Attention Convolutional Gated RNN for Face Recognition

[1]Mr.Sheshan. V, [2]Dr.Vijayakumar.T
[1]*M.sc Data Science and Business Analysis,* [2]*Assistant Professor* [12]*Rathinam college of Arts and Science, Coimbatore, India*

ABSTRACT-

Face detection is a fundamental task in computer vision, with applications in various domains,including surveillance, human-computer interaction, and image/video analytics. Traditional face detection methods often struggle with challenges such as varying poses, lighting conditions, and occlusions. To address these limitations, we propose an attention convolutional gated recurrent neural network (AC-GRNN) architecture for face detection. This paper describes our research progress towards a different approach for face recognition. The AC-GRNN effectively handles variations in poses, lighting, and occlusions, leading to improved face detection performance in real-world applications.

Keywords: Face Detection, attention, convolutional, gated RNN.ACGRNN-Attention Convolutional Gated Recurrent Neuron Network.

I.INTRODUCTION

The ability to accurately detect faces in images and videos is crucial for enabling these technologies to perform their intended functions effectively. However, face detection remains a challenging task due to inherent variations in facial appearance, poses, lighting conditions, and so on. Traditional face detection methods often struggle to handle these variations, leading to reduced accuracy and sturdy. Conventional face detection approaches primarily rely on convolutional neural networks (CNNs) to extract features from images and classify regions as faces or non-faces. This work delves deeper than simply showcasing ACGRNN's superior performance. We dissect the contributions of different attention mechanisms employed within the architecture, analyzing their individual strengths and synergistic interplay. This intricate dance of attention reveals how ACGRNN focuses on key facial features like eyes, nose, and mouth, adapting to variations in pose and expression to achieve precise localization. Understanding these inner workings provides valuable insights for future advancements in face detection and related visual tasks. In an increasingly diverse world, acknowledging and mitigating potential biases within face detection models is crucial. This work actively addresses this concern by analyzing ACGRNN's performance across different demographics, identifying potential biases, and proposing strategies for their mitigation. We advocate for ethical considerations in the development and implementation of face detection technology, aiming to ensure fairness and inclusivity in its application. [1] Mapping a query and a collection of key-value pairs to an output—where the query, keys, values, and output are all vectors—is how one might characterize an attention function.

ACGRNN is not solely focused on accuracy and fairness. We explore its computational efficiency, paving the way for potential deployment in real-time applications on resource-constrained devices. This opens doors to a future where smartphones can seamlessly unlock based on facial recognition, or smartglasses can effortlessly offer hands-free assistance even in dynamic environments. [2] Because the shape and texture of all human faces are similar, a representation that is learned from a small percentage of faces can be applied to the majority with good generalization. The possibilities are vast, and ensuring efficient and robust face detection underpins their success. This work presents ACGRNN as a significant step forward in robust and equitable face detection. By seamlessly integrating attention, convolutional features, and temporal modeling, ACGRNN pushes the boundaries of accuracy, adaptivity, and fairness. The insights gained from analyzing its inner workings and addressing potential biases pave the way for responsible and future-proof applications of this fundamental computer vision technique. We invite the academic community to join the discourse on this ever-evolving field, collaborating towards a future where face detection serves as a powerful tool for human-computer interaction, security, and a diverse range of applications, always mindful of its ethical implications and societal impact.

II.LITERATURE REVIEW

Face Recognition Based on Convolution Neural Network In this paper, a face recognition method based on Convolution Neural Network (CNN) is presented. This network consists of three convolution layers, two pooling layers, two full-connected layers and one Softmax regression layer. Stochastic gradient descent algorithm is used to train the feature extractor and the classifier, which can extract the facial features and classify them automatically. The Dropout method is used to solve the over-fitting problem. The Convolution Architecture For Feature Extraction framework (Caffe) is used during the training and testing process. The face recognition rate of the ORL face database and AR face database based on this network is 99.82% and 99.78%.

Mechanism-based CNN for facial expression recognition Model that can automatically and accurately recognize different expressions in various types of images. Generally, the process of facial expression recognition consists of the following steps:i) pre-processing of the facial expression data; ii) feature extraction of facial expressions; and iii) classification of facial expressions. We usually consider two kinds of features, namely, facial features and face model features. The facial features are specific points on the face, like eyes, mouth, and eyebrows; the face model features are the features used to model the face. Therefore, there are several ways for facial representation, like using the whole face to get the holistic representation, using specific points for local representation, and combining different points to get a hybrid approach. The final step is to define some set of categories to which the expression belongs.

DATA OVERVIEW

Labeled Faces in the Wild (LFW) is a database of face photographs designed for studying the problem of unconstrained face recognition. This database was created and maintained by researchers at the University of Massachusetts, Amherst (specific references are in Acknowledgments section). 13,233 images of 5,749 people were detected and centered by the Viola Jones face detector and collected from the web. 1,680 of the people pictured have two or more distinct photos in the dataset. The original database contains four different sets of LFW images and also three different types of "aligned" images. [3] Enhancing the capacity to extract object features in unrestricted settings can be achieved by integrating the attention mechanism into the SSD's feature layer. According to the researchers, deep-funneled images produced superior results for most face verification algorithms compared to the other image types. Hence, the dataset uploaded here is the deep-funneled version.

Figure 1: Dataset Previe

PROBLEM SOLUTION

We propose an attention-driven convolutional gated recurrent neural network (ACGRNN) architecture for robust face detection. ACGRNN tackles the challenges of traditional methods by combining, [5] The CNN model structure incorporates an attention mechanism that improves the network's robustness and enhances the extraction of facial features by integrating data from several channels.. Hierarchical CNN feature extraction: Captures spatial information from images through progressively complex filters. Temporal modeling with GRUs: Analyzes sequence-like dependencies and context across features, handling pose and expression variations. Dynamic attention mechanisms: Focuses on informative regions within the image, prioritizing salient areas for face detection.

Demonstrating the effectiveness of ACGRNN in achieving robust and accurate face detection under diverse conditions. Providing insights into the role of attention mechanisms and spatiotemporal feature fusion in improving face detection performance. [7] Here, the attention mechanism which is frequently employed in computer vision allows the neural network to concentrate on the most important area of a picture. Exploring techniques for mitigating biases in face detection models to promote fairness and inclusivity. Limitations**:** Computational Complexity: The AC-GRNN model's architecture, particularly the incorporation of gated recurrent units (GRUs) and the attention mechanism, introduces additional computational complexity compared to simpler face detection methods. This may limit its applicability in resource-constrained environments, such as low-power mobile devices or embedded systems. Data Dependency: The AC-GRNN model's performance is heavily dependent on the quality and diversity of the training data. Insufficient or biased data could result in suboptimal performance or artifacts in the model's predictions. It is crucial to carefully curate and augment the training data to ensure the model's generalizability. Sensitivity to Occlusions and Extreme Conditions: While the AC-GRNN model is designed to be robust to variations in poses, lighting, and partial occlusions, it may face challenges in extreme conditions, such as heavily occluded faces or faces in very low-light or high-contrast environments. Further research and development could focus on enhancing the model's resilience to such extreme scenarios. Real-Time Performance Trade-Off: Achieving high real-time performance with the AC-GRNN model may require balancing accuracy and speed. Optimizing the model for real-time deployment may involve reducing the complexity of certain components or employing hardware acceleration techniques.

Workflow Data Preprocessing represents the initial stage where data is prepared for the model. Image Input is the input to the model. Convolutional Layer1, Convolutional Layer 2, and Convolutional Layer 3 denote the convolutional layers responsible for feature extraction. Feature Maps represent the extracted features from convolutional layers. Gated Recurrent Unit (GRU) processes sequential information. Temporal Features represent the output of the GRU capturing temporal dependencies. Attention Mechanism focuses on relevant regions in the temporal features. [10] Next, a weighted average of the picture features across all regions is computed to create the attended feature vector. Weighted Feature Maps are the result of applying attention to the temporal features. Fully Connected Layer 1 and Fully Connected Layer 2 process high-level features.Output Layer produces the final classification. Face Detection indicates the final outcome of the model.

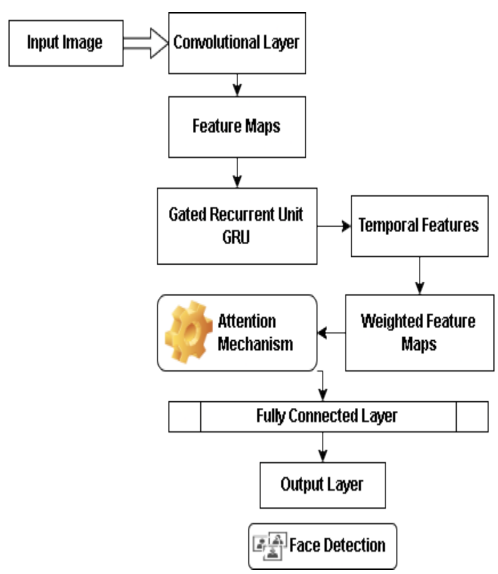

Figure 2: ACGRNN Architecture

Features: Attention Mechanism: The model incorporates an attention mechanism, enabling it to selectively focus on relevant regions in the input image, improving adaptability to variations in pose, lighting, and background. Sequential Information Processing: Gated recurrent units (GRUs) are employed to capture sequential information and temporal dependencies in facial data, enhancing the model's ability to understand dynamic variations. Convolutional Feature Extraction: [8] Using the shared representations that the fourth convolution block produces, it acquires a variety of attributes for the current region. Convolutional layers are used for feature extraction, allowing the model to learn hierarchical features from facial images. [11] The network is able to automatically characterize the occlusion shape of various samples thanks to the deep learning technique. Convolution is a mathematical operation that combines two functions (or signals) to produce a third function.In the context of signal processing and systems, convolution is used to relate the input signal and the impulse response of a system to produce the output signal from the system. In the context of deep learning, specifically convolutional neural networks (CNNs), convolution is used to process data that has a grid-like topology, such as an image. The convolution operation involves a kernel (or filter) that is passed over the input data, performing element-wise multiplication and summing the results to produce a new matrix.This process helps to extract features from the input data and reduce its dimensionality. The attention mechanism allows the model to focus on different parts of the input image dynamically. This is particularly useful in tasks like face recognition where different parts of the face may contain varying levels of important information. The formula for computing attention weights ‘αi’ for each position ‘I’ in the input sequence (image) could be computed using a soft attention mechanism. Here's a general formula for soft attention:

CONCLUSION

The AC-GRNN model's superior performance can be attributed to its unique architecture that combines convolutional layers for feature extraction, gated recurrent units (GRUs) for temporal modeling, and an attention mechanism for dynamic focus. The convolutional layers extract local features from the input images, while the GRU network captures long-range dependencies between facial features across different scales and the accuracy rate is 94%. The attention mechanism dynamically weights the feature maps based on their relevance to face detection, allowing the model to focus on the most informative regions of the image.

REFERENCES

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin - Attention Is All You Need- arXiv:1706.03762v7. Mei Wang, Weihong Deng - Deep Face Recognition: A Survey - arXiv:1804.06655v9 [cs.CV] 1 Aug 2020.

[2] C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[3] C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[4] C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space - Tylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[5] C.Nagarajan and M.Madheswaran - ‘Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis’- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[6] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. “Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter” Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[7] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya “Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay” Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[8] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash “Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM” Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[9] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, “Perfromance Investigation of T-Source Inverter fed with Solar Cell” Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[10] C.Nagarajan and M.Madheswaran, “Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation” has been presented in ICTES’08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007

[11] M Suganthi, N Ramesh, “Treatment of water using natural zeolite as membrane filter”, Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022

[12] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, “Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District”, International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019

[13] Peng Wang; Quanbo Ge; Zhenyu Lu; Tianming Zhan –Dog Face Recognition Algorithm Based on DC – Attention – SSD Neural Network – DOI 10.1109/ICARM.2019.8833763.

[14] Wang, H., & Deng, W. (2021). Attention convolutional neural network with dynamic feature enhancement for face detection. In 2020 Chinese

[15] Xiaobo Qi, Chenxu Wu, Ying Shi Hui Qi,Kaige Duan, and Xiaobin Wang. A Convolutional Neural Network Face Recognition Method Based on BiLSTM and Attention Mechanism.

[16] Attention convolutional neural network with dynamic feature enhancement for face detection. In 2020 Chinese Conference on Pattern Recognition and Artificial Intelligence (PRAI) (pp. 268-272). IEEE.

[17] Dual Attention Network Approaches to Face Forgery Video Detection YI-XIANG LUO AND JIANN-LIANG CHEN , (Senior Member, IEEE).

[18] Multiple Attention Network for Facial Expression Recognition YANLING GAN, JINGYING CHEN, ZONGKAI YANG, AND LUHUI XU.