Beyond Pixels: A Mathematical Framework for Tracking Visual Distortions in 3D

Ramon Visser, Coert van Gemeren, Alpha One

1. Egetvelitaliquetsagittisid consecteturpurus, Utrecht, the Netherlands| 2. Ut faucibuspulvinar elementuminteger, Sofia, Bulgaria|3. Enimnequevolutpatac tinciduntvitae semper quis, Budapest, Hungary

ABSTRACT

Logos are powerful brand symbols, instantly recognizable to consumers. However, accurately identifying logos in real-world situations, like videos of products moving in a store, presents a significant challenge. Existing logo recognition systems struggle when logos are distorted, such as when seen from an angle or partially obscured. Furthermore, high-quality, annotated data for logo recognition is often scarce, particularly for niche markets or specific regions like the Netherlands. This research aims to improve logo recognition by enhancing convolutional neural networks (CNNs). The approach involves incorporating mathematical equations into the CNN architecture to better understand and account for visual distortions in moving imagery. The research aims to find a solution to the question of:" How can a convolutional neural network architecture be expanded with mathematical equations in order to detect visual distortion in moving imagery in order to increase the functionality and explainability of the architecture?" By tracking how visual elements, including logos, are distorted in a 3D environment, we can create a more robust and explainable CNN model for accurate logo recognition, even in challenging real-world scenarios.

LITERATUREREVIEW

Tracking objects over time is difficult. Multiple object tracking (MOT) involves tracking multiple objects simultaneously in a video. This is difficult because objects often overlap, disappear and reappear from view, or change their appearance. The problem is exacerbated when many objects are in view at once. Thus, we need more robust methods that can track objects over time, even when they are partially or fully obscured [2][3]. Deep learning is computationally and time-consuming. Traditional deep learning models often require huge amounts of data, computational power (such as dedicated graphics cards or GPUs), and long training times. This makes it difficult to use these models in real-world devices or situations where fast responses are required. Thus, there is a need for more efficient methods that require fewer resources and train faster [4][5]. Logo recognition in documents is difficult due to poor quality scans and photos, obscured or distorted logos. and poor logo segmentation (cutting out the logo from the background). Methods that work well on perfect photos do not work well in these cases. Identifying logos is not simply a matter of "finding a shape," especially when that shape is distorted or blurred. Developing techniques that can withstand this challenge is crucial[1].

RESEARCH METHODOLOGY

The research will be carried out in 10 weeks and will begin by thoroughly examining the current convolutional neural network (CNN) architecture used for logo recognition. This investigation will focus on understanding how the network currently processes visual information and identifying its limitations, particularly in handling distorted logos. Based on the analysis performed, a set of mathematical formulas will developed that can accurately model and predict the effects of various distortions on visual elements within an image. These formulas will consider factors such as perspective shifts, rotations, and occlusions. The research will then integrate these mathematical formulas into the existing CNN architecture. This expansion will allow the network to not only recognize logos but also understand and compensate for the effects of distortions, improving its accuracy and robustness. Finally, the research will rigorously test the enhanced network. The research will evaluate its performance in terms of accuracy, speed, and robustness. Furthermore, this research will investigate methods to make the network's decision-making process more transparent and understandable, enhancing its explainability and building trust in its results.

PRELIMINARY CONSIDERATIONS

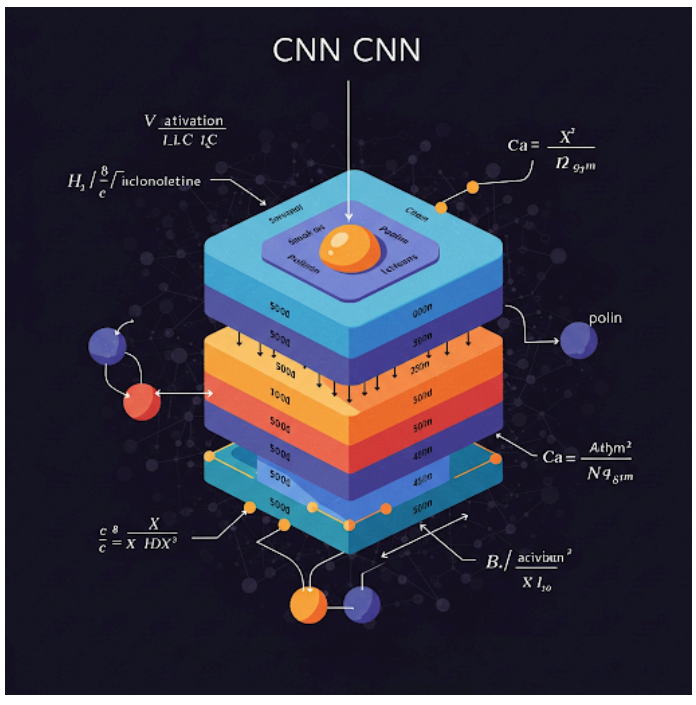
This research can lead to improved CNN architectures, such as those with more efficient convolutional layers or novel attention mechanisms, can lead to significant gains in accuracy, speed, and resource efficiency across various applications. This includes advancements in image recognition, object detection, medical imaging analysis, and natural language processing. Furthermore this research can lead to better explainability of CNNs which is crucial for building trust in AI systems. Techniques like visualizing feature maps, generating attention heatmaps, and developing simpler, more interpretable architectures will allow for greater understanding of how CNNs make decisions, leading to more reliable and trustworthy AI systems.t

CONCLUSIONS

Logos are powerful brand symbols, instantly recognizable to consumers. However, accurately identifying logos in real-world situations, like videos of products moving in a store, presents a significant challenge. But tracking objects over time is difficult. Multiple object tracking (MOT) is about tracking multiple objects simultaneously in a video. This is difficult because objects often overlap, disappear and reappear, or become distorted. This research aims to improve the general architecture of the CNN, its explainability and the amount of training data required to train CNNs. The research will be carried out over ten weeks.

References.

1. Zhao, Z., Zheng, P., Xu, S., & Wu, X. (2018). Object Detection with Deep Learning: A Review. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1807.05511
2. Alaei, A., & Delalandre, M. (2014). A Complete Logo Detection/Recognition System for Document Images. 11th IAPR International Workshop On Document Analysis Systems, 324–328. https://doi.org/10.1109/das.2014.79
3.Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T. (2020). Multiple object tracking: A literature review. Artificial Intelligence, 293, 103448. https://doi.org/10.1016/j.artint.2020.103448
4.Chen, W., Li, Y., Tian, Z., & Zhang, F. (2023). 2D and 3D object detection algorithms from images: A Survey. Array, 19, 100305. https://doi.org/10.1016/j.array.2023.100305
5.Angelov, P., & Soares, E. (2020). Towards explainable deep neural networks (xDNN). Neural Networks, 130, 185–194. https://doi.org/10.1016/j.neunet.2020.07.010



Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This Poster Was Created As Part of the Blended IP Event Organized Under Erasmus + Programme of the European Union.