Machine Learning for Basic Visual Research in Graphic Design

Jan-Henning Raff

Department of Design, HMKW University of Applied Sciences for Media, Communication and Management, Berlin, Germany

* Corresponding author's email: j.raff@hmkw.de

doi: https://doi.org/10.21467/proceedings.168.26

ABSTRACT

This paper explores the intersection of machine learning and graphic design, aiming to enhance visual analysis methodologies through the integration of domain-specific knowledge. A critical examination of existing machine learning approaches for visual analysis reveals their limitations and the need to integrate more design specific knowledge. The paper proposes two approaches to analyze spatial aspects of graphic design. The application of the proposed methods demonstrates the potential of machine learning to reconstruct the intuition of graphic designers and to automate visual analysis tasks. The relevance of this study lies in its contribution to both academic research and practical applications in graphic design. By bridging the gap between computational methods and design theory, the study offers new perspectives on visual communication and provides tools for designers and researchers alike. Moving forward, interdisciplinary collaborations between machine learning experts and graphic designers will be essential to refine methodologies and unlock the full potential of machine learning in visual communication.

Keywords: machine learning, graphic design, embodied vision

1 Introduction

Recent machine learning technology has brought advances in fields that affect the discipline of graphic design, most importantly with image recognition, and image generation. The technology is built on an assumed parallelism between computer models and the human brain, for example, it is said to include a "neural network". Vision related models, such as convolutional neural networks are based on insights from neuroscience about early visual processing. If such parallelism is the case, then practitioners and researchers might profit from machine learning to learn about visual communication. What does a machine "learn" about images? Is the "intelligence" comparable to the intuition of designers? How can the technology be used as an analytical tool for research and practice? The appeal of machine learning for visual analysis also lies in its capability to automate analysis; it is possible to process thousands of graphic designs as digital images. There are several aspects of analysis and synthesis of graphic design that may profit from machine learning [1], here I want to take up just one of those aspects. It is the emotional evaluation of a graphic design which again can be subdivided into several aspects. One of those is the evaluation of a graphic design by the eye in conjunction with the body where vision is supported by the proprioception of muscles, tendons, bones. Such embodied vision is important to evaluate the spatial aspects, the composition of a graphic design. The embodiment of vision has been explored since long by several researchers, notably Gibson [2]; its relevance for graphic design has been discussed e.g., by Kepes [3].

2 Limitations of computer vision for visual analysis

Before turning to existing machine learning approaches for visual analysis, I want to discuss some limitations of computer vision for visual analysis. One key technology of computer vision is the convolutional neural network (CNN). A CNN takes a digital image as input and decomposes it into several

Proceedings DOI: 10.21467/proceedings.168; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-970666-6-5

feature maps by using so-called kernels or filters. Trained with enough labelled data a CNN can be used for image recognition. The idea of filters comes from neuroscience that discovered orientation sensitive receptive fields in the visual cortex of cats [4]. Today, CNNs are superseded by so-called Transformers [5] which are a more flexible technology. However, at the early stages of processing, Transformers seem to operate like CNNs using oriented edge detectors [6]. Existing models, based on CNNs or Transformers are very good at object recognition as they have been trained on large datasets like ImageNet [7]. But because of their specific training with real-world examples these models fail to recognize abstract shapes that would be part of a graphic design, instead they enforce a real-world interpretation (Figure 1).

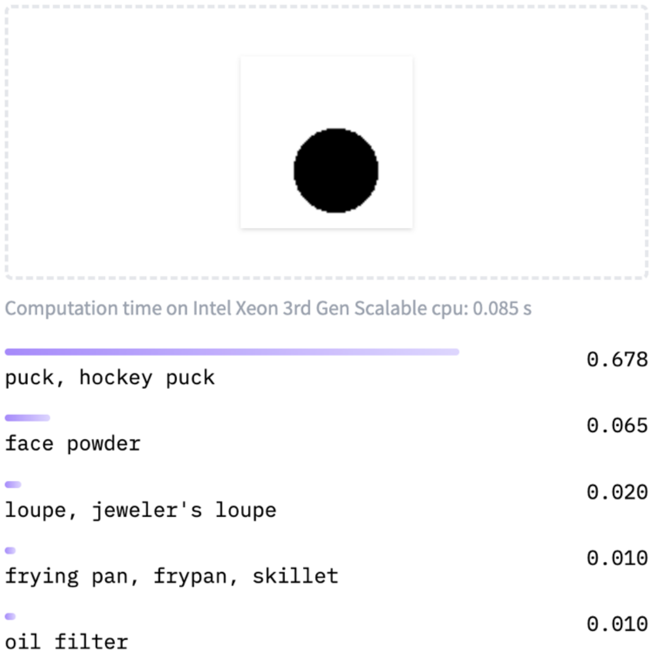

Figure 1: *A CNN forcibly predicts a black disc as a real-world object [8].*

Such forced interpretation can of course be tamed by training the models accordingly with new data. Still, such image recognition models seem too specialized for basic phenomena that are of interest in graphic design where the expressiveness of abstract shapes is often taken advantage of. The features that are found by the network in between are of greater interest. This calls for a look into the "hidden" layers of a neural network that can reveal more basic visual information. Research has also pointed out an important difference between human and computer vision. Human visual processing works with perceptual grouping as described by Gestalt psychologists [9], where the "[t]he whole is *different* from the sum of its parts" [10]. However, with CNNs, no such grouping happens; a square will also be identified with just its parts detected (Figure 2).

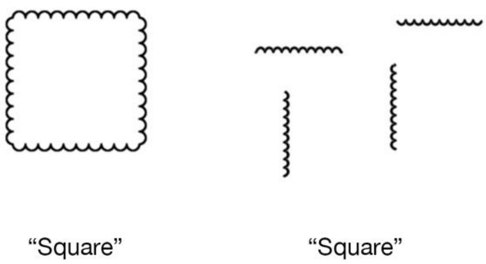

Figure 2: *A CNN predicts both configurations as squares [11].*

"[U]nlike human vision, CNNs do not compute representations of object geometry." [12]. This striking difference to human vision is less accentuated with Transformers that expose some kind of perceptual grouping [13]. Anyway, it calls for caution comparing computer to human vision. Apparently, as of today, only the early stages of vison are similar.

3 Existing machine learning approaches for visual analysis

I want to briefly highlight two main approaches here to characterize the research in the field. This is not an exhaustive overview of research that is going on, I rather want to demonstrate by example two main machine learning approaches to analyze images and discuss their advantages and limitations. The first approach (3.1) is working with engineered features, thus a sensible choice of calculations that are assumed to capture specific properties of an image. The second approach (3.2) is using supervised, deep learning to discover specific visual properties.

3.1 Machine learning with feature engineering for visual analysis

This approach takes one aspect of images and tries to engineer features to detect it. A more or less vague idea about images and their composition guides the construction of the analytical tool. For example, to account for the left-right or vertical axial symmetry of music album covers, researchers have adapted a standard CNN architecture that was trained for image recognition [14]. Only at the last stages the output was "cut" into two halves to compare the features of each side (Figure 3).

(a) Calculated Symmetry: 0.98 **(b)** Calculated Symmetry: 0.53 **(c)** Calculated Symmetry: 0.23

Figure 3: *Three evaluations of the vertical axial symmetry by the CNN. The closer the value is to one, supposedly the more symmetric is the image [14].*

The approach is straightforward as it takes a trained CNN, assuming that it "sees" the salient parts of an image and then proceeds by simply comparing both sides' features. This comes close to human visual perception that is also very much sensible to the sides of the field of view. At the same time this approach is still narrow, almost technical, accounting for one "ideal" state of balance in a composition. Indeed, vertical axial symmetry is but one option of balancing a composition in graphic design. With no symmetry there can still be balance [3], [15]. Still, this research is relevant for visual communication as it explores the spatial aspects of an image with machine learning.

3.2 Deep learning for visual analysis

Another approach for visual analysis is solely depending on deep learning, i.e. labelled training data is used to then predict such labels. But instead of real-world objects, other aspects of images make up the labels of the training data. A pretrained model is used that is retrained with the new data, which makes this a case of so-called transfer learning. One example of this research is "Recognizing image style" [16] where labels from an internet-based photo community were taken to train a CNN. Based on this training an image search system was built (Figure 4).

Proceedings DOI: 10.21467/proceedings.168 Series: AIJR Proceedings
ISBN: 978-81-970666-6-5 ISSN: 2582-3922

Figure 4: *Found images for the labels "Minimal" (top row) resp. "Melancholy" (bottom row) [16].*

This approach is different from the first, as it does not presuppose an idea about images, instead uses the CNN to learn about the images using labels. Apparently, the CNN "learns" about certain styles and moods in an image – but what do we, humans learn in return?

"The downside of deploying deep models and end-to-end feature learning […] is that a model's performance does not always allow drawing conclusions on how the model's features bring about this performance. In other words, while it is possible to predict whether an image will be considered to be aesthetic or an artwork, it is not easy to understand why this might be the case." [17]. Regarding the examples of figure 4 at the bottom row, we can't be sure if "Melancholy" is just attributed to black and white and very desaturated images.

4 Two propositions for less technical and more insightful visual analysis with machine learning

The short review of two different approaches to visual analysis with machine learning calls for approaches that a more informed by criteria from graphic design practice and provide more insights into the structure of the design. In the remainder of this article, I want to propose two approaches that point into this direction.

4.1 "Above and Below" according to Kandinsky

In 3.1 I have shown how researchers divide the visual plane into two lateral halves to calculate symmetry. This approach is intriguing as it exploits a commonality between the digital image and the human field of view. The digital image as a two-dimensional array shares the same coordinates as our field of view: top and bottom, left and right. By dividing the image vertically and horizontally we obtain a basic structure that can easily be computed and is of great importance for the spatial evaluation of an image. The expressiveness of top and bottom, left and right in a composition has been discussed by visual communication researchers [18]. Already Kandinsky explored the expressiveness of "Above and Below" on the visual plane in an embodied way: "The 'above' gives the impression of a great looseness, a feeling of lightness, of emancipation and, finally, of freedom. […] The effect of 'below' is completely contrary: condensation, heaviness, constraint." [19]. He even provided a formula where the same elements on the lower half are attributed a "heavier", double weight. It seems straightforward to do such a calculation, summing up the brightness values of all pixels, and comparing the top and bottom half. As the sheer amount of color of a graphic design is just one measure of its "weight", also horizontal and vertical edges are summed up to account for weighty visual elements, similar to the discussed approach in 3.1 [14]; this is done with a convolution operation with four predefined filters for top, bottom, left, and right edges. Both computations are executed on the grayscale version of the poster-image (Figure 5).

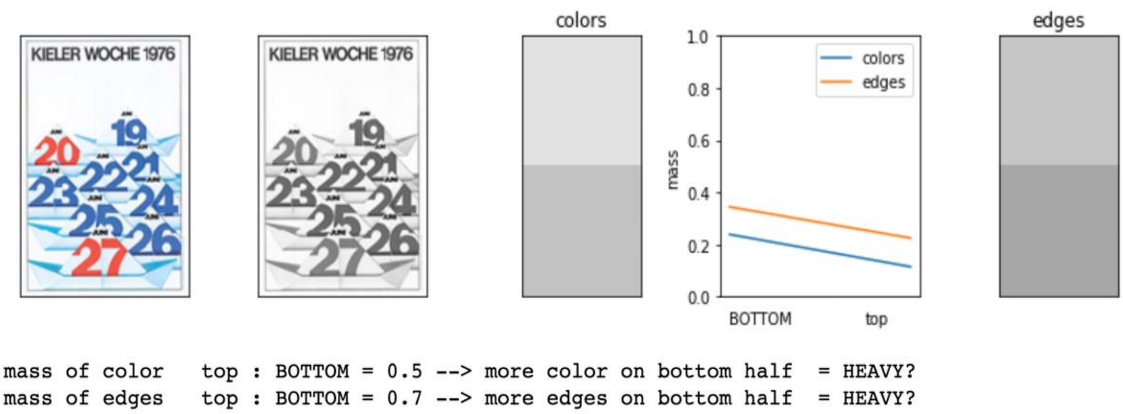

Figure 5: *Counting brightness values of pixels and edges of top and bottom halves of a poster – with a greater portion of both at the bottom, does it feel heavy?*

Figure 5 shows that a disbalance between top and bottom halves can be expressive, as formulated by Kandinsky [19]. It also proves that both measures – brightness values of all pixels and horizontal/vertical edges in the poster – result in about the same evaluation. However, when a poster has inverted colors, i.e., when foreground elements are bright and the background is dark, the evaluation of brightness values is reversed. Therefore, the measure of edges is more reliable. But edges are just one indicator for weight; Klee had already shown that also a smaller patch of an active color appears "heavier" [15]. Also, the attribution of labels like "heavy" and "light" needs to be done with caution as a sparsely populated design would also be evaluated by humans as "light" regardless of its spatial distribution. Still, this simple calculation delivers interesting statistics when using more data (Figure 6).

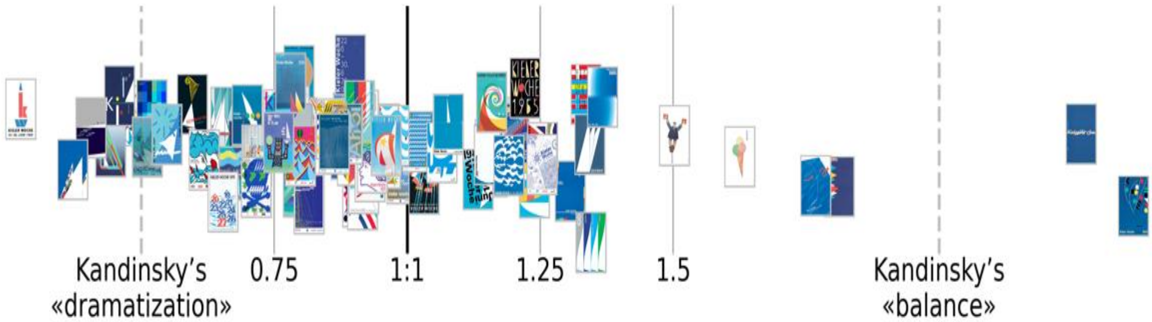

Figure 6: *Computation of edges at top and bottom halves of 78 posters for "Kieler Woche" [20] from 1948 to 2023. To the left posters with more weight on the bottom half, to the right those with more weight on the top half. The vertical distribution is random for maximum visibility.*

The diagram shows a distinct cluster of posters with an equal amount of edges at the top and bottom halves. Furthermore, there are more "heavy" than "light" posters; Kandinsky may have observed "dramatization" there. Very few posters are situated where Kandinsky supposed balance. Kandinsky had speculated: "It can be assumed that in time, perhaps, means will be found of accomplishing measurements in the above sense with more or less exactness. At all events, the formula which I have just roughly outlined could be corrected in such a way that the relative nature of balance would stand out with clarity." [19]. Do we have to correct Kandinsky now? In any case, it is striking to find more "heavy" posters according to Kandinsky here (with an equal number of edges at top and bottom halves he still observed more gravity towards the bottom). This is in accordance with the idea of embodied vision with the ground having a gravitational pull. Can we reproduce this statistic with another set of posters? Indeed, taking a set of contemporary posters designed for different purposes reveals a similar distribution (Figure 7):

Proceedings DOI: 10.21467/proceedings.168
ISBN: 978-81-970666-6-5

Series: AIJR Proceedings
ISSN: 2582-3922

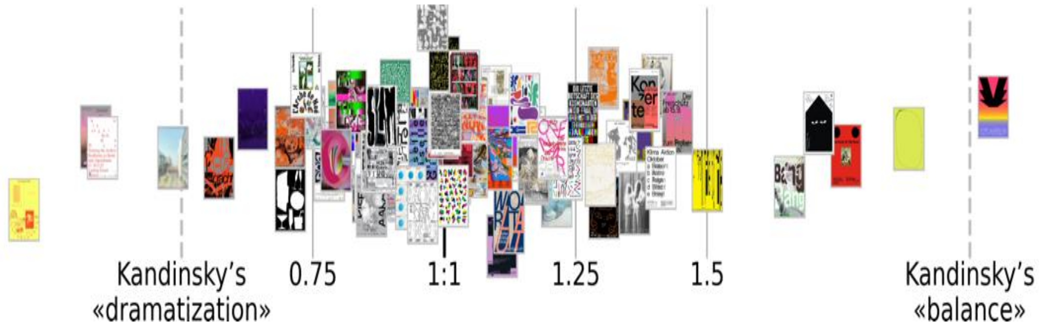

Figure 7: *Computation of edges at top and bottom halves of "100 best Posters 2022" [21]. To the left posters with more weight on the bottom half, to the right those with more weight on the top half. The vertical distribution is random for maximum visibility.*

Here the equal distribution of edges is more distinct; most posters have an equal sum of edges at the top and bottom halves. We can conclude that designers prefer an equal distribution of weights which following the idea of embodied vision still provides an emphasis of the bottom part, thus grounding the composition – without too much of Kandinsky's "dramatization". An extreme emphasis of top or bottom half is rarely found. The potential expressiveness of the portrait format is not exploited. Although top and bottom offer a semiotic resource for meaning making – "fundamental concepts are organized in terms of one or more spatialization metaphors" [22] – its use may be limited due to other constraints on the design. We must also be aware that the simple measure of edges is but one to account for weighty elements. As already mentioned, color contrasts play a role in balancing too. Color has been discarded here as the grayscale version the posters was computed. In any case, this simple calculation shows how a simple spatial division of the format can advance visual analysis.

4.2 Spirit level vision for "spatial feelings" [3]

The edge filters used for the computation presented in 4.1 are at the same time telling for the orientation of the edges, in this case of their horizontal and vertical orientation. Indeed, finding and edge means finding an oriented edge, especially in early human vision. The sensitivity to horizontal and vertical orientations is enforced by the body. Our whole skeleton is symmetrically built; eyes sit horizontally in our head; our main posture is upright. Our body reinforces perpendicularity in vision, we have "spirit level vision". This aspect must be of interest in graphic design. There are certainly distinct emotional assessments possible at the extremes of verticality and horizontality.

Early human visual processing not only accounts for edges and their orientation but perceptually groups these according to their "factor of direction" [9]. As mentioned in 2 this "Gestalt" operation is not found explicitly in CNNs or Transformers. For an analysis of graphic design this aspect is of importance as text is a common ingredient. Text blocks are perceptually grouped around their outer shape mostly resulting in a horizontal block; the inner vertical edges are not perceived as contributing to general verticality. One way to enforce the "factor of direction" in CNNs is to "dilate" the filters so they reach out over a greater area, filtering out shorter, non-continuing edges. As a first measure, verticality and horizontality are compared. Again, only the grayscale version the posters is computed (Figure 8).

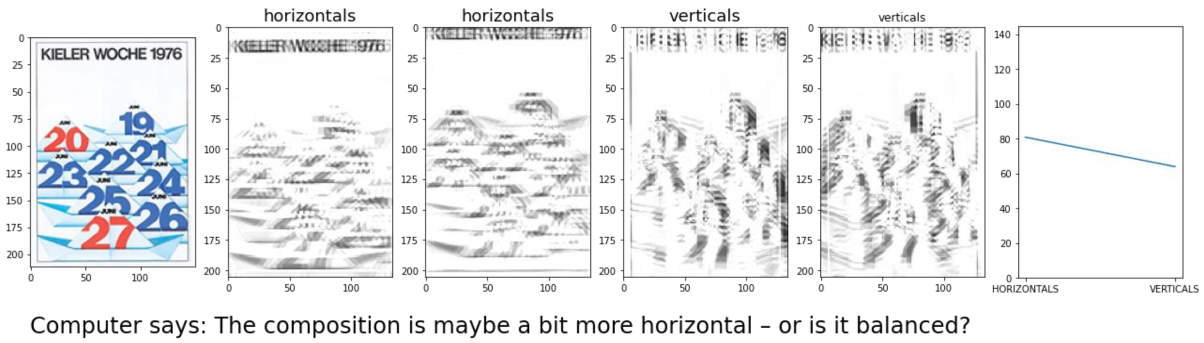

Figure 8: *A poster and its bottom, top, right, and left edges computed. The diagram at the right shows the sum of horizontal (bottom and top) versus vertical (right and left) edges. With more horizontal edges does the poster feel more horizontal? The grouping by the "factor of direction" can best be observed with the top text block at the feature map of the horizontal edges.*

Computing horizontals versus verticals with a set of posters results in a diagram as in figure 9.

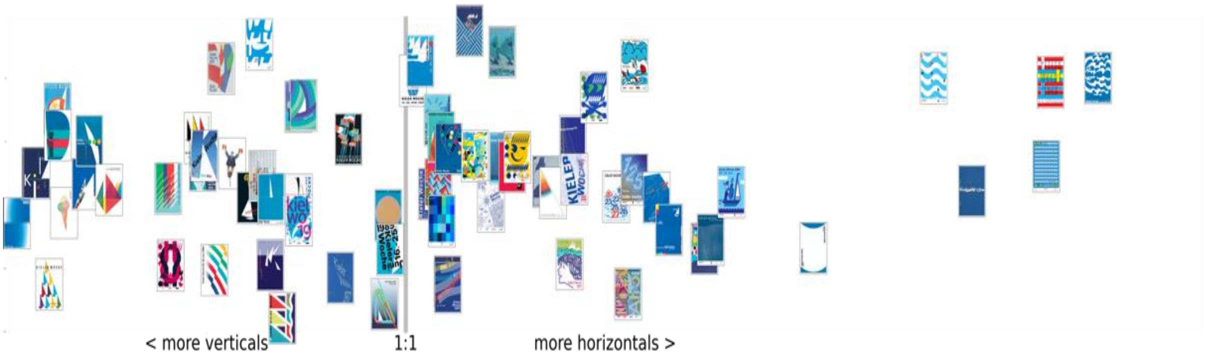

Figure 9: *Computation of the sum of horizontal (bottom and top) versus vertical (right and left) edges of 79 posters for "Kieler Woche" [20] from 1948 to 2023. To the left posters with more verticals, to the right those with more horizontals. The vertical distribution of posters is random for maximum visibility.*

The diagram reveals a cluster around the equal distribution without concentration there; noteworthy are some clearly horizontal oriented posters. In this case, as all posters promote the same sailing event this may be due to the representation of water and waves. The vertical ones often have shapes of a sail. Can the distribution be reproduced with other posters? The same set of posters as in 4.1 is used for comparison (Figure 10).

Figure 10: *Computation of the sum of horizontal (bottom and top) versus vertical (right and left) edges of "100 best Posters 2022" [21]. To the left posters with more verticals, to the right those with more horizontals. The vertical distribution of posters is random for maximum visibility.*

Proceedings DOI: 10.21467/proceedings.168
ISBN: 978-81-970666-6-5

Series: AIJR Proceedings
ISSN: 2582-3922

Here, a more pronounced clustering near the equal distribution of horizontals and verticals can be observed – most posters appear near the "1:1" line. Again, some posters are clearly more horizontally oriented, more than clearly vertically oriented ones. This may come as a surprise, as the poster format, the portrait format is a vertical one. Why is verticality not exploited more? One reason may be that text which is an important element in every graphic design emphasizes horizontality. As I have mentioned earlier, a text block is rather perceived as a horizontally stretched rectangle, as described by Gestalt principles. Also, the vertical format itself constrains the design process; with several elements to be laid out, the format "suggests" a vertical stacking of elements from top to bottom which may result in some horizontal clusters.

5 Results and Discussion

The two presented analyses show that the simple features – division of the format, and oriented edge detectors – can reveal interesting insights. As they come close to human embodied vision, they are of relevance for visual communication research and practice. Some tacit knowledge of graphic designers can be reconstructed and discussed with the technology. The machine learning is "unsupervised" as it is not working with predefined labels as with deep learning. This makes analyzes simple and fast; large amounts of images can be processed on a standard computer.

6 Conclusions

My exploration sheds light on the intersection of machine learning and graphic design, offering insights into the potential applications and limitations of computer vision in visual analysis. The analyses I present highlight the need for a more nuanced and informed approach to visual analysis, one that takes into account the embodied nature of human vision and the complex spatial relationships within graphic compositions. By proposing two approaches that implement criteria of graphic design practice like balance and orientation I underscore the importance of integrating domain-specific knowledge into machine learning frameworks. Through the application of these approaches, I demonstrate how machine learning can be leveraged to uncover tacit knowledge and provide deeper insights into the spatial aspects of graphic design. By bridging the gap between computational methods and design theory, I believe researchers and practitioners can gain a richer understanding of visual communication and enhance their analytical capabilities. Moving forward, it is imperative to continue exploring interdisciplinary collaborations between machine learning experts and graphic designers to refine existing methodologies and develop new tools for visual analysis. By embracing a more holistic approach that combines computational techniques with domain-specific expertise, we can unlock new possibilities for advancing the field of graphic design and harnessing the full potential of machine learning in visual communication.

7 Declarations

7.1 Competing Interests

The author declares no conflict of interest.

7.2 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in in published maps and institutional affiliations.

How to Cite

Jan-Henning Raff (2024). Machine Learning for Basic Visual Research in Graphic Design. *AIJR Proceedings*, 232-240. https://doi.org/10.21467/proceedings.168.26

References

[1] J.-H. Raff, "A model to identify the data needed for ai-driven graphic design," in *Designing with Artificial Intelligence*, 2021.

[2] J. J. Gibson, *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.

[3] G. Kepes, The Language of Vision: Painting, Photography, Advertising-Design. Chicago: Paul Theobald, 1969.

[4] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," The Journal of physiology, vol. 148, no. 3, pp. 574–591, Oct. 1959.

[5] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.

[6] A. Ghiasi *et al.*, "What do Vision Transformers Learn? A Visual Exploration," *arXiv e-prints*, pp. 1–44, 2022.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[8] "ConvNeXT (tiny-sized model)," *Hugging Face*, 2022. [Online]. Available: https://huggingface.co/facebook/convnext-tiny-224. [Accessed: 18-Sep-2023].

[9] M. Wertheimer, "Principles of Perceptual Organization," in *Readings in perception*, D. C. Beardslee and M. Wertheimer, Eds. Princeton: D. Van Nostrand, 1958, pp. 115–135.

[10] S. E. Palmer, "Perceptual Organization in Vision," in *Stevens' Handbook of Experimental Psychology.Volume 1: Sensation and Perception*, 3rd ed., H. Pashler and S. Yantis, Eds. New York: John Wiley & Sons, 2002, pp. 177–234.

[11] N. Baker, H. Lu, G. Erlikhman, and P. J. Kellman, "Local features and global shape information in object classification by deep convolutional neural networks," *Vision Research*, vol. 172, no. March, pp. 46–61, 2020.

[12] D. Heinke, P. Wachman, W. van Zoest, and E. C. Leek, "A failure to learn object shape geometry: Implications for convolutional neural networks as plausible models of biological vision," *Vision Research*, vol. 189, pp. 81–92, 2021.

[13] P. Mehrani and J. K. Tsotsos, "Self-attention in Vision Transformers Performs Perceptual Grouping, Not Attention," *arXiv e-prints*, pp. 1–30, 2023.

[14] A. Brachmann and C. Redies, "Using Convolutional Neural Network Filters to Measure Left-Right Mirror Symmetry in Images," *Symmetry*, vol. 8, no. 12, Dec. 2016.

[15] P. Klee, *Notebooks Volume 1: The thinking eye*. London: Lund Humphries, 1973.

[16] S. Karayev et al., "Recognizing image style," BMVC 2014 – Proceedings of the British Machine Vision Conference 2014, pp. 1–20, 2014.

[17] A. Brachmann, E. Barth, and C. Redies, "Using CNN features to better understand what makes visual artworks special," *Frontiers in Psychology*, vol. 8, no. May, pp. 1–14, 2017.

[18] G. Kress and T. van Leeuwen, *Reading Images. The Grammar of Visual Design*, 2nd ed. London: Taylor & Francis, 2006.

[19] W. Kandinsky, Point and line to plane. Contribution to the analysis of the pictorial elements. New York: Solomon R. Guggenheim Foundation, 1947.

[20] "Kieler-Woche-Design – weltweit bekannt," *Kieler Woche - Keine Woche ist wie diese*, 2023. [Online]. Available: https://www.kieler-woche.de/de/service/design.php#plakate. [Accessed: 29-Sep-2023].

[21] "Plakate des Jahres 2022," *100 Beste Plakate e. V. | Website zum Wettbewerb 100 beste Plakate*, 2023. [Online]. Available: https://100-beste-plakate.de/archiv-plakate/?jahr=2022. [Accessed: 29-Sep-2023].

[22] G. Lakoff and M. Johnson, *METAPHORS We Live By*. Chicago and London: The University of Chicago Press, 1980.