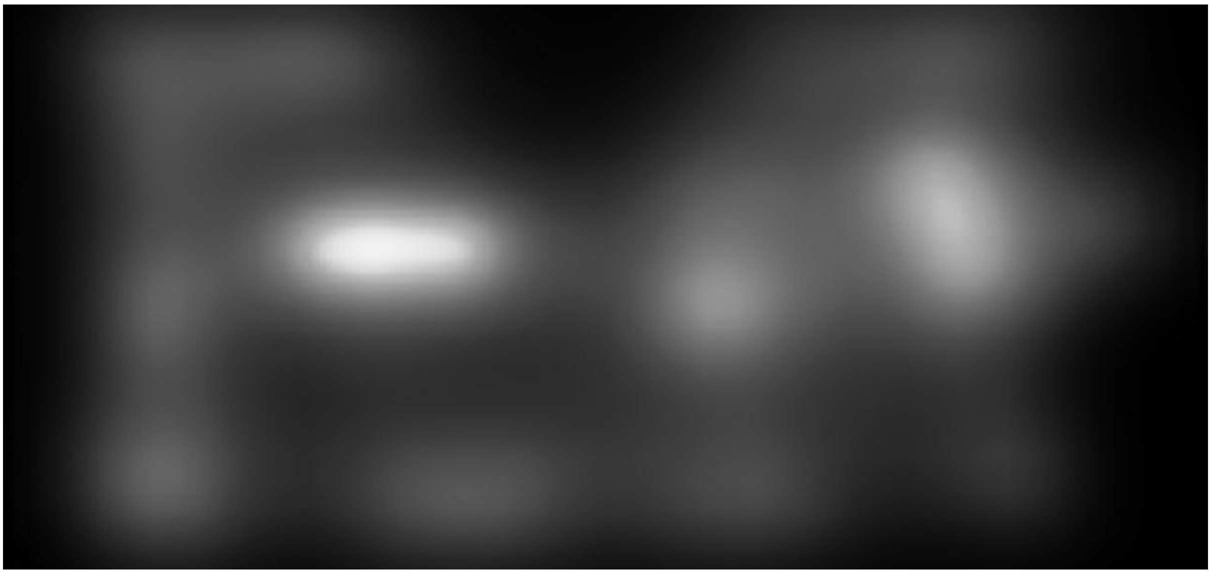


Figure 6: *Heatmap generated by the code for the Amazon homepage*

Table 1: *Amazon's Metrics*

AUC	0.79
NSS	1.30
KL-Div	0.00
CC	1.00
SIM	0.99

The Amazon platform results showed an AUC of 0.79 and an NSS score of 1.30, indicating good predictive performance. The KL-Div score was 0.00, with a perfect CC of 1.00 and a SIM score of 0.99≈1, confirming the model's reliability. The heatmap generated for Amazon highlighted product images and call-to-action buttons as the main focal points, indicating that the algorithm can capture key elements that attract users' attention. However, some less prominent areas, such as product description sections, were underestimated, suggesting the need for further optimisation for a more comprehensive detection of areas of interest. That means the saliency model's output accounts for some false negatives.

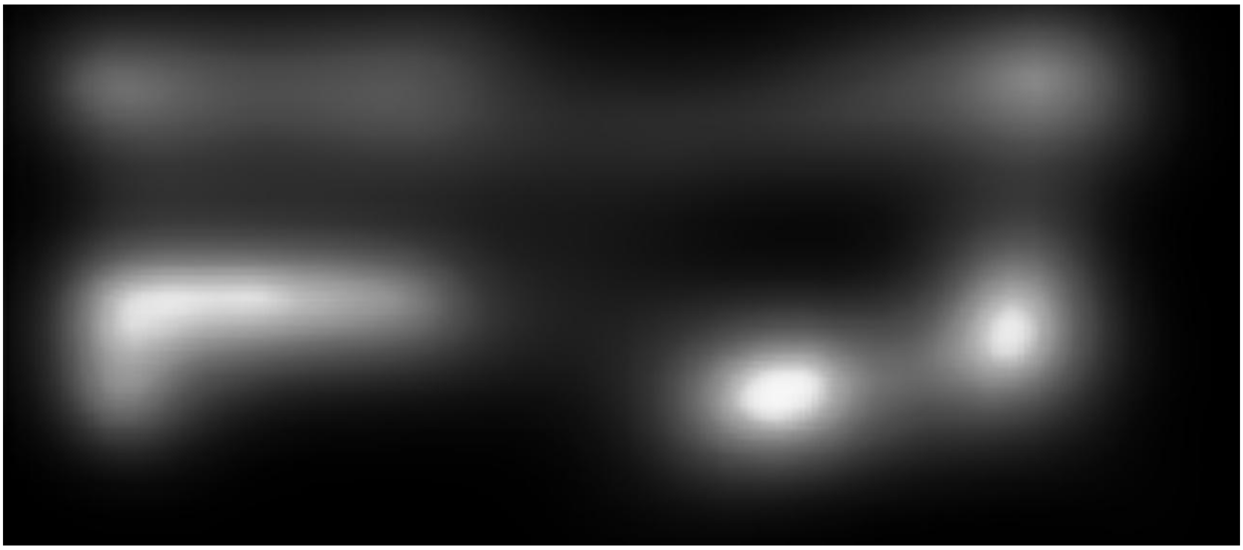


Figure 7: *Heatmap generated by the code for the eBay homepage*

Table 2: *eBay's Metrics*

AUC	0.87
NSS	1.83
KL-Div	0.00
CC	0.99
SIM	0.99

In the case of eBay, the predictive saliency model achieved an AUC rate of 0.87 and an NSS score of 1.83, demonstrating high accuracy. The KL-Div score was 0.00, with a CC of 1.00 and a SIM score of 0.99≈1, underscoring the model's effectiveness in capturing users' attention accurately. The heatmap generated for eBay showed good alignment with the central areas of the page, where the most viewed products are located. However, there are discrepancies in the peripheral regions and navigation sections, where the model's predictions do not fully match the eye-tracking data. That suggests that while the algorithm performs well for the main areas of interest, it may require improvements to capture the entire spectrum of user fixations, including secondary elements that could influence the browsing experience.

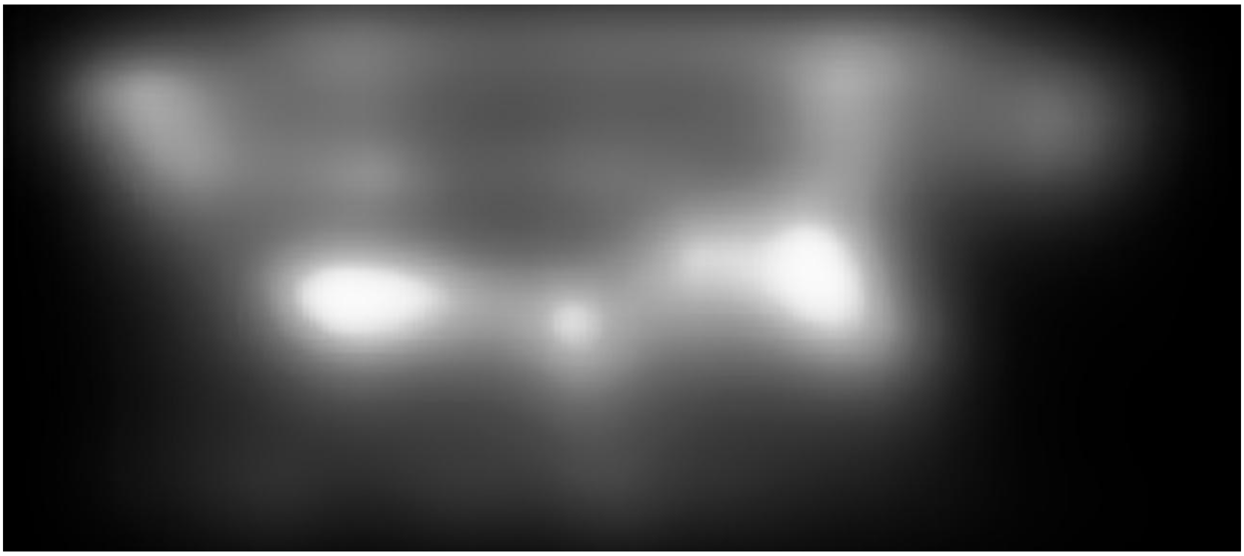


Figure 8: *Heatmap generated by the code for the Shein homepage*

Table 3: *Shein's Metrics*

AUC	0.82
NSS	1.25
KL-Div	0.00
CC	1.00
SIM	1.00

For Shein, the predictive model achieved an AUC value of 0.82, indicating a high accuracy rate in distinguishing salient from non-salient regions. The NSS score was 1.25, showing a moderate alignment between the predicted saliency and actual user attention. Similarly, the KL-Div score was 0.00, the CC was 1.00, and the SIM score was 1.00, reflecting a substantial similarity between the predictive model and the eye-tracking data. The results reflected the quality of the predictive map, which revealed significant attention to product listings and promotional images, which are crucial elements for a fashion-focused e-commerce site. However, the predictive maps did not highlight some areas, such as search filters and navigation sections. That suggests that the algorithm could benefit from further optimisation to better capture the various visual preferences of users, especially in a dynamic context like fashion.

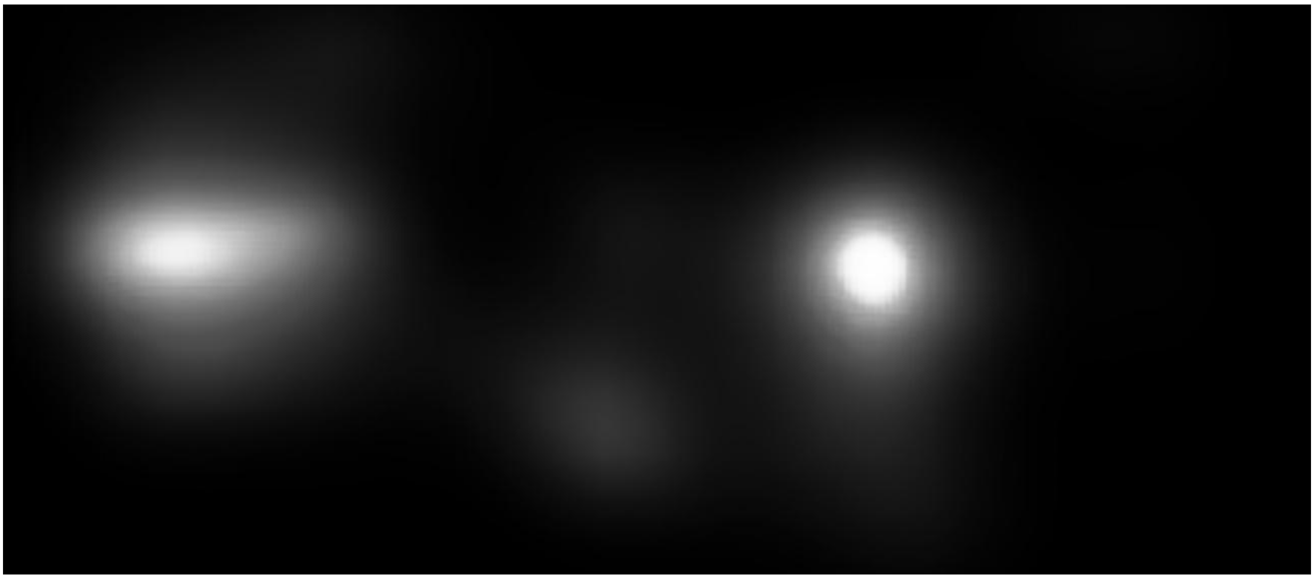


Figure 9: *Heatmap generated by the code for the Vinted homepage*

Table 4: *Vinted's Metrics*

AUC	0.95
NSS	3.36
KL-Div	0.00
CC	0.99
SIM	1.00

Finally, for Vinted, the predictive saliency maps are well-aligned with the eye-tracking data in sections where user-generated content is displayed, such as product images uploaded by users. AUC reaching 0.95 and NSS scoring 3.36 indicate a strong correlation with the eye-tracking data. The KL-Div score remained at 0.0, while the CC was 0.99≈1, and the SIM score was 1.0, further validating the model's accuracy. However, some discrepancies emerge in areas with mixed content, where the model's predictions only partially align with the user fixation data. That indicates that while the algorithm effectively predicts the main areas of interest, it may require further optimisation to better process the complexity of mixed content.

As observed, even though the algorithm used was not fine-tuned to capture the most salient parts of the sites, it still performed well by identifying the critical areas of interest. In fact, as we can see from these metrics, the high AUC and NSS scores and the perfect or near-perfect CC and SIM scores indicate that the models are highly effective in replicating actual user attention patterns. However, the moderate NSS scores for platforms like Amazon and Shein suggest that while the models perform well, there is still room for improvement in fine-tuning the predictions to better capture all nuances of user behaviour. That demonstrates the algorithm's robustness in highlighting essential elements, such as product images and call-to-action buttons, which are crucial for user engagement.

The differences in the salient areas can be attributed to the diverse nature of the e-commerce sites analysed. Each site caters to different product categories and thus employs distinct site structures and design elements. For example, fashion-focused sites like Shein emphasise product listings and promotional images. In contrast, a marketplace like eBay may have a broader focus that includes search filters and navigation tools.

These variations in site design and content emphasis naturally lead to differences in what is considered visually salient for users. Despite not being fine-tuned explicitly for each site, the algorithm's ability to adapt to these different contexts highlights its potential for general applicability across various types of e-commerce platforms. That underscores the importance of considering different product categories' unique characteristics and user behaviours when developing and optimising predictive saliency models.

Moreover, I have analysed the results of the eye-tracking experiment. Fixation points, where the gaze is held steadily for around 200-300 milliseconds (Rayner, 1998), indicate areas of interest and attention, revealing which elements on the webpage attract the user's focus. Saccades, which are rapid movements between fixation points, help identify the scanning behaviour and how users navigate the visual information. The heatmaps obtained provide a detailed representation of user fixations on various e-commerce sites, offering a more comprehensive analysis of users' visual habits.

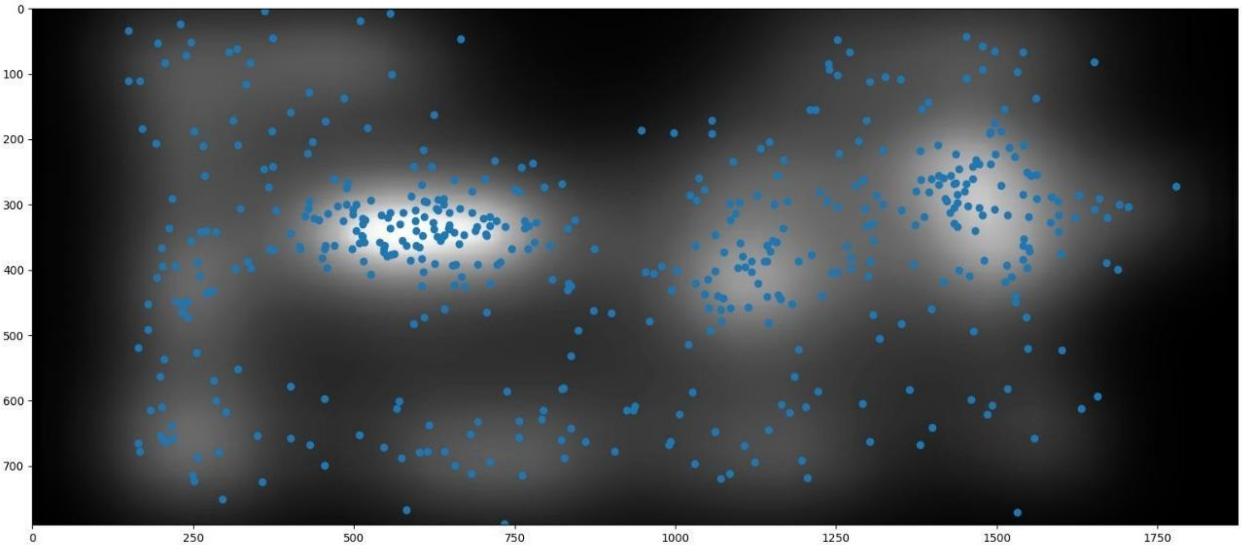


Figure 10: *Fixation points (blue circles) overlaid with saliency map for Amazon homepage*

For Amazon, the eye-tracking data showed a substantial concentration of fixations on product images and call-to-action buttons, confirming the importance of these elements in capturing users' attention. However, the heatmaps also revealed significant interest in product description sections and user reviews, which were not as well highlighted by the predictive saliency maps.

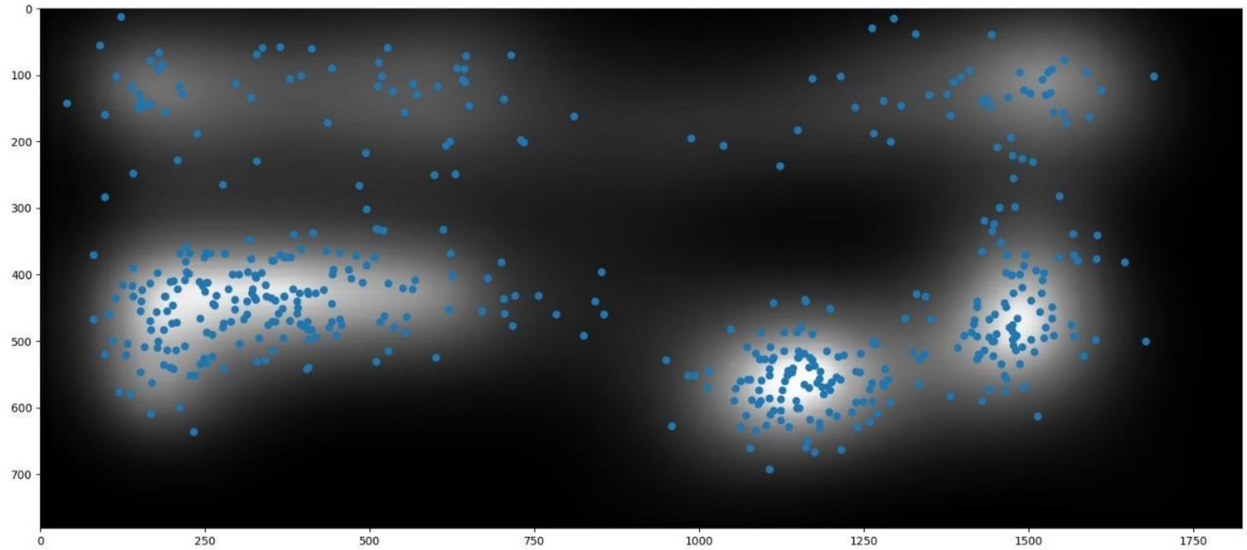


Figure 11: *Fixation points (blue circles) overlaid with saliency map for eBay homepage*

In the case of eBay, user fixations spread in the central areas with the most viewed products but with more significant dispersion in the peripheral regions compared to the model's predictions. That suggests that users also explore navigation sections and search filters, which are crucial for the shopping experience but are not always captured by the predictive maps. The eye-tracking heatmaps thus indicate a more complex and distributed exploration behaviour, highlighting the importance of optimising all sections of the page to improve overall user interaction.

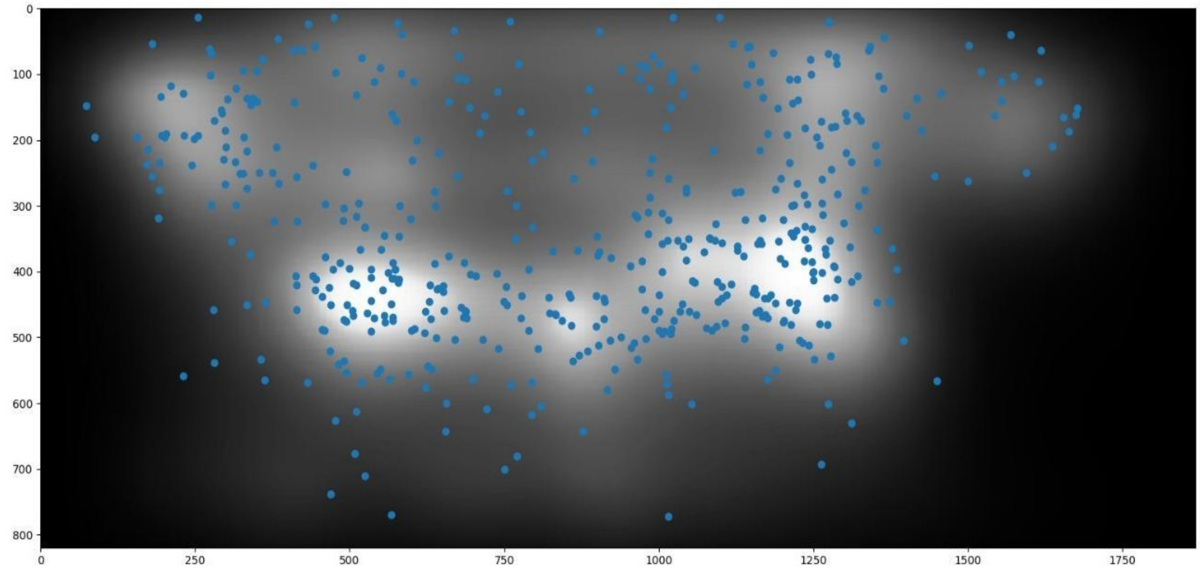


Figure 12: *Fixation points (blue circles) overlaid with the saliency map for the Shein homepage*

Regarding Shein, the eye-tracking data revealed greater attention to search filters, navigation sections, and product listings. That suggests that users find these areas equally important, an aspect that the predictive saliency code did not fully capture. The heatmaps indicate that, besides products, users frequently interact with search and filter tools to customise their shopping experience. Future method optimisations should tackle the abovementioned aspects.

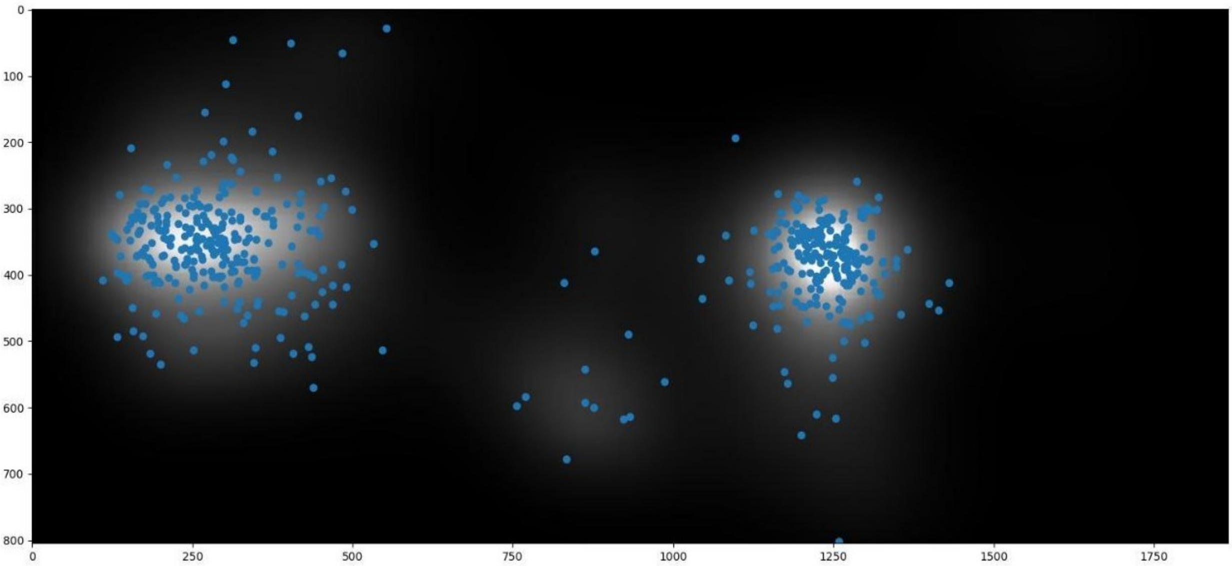


Figure 13: *Fixation points (blue circles) overlaid with saliency map for Vinted homepage*

Finally, for Vinted, the eye-tracking heatmaps aligned strongly with the predictive maps in sections with user-generated content, such as product images. However, the heatmaps also highlighted significant interest in navigation sections and product details. Those areas that were not highlighted as well as by the predictive saliency maps. That suggests that while the algorithm effectively predicts the main areas of interest, it must be refined to capture the full spectrum of users' visual interactions with higher accuracy.

The eye-tracking heatmaps provide invaluable insights into user fixations across various e-commerce platforms, revealing a nuanced picture of users' visual habits. They highlight not only the primary areas of interest, such as product images and call-to-action buttons, but also secondary elements, like product descriptions, user reviews, and navigation tools, that are equally crucial for a comprehensive user experience. While the predictive saliency maps performed well in identifying critical focal points, the eye-tracking data uncovered additional areas of interest that the algorithm did not fully capture. This discrepancy underscores the importance of integrating eye-tracking data into the development and refinement of predictive models. By doing so, we can ensure a more holistic understanding of user behaviour and optimise e-commerce sites to enhance user interaction and satisfaction across different product categories and site structures.

5. Conclusions

This research explored the integration of artificial intelligence (AI) into the field of visual saliency modelling, specifically within the context of e-commerce. The primary objective was to enhance the understanding of how AI models can replicate human attention mechanisms, thereby improving user interaction and engagement with online platforms. Visual saliency plays a crucial role in identifying areas of interest in visual scenes, both in human vision and AI systems. This study focused on assessing how AI-driven models, like TranSalNet, could predict user attention and saliency across various e-commerce websites.

The findings revealed that the model effectively identified central areas of visual interest, particularly product images and promotional content. However, some discrepancies were observed in peripheral elements, such as navigation tools and search filters. That suggests that while AI-driven models can replicate core areas of user focus, further improvements are needed to capture more subtle or peripheral elements equally important for user interaction. Including top-down processes, which integrate user intent and cognitive context, could significantly enhance the predictive power of these models.

Another key insight from the study was the role of personalisation in saliency modelling. Individual user preferences, browsing behaviour, and cultural differences contribute to variations in visual attention patterns. AI models that incorporate real-time user data and adaptive mechanisms have the potential to provide personalised predictions, ensuring that websites are optimised for different user groups. Future work in this area could focus on developing models that dynamically adjust to specific user profiles, enhancing the overall user experience.

Moreover, as e-commerce platforms continue to integrate dynamic and interactive content, it becomes essential for saliency models to handle both spatial and temporal dynamics. The results suggest that optimised for static images, the current models could benefit from advancements in handling time-based interactions, such as scrolling, animations, and videos. Future research should prioritise developing models capable of processing these temporal aspects to provide a more comprehensive understanding of user attention.

In conclusion, this research highlights the potential of AI models like TranSalNet in optimising the design of e-commerce platforms by predicting user attention patterns. While significant progress has been made, there remains room for improvement, particularly in terms of personalisation, peripheral element detection, and the incorporation of temporal dynamics. The continued advancement of AI-driven saliency models will improve user experiences and provide valuable insights for enhancing web design and user interface optimisation across various industries.

References

[1] Russell, S.J. and Norvig, P., 2016. Artificial intelligence: a modern approach. Pearson.

[2] Li, J. and Gao, W. eds., 2014. Visual saliency computation: A machine learning perspective (Vol. 8408). Springer

[3] Itti, L., Koch, C. and Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, pp.1254-1259.

[4] Tatler, B.W., Baddeley, R.J. and Gilchrist, I.D., 2005. Visual correlates of fixation selection: effects of scale and time. *Vision Research*, 45(5), pp.643-659.

[5] LJOVO TranSalNet GitHub repository. Available: https://github.com/LJOVO/TranSalNet. Accessed: Sep. 17, 2024.

[6] Fang, C., Tian, H., Zhang, D., Zhang, Q., Han, J. and Han, J., 2022. Densely nested top-down flows for salient object detection. Science China Information Sciences, 65(8), p.182103.

[7] Hou, X. and Zhang, L., 2007. Saliency detection: A spectral residual approach. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8.

[8] Harel, J., Koch, C. and Perona, P., 2007. Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 19, pp.545-552.

[9] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444

[10] Tliba, M., MA, K., Ghariba, B., Chetouani, A., Çöltekin, A., MS, S. and Bruno, A., 2022. SATSal: A Multi-Level Self-Attention Based Architecture for Visual Saliency Prediction. IEEE ACCESS, 10, pp.20701-20713.

[11] Liu, X., Huang, G., Yuan, X. et al. Weakly supervised semantic segmentation via saliency perception with uncertainty-guided noise suppression. Vis Comput (2024). https://doi.org/10.1007/s00371-024-03574-1

[12] Zhong, R., Xiao, D., Dong, S. and Hu, M., 2021. Spatial attention model-modulated bi-directional long short-term memory for unsupervised video summarisation. Electronics Letters, 57(6), pp.252-254.

[13] Bruno, A., Gugliuzza, F., Ardizzone, E., Giunta, C.C. and Pirrone, R., 2019. Image content enhancement through salient regions segmentation for people with color vision deficiencies. i-Perception, 10(3), p.2041669519841073.

[14] Wang, C., Niu, L., Zhang, B. and Zhang, L., 2023. Image Cropping With Spatial-Aware Feature and Rank Consistency. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10052-10061).