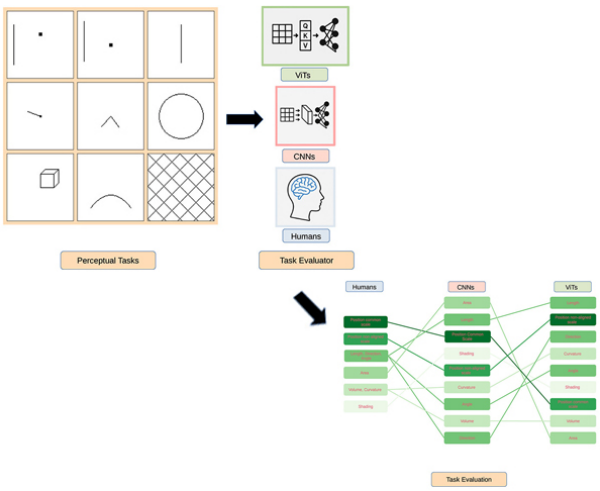
ViRVIG
Visualització, Realitat Virtual i Interacció Gràfica

www.virvig.eu
C. Llorens i Artigas 4-6
Edifici U, 1er pis
E- 08028- Barcelona

Evaluating graphical perception capabilities of Vision Transformers

Poonma, P.; Vazquez, Pere-Pau; Ropinski, T.



Vision Transformers (ViTs) have emerged as a powerful alternative to convolutional neural networks (CNNs) in a variety of image-based tasks. While CNNs have previously been evaluated for their ability to perform graphical perception tasks, which are essential for interpreting visualizations, the perceptual capabilities of ViTs remain largely unexplored. In this work, we investigate the performance of ViTs in elementary visual judgment tasks inspired by Cleveland and McGillâ€™s foundational studies, which quantified the accuracy of human perception across different visual encodings. Inspired by their study, we benchmark ViTs against CNNs and human participants in a series of controlled graphical perception tasks. Our results reveal that, although ViTs demonstrate strong performance in general vision tasks, their alignment with human-like graphical perception in the visualization domain is limited. This study highlights key perceptual gaps and points to important considerations for the application of ViTs in visualization systems and graphical perceptual modeling.

http://dx.doi.org/10.1016/j.cag.2025.104458

UPC