Original scientific paper http://doi.org/10.24867/JGED-2023-1-029

Using artificial intelligence for predictive eye-tracking analysis to evaluate photographs

ABSTRACT

The goal of the study was to determine how close the eye-tracking results predicted by the AI model are to actual measurements and whether they can be used in scientific research or in real business cases. The study was based on a carefully prepared photo database of 30 photos of varying complexity and colour. The photos were shown to 110 participants (age and gender evenly distributed), and eye-tracking device (Tobii X120) was used to measure how the photos were viewed. In comparison, the same photos were tested using an AI-based application (Expoze.io). The final results show the comparison between the heatmaps and transparent gaze visualisations of the collected data with the two used measurement methods. Suggestions are made in which cases and how the two described methods should be used.

KEY WORDS

Artificial intelligence, communication value, photography evaluation, predictive eye-tracking, SSIM

Jure Ahtik

University of Ljubljana,
Faculty of Natural Sciences and Engineering, Ljubljana, Slovenia

Corresponding author:
Jure Ahtik
e-mail: jure.ahtik@ntf.uni-lj.si

First received: 8.7.2022.
Revised: 26.8.2022.
Accepted: 30.8.2022.

Introduction

When a photograph is going to be published, three different aspects are considered: the technical, the artistic and the content/meaning. The technical aspect is usually not a problem nowadays. Digital cameras are now so advanced that an experienced photographer can be confident that exposure, sharpness, colour reproduction etc. are almost always correct, even under the most difficult conditions. The second aspect is an understanding of artistic variables, such as composition, which usually comes with experience. Both the technical and artistic aspects should be at a high level of quality when it comes to professional photography that can be considered for publication. The final decision on what to publish is based in most cases on the third aspect – content/meaning.

Photography is mainly used for communication purposes and has become one of the main sources of information with more than 54,000 photos taken each second (Photutorial, 2022). An author or editor wants to transfer the information to the end user, and confidence if a photograph can do this is critical. Deciding which photo will best convey the message is the biggest part of the decision-making process and predicting it can sometimes be difficult. It is important to understand which elements of the photo the user spends the most time on when viewing it. If the user is distracted by secondary elements, the main message may not be transferred. In the past, many different subjective and objective methods have been used to determine the communication value of photographs, one of which is eye-tracking (Ahtik & Starešinič, 2017; Molina et al., 2018; Mañas-Viniegra, Veloso & Cuesta, 2019; Mitu & Bota, 2021; Loceye, 2022). In order to conduct such a study, many resources are needed: the main difficulties are the number of people we need to test in order to get a representative result and the time needed for the measurements. The procedure can be also expensive. In scientific research, all this is not a problem in most cases. But if you are trying to conduct such a study in an advertising agency or a graphic design studio, for example, the available resources are in most cases unfavourable. The industry needs the results at the moment when new information/

products are created, and the use of reliable eye-tracking measurements is not possible in most cases.

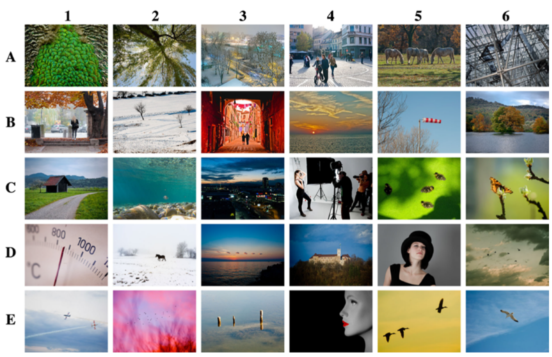
In recent years, artificial intelligence (AI) has proven to be useful in many fields, and photography is no exception (Kong et al., 2022). There are well-known and commonly used cases of computational photography and image recognition and we can even see some algorithms capable of generating a photographic-looking images based on just a few keywords (OpenAI, 2022; Dechterenko & Lukavsky, 2016). There are also new approaches to analysing the perception of photos based on content, composition, and appearance (Müller, Kappas & Olk, 2012). To speed up the photo evaluation process and make it more available, machine learning methods based on neural networks trained by a large number of eye-tracking measurements have been introduced. We call this method *predictive eye-tracking* (Expoze.io, 2022).

The aim of the study was to determine whether traditional eye-tracking measurements could be replaced by predictive eye-tracking for assessing the communication value and visual attention of photographs. The study involved selecting an image database, conducting eye-tracking measurements, performing predictive eye-tracking calculations, comparing data collected by both methods, and final data analysis.

Materials and methods

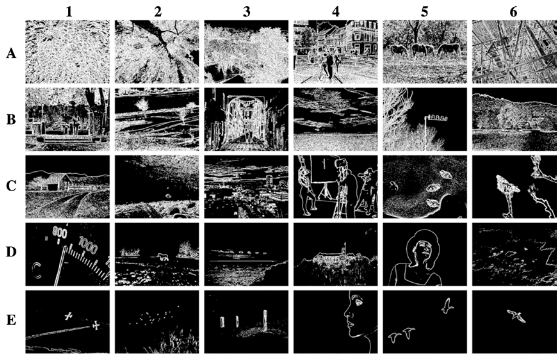
Materials

The study was conducted using the novel image database first introduced in 2017 (Ahtik, Muck & Starešinič, 2017; Ahtik & Starešinič, 2017). The database consists of 30 reference photographs (Figure 1), which differ mainly in content complexity and colour variety. In addition, the photo database also contains 300 different modified images (10 per reference) for image quality assessment, which were not used in this study. Photographs were carefully selected for the novel image database to examine the impact of image complexity on the way we perceive them (see Figure 2 for details overview).



» **Figure 1:** *Image database*

Photos are ranked in order of highest complexity (Figure 2: A1) to lowest complexity (Figure 2: E6). The range of overall detail coverage (complexity), was calculated as the average pixel value of the images with exposed edges as shown in Figure 2, is from 99 % (Figure 2: A1) to 22 % (Figure 2: E6), with an average of 77 %. This is significantly better than the image database most commonly used in other research – TID2008 (Ponomarenko et al., 2009).



» **Figure 2:** *Image database with exposed detail complexity.*

Eye-tracking

A Tobii X120 eye tracker, a HP ZR24W LCD monitor, a PC, a controlled dark room environment, and Tobii Studio 3.4.4 software were used to perform the eye-tracking measurements. The photographs were displayed in the centre of a black screen with a fixed resolution of 840 × 630 px, and each observer had the same sitting position and viewing angle of the photographs (Figure 3). Each photograph was displayed for 5 seconds, followed by 2 seconds of dark screen. The entire test lasted around 4 minutes. The only instruction to observers was to observe, after which no questions were asked. There were 110 participants in total, 50 % female and 50 % male, 50 % under 30 years of age and 50 % over 30 years of age. The average age of all participants was 33.39 years. All participants were from Slovenia and had normal or corrected-to-normal vision. The final results were exported as an average of all observers in the form of heat maps and transparent gaze visualisations.

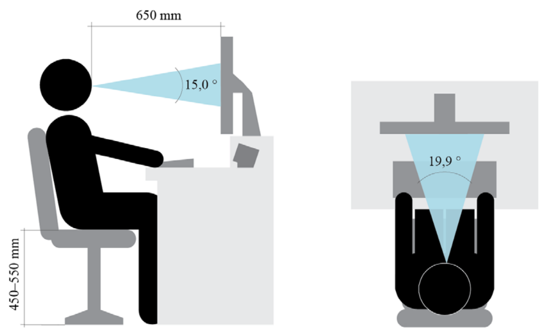


Figure 3: *The eye-tracking measuring condition.*

Predictive eye-tracking

Predicting what and how observers would look was done using an artificial intelligence application called Expoze.io. This is a subscription-based online application that allows the analysis of static images and video recordings. The attention prediction generated by Expoze.io is created using a generative adversarial network (GAN). The input are the RGB values of image or video and the output is the attention prediction. On the input side, a ConvNet (Convolutional Neural Network) is used, which has been trained to recognise objects with over 14 million images (The data science blog, 2016; Expoze.io, 2022). The neural network was than trained on the attention data of thousands of participants that looked at more than 10,000 images (Expoze.io, 2022).

The application has been tested on various graphic design examples, user interfaces, packaging, and commercials, and according to a study (Expoze.io, 2022), it offers 95 % accuracy compared to eye-tracking. This probability provides a high enough confidence factor for the method to be taken seriously. The method has not yet been tested for analysing communication value on a photographs before. We used the same photographs that were used for the eye-tracking measurements. The settings for generating the final heat maps and the gaze opacity images for export were standard.

Analysis and data evaluation

The final data were exported from both used applications (Tobii Studio 3.4.4 and Expoze.io) in similar and comparable ways. There are three common ways to display eye-tracker data in general: numerically, as a heatmap, and as a gaze plots. Tobii Studio offers all possible exports with many additional options, while Expoze.io only offers export of heatmaps (*Normal* setting) and a transparent view on gaze plots (*Reveal* setting), where only the parts of the photos that were viewed most are visible. Expoze.io doesn't offer numerical export, so the evaluation had to be done based on visualised data exports.

The analysis of the measured and calculated data was then performed in two ways:

a. the data presented in the form of heat maps was evaluated by subjective visual assessment,
b. the data presented in the form of transparent gaze visualisations was evaluated with an objective assessment using the structural similarity (SSIM) index, which has been-shown in previous studies to be a suitable method for comparing two similar images (Tong, 2005; Ahtik, Muck & Starešinič, 2017). Expoze.io also did their similarity studies using the same metric (Expoze.io, 2022).

The SSIM index is calculated on various windows of an image. Equation (1) shows the measure between two windows *x* and *y* of common size N×N, where μ_x is the average of *x*, μ_y is the average of y, σ_x^2 is the variance of *x*, σ_y^2 the variance of *y*, σ_{xy} is the covariance of *x* and *y*, $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ are two variables to stabilize the division with the weak denominator, and *L* the dynamic range of pixel-values, $k_1 = 0.01$ and $k_2 = 0.03$ by default.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{1}$$

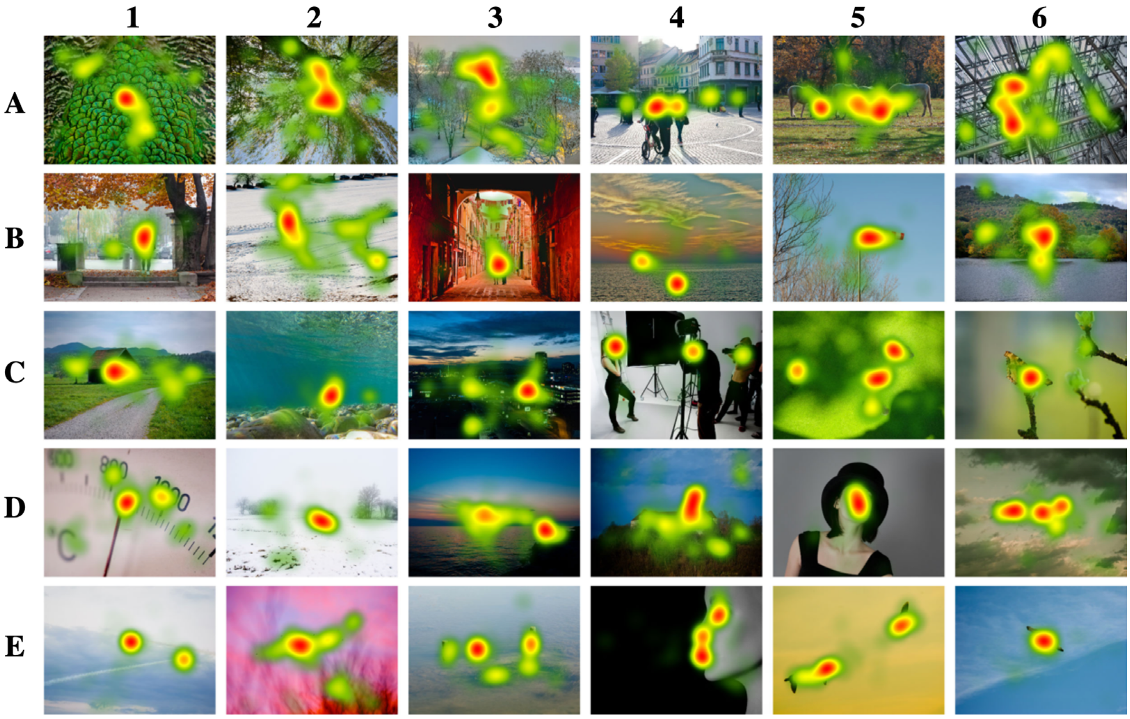
Results and discussion

Heat map visualisations

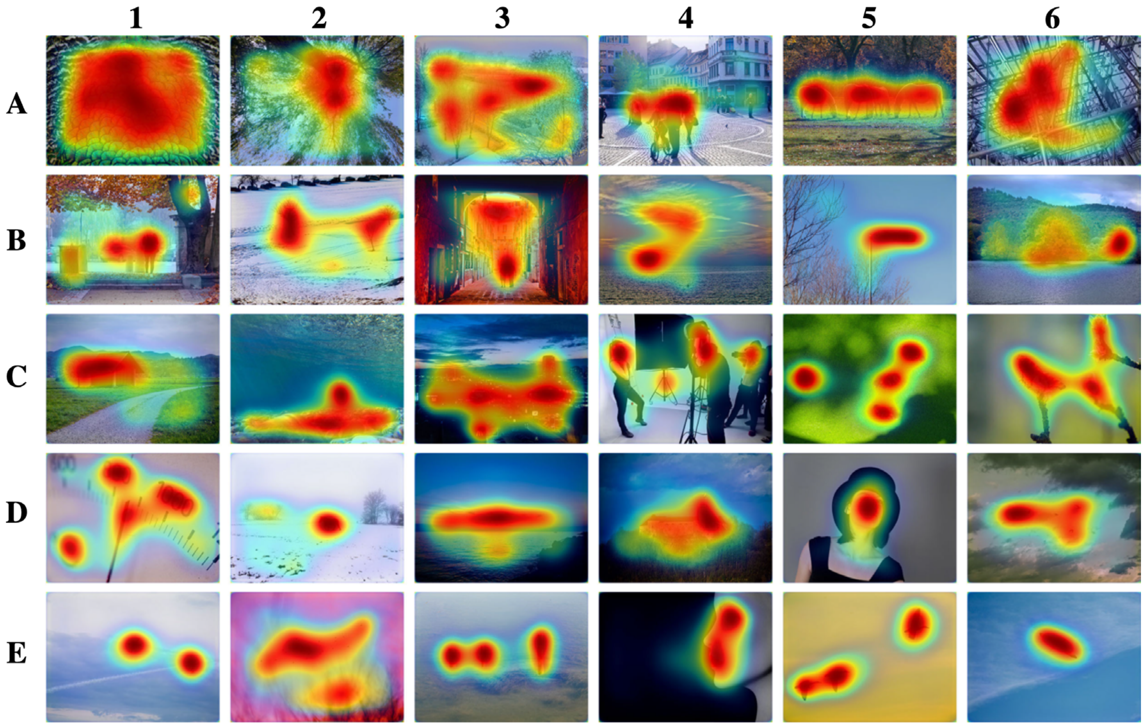
The results of the eye-tracking measurement are shown in Figure 4, and the results of the predictive eye-tracking calculations are shown in Figure 5. In both cases, the results are presented as visualised heat maps of the areas viewed and the areas predicted to be viewed. The exported visualisations are similar and can be compared visually, as this type of visualisation is intended for visual evaluation. Red areas represent the elements that are seen the most, and green/blue areas the least. Areas that are not coloured in the heat map were not viewed or are not predicted to be viewed by the observers.

When comparing the visualisation of the two methods, the main interest is whether and how the main areas of interest were seen and observed. Some photos in the image database have clearly visible elements that stand out from the background (e.g., A4, A5, B1, B5, C5, D2, D5, E1, E3-6 ...), and some photos have less clearly visible elements (e.g., A1-3, A6, B4, C2, D3 ...). The visual comparison shows that both methods correlate most with photos that have clearer areas of interest, while there is less or no correlation for some other examples where there are fewer elements on which to fixate the eyes. Since predictive eye-tracking separates clear elements from the background and predicts that the human eye will fixate on these elements the most, the result is predictable. For the photos with less recognisable elements, especially example A1, the prediction is unclear and has no real relation to the real measurements.

The other conclusion that can be drawn is that prediction is more successful with less complex photographs (Figure 2). Less detail means that the machine learning algorithm can more easily distinguish between elements and background. Since Expoze.io was originally developed for evaluating graphic designs, where the elements are usually clearly visible – which is the main function of the design process itself – the results are also not surprising.



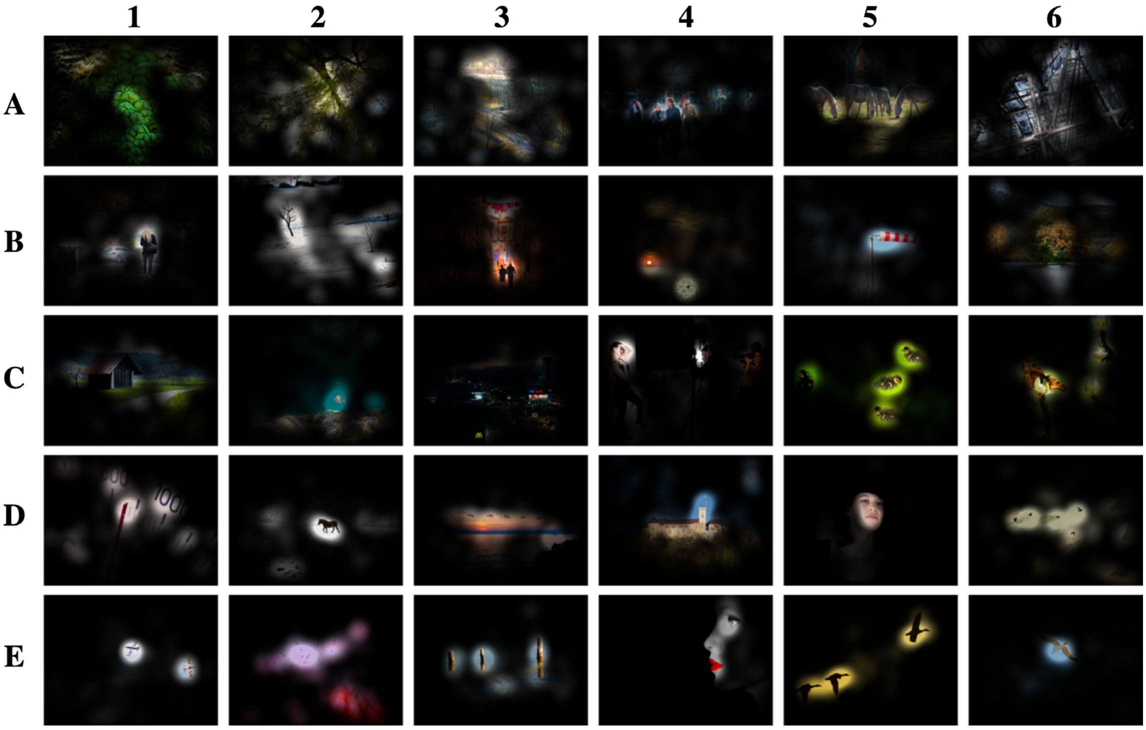
» **Figure 4:** *Data measure by the eye tracker Tobii X120 presented as heat maps*



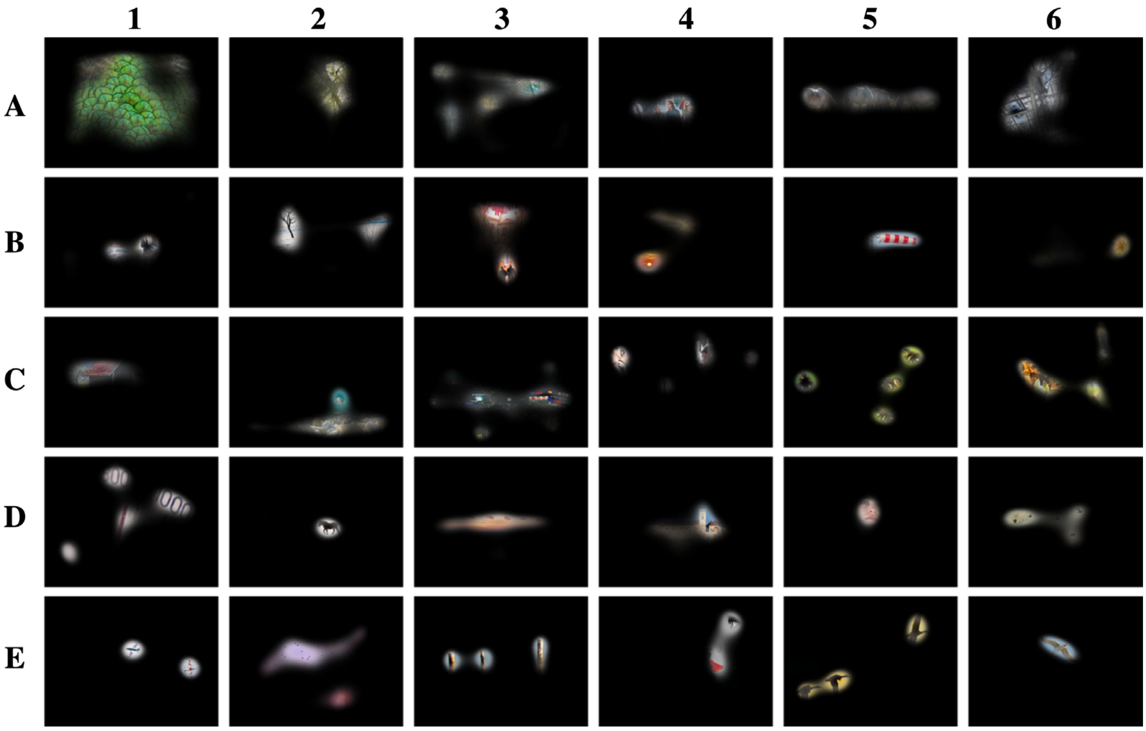
» **Figure 5:** *Data calculated by predictive eye-tracking application Expoze.io presented as heat maps*

Transparent gaze visualisations

Unlike heat maps, where the exported visualisations from the two applications used are not fully comparable, there is a better way to do this when analysing transparent gaze visualisations. Measured transparent gaze visualisations are shown in Figure 6 and predicted transparent gaze visualisations in Figure 7. In the visual evaluation of the visualisations of both methods, the similar conclusions can be drawn as for the heat maps. But the greater similarity of the visualisations gives us the opportunity to make more objective calculations. The SSIM index gives us an objective result of how structurally similar the images are to each other. We compared predictive eye-tracking visualisations to eye-tracking, and a higher result means better similarity or in this case a better pre-



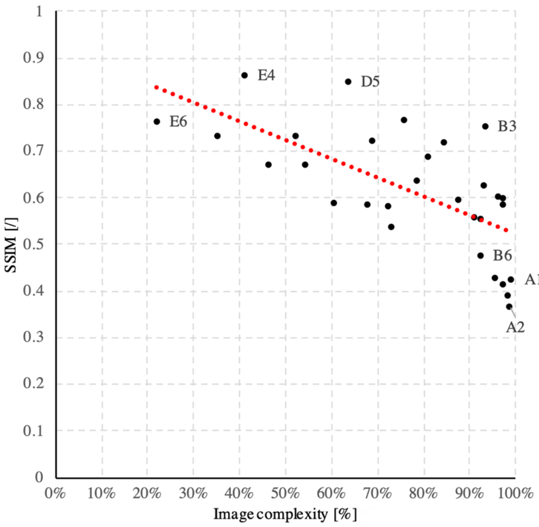
» **Figure 6:** *Data measured by the eye-tracker Tobii X120 presented as transparent gaze*



» **Figure 7:** *Data calculated by predictive eye-tracking application Expoze.io presented as transparent gaze*

diction. The calculations are consistent with the results of visual evaluation, where dependence on image complexity has also shown that prediction is better when photos are less complex. As shown in Figure 8, in some cases, e.g. E4 or D5, the SSIM index is very high, almost 0.9, while in other cases, e.g. A1 or A2, it is much lower, even below 0.4. The trend is clearly in favour of less complex and is steeply dropping towards more complex images.

are less complex. As shown in Figure 8, in some cases, e.g. E4 or D5, the SSIM index is very high, almost 0.9, while in other cases, e.g. A1 or A2, it is much lower, even below 0.4. The trend is clearly in favour of less complex and is steeply dropping towards more complex images.



» **Figure 8:** *SSIM index calculated between transparent gaze plot images of both used methods compared to image complexity of tested images.*

Conclusions

The study showed that predictive eye-tracking technology based on machine learning with neural networks can be used to predict how users will view a photo when the content of the photo is distinguishable from its background or the structure of the photo is not very complex. Considering that the image database used for the research contains some examples of highly complex photographs that are not usually used as main communication elements in advertising campaigns or graphic designs, the conclusion may be less rigorous. In any case, when working with complex or abstract photographs, we cannot rely on an accurate prediction.

The recommendation that can be made based on the following study is that predictive eye tracking is not recommended for scientific purposes because it is not suitable for thorough image evaluation. However, for real-world business cases where a quick result is needed, the technology provides useful and reliable results. We expect the technology to evolve, with respect to some other use cases of machine learning technologies that are currently available to a wider audience. Accordingly, in the future, creators will increasingly become curators of content generated or evaluated by artificial intelligence, which will improve the quality of visual communication.

Acknowlegment

This work was supported by the Slovenian Research Agency (Infrastructural Centre RIC UL-NTF).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Ahtik, J., Muck, D. & Starešinič, M. (2017) Detail diversity analysis of novel visual database for digital image evaluation. *Acta Polytechnica Hungarica*. 14 (6), 115-132.

Ahtik, J. & Starešinič, M. (2017) Eye movement analysis of image quality parameters compared to subjective image quality assessment. *Technical Gazette*. 24 (6), 1833-1839. Available from: doi: 10.17559/TV-20161213185321

Expoze.io (2022) *Predict attention using AI – Our platform explained*. Available from: https://www.expoze.io/uploads/90/69/17320a49c45e49baf-cb2539b001895bd.pdf [Accessed 30th August 2022]

Dechterenko, F. & Lukavsky, J. (2016) Predicting eye movements in multiple object tracking using neural networks. In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, ETRA '16, 14-17 March 2016, Charleston, South Carolina*. New York, Association for Computing Machinery. pp. 271-274. Available from: doi: 10.1145/2857491.2857502

Kong, P., Mancas, M., Gosselin, B. & Po, K. (2022) DeepRare: Generic Unsupervised Visual Attention Models. *Electronics*. 11 (11), 1-36. Available from: doi: 10.48550/arXiv.2109.11439

Loceye (2022) *How it works*. Available from: https://www.loceye.io/how/ [Accessed 4th July 2022]

Mañas-Viniegra, L., Veloso, A. I. & Cuesta, U. (2019) Fashion promotion on Instagram with eye tracking: curvy girl influencers versus fashion brands in Spain and Portugal. *Sustainability*. 11 (14), 1-18.

Mitu, F. G. & Bota, M. (2021) Consumers'perception: Discovering The" Law of Attraction" In Art Using Eye-Tracking. *Studia Universitatis Babes-Bolyai, Negotia. 66* (4), 19-44. Available from: doi: 10.24193/subbnegotia.2021.4.02

Molina, A. I., Navarro, Ó., Ortega, M. & Lacruz, M. (2018) Evaluating multimedia learning materials in primary education using eye tracking. *Computer Standards & Interfaces*. 59, 45-60. Available from: doi: 10.1016/j.csi.2018.02.004

Müller, M. G., Kappas, A. & Olk, B. (2012) Perceiving press photography: a new integrative model, combining iconology with psychophysiological and eye-tracking methods. *Visual Communication*. 11 (3), 307-328. Available from: doi: 10.1177/1470357212446410

OpenAI (2022) DALL·E 2. Available from: https://openai.com/dall-e-2/ [Accessed 7th July 2022]

Photutorial (2022) Number of Photos (2022): *Statistics, Facts, & Forecasts*. Available from: https://photutorial.com/photos-statistics/ [Accessed 7th July 2022]

Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M. & Battisti, F. (2009) TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*. 10 (4), 30-45.

The data science blog (2016) *An Intuitive Explanation of Convolutional Neural Networks*. Available from: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/ [Accessed 7th July 2022]

Tong, W. (2005) An evaluation of digital image correlation criteria for strain mapping applications. *Strain*. 41 (4), 167-175. Available from: doi: 10.1111/j.1475-1305.2005.00227.x.