IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A REVIEW OF FACE OCCLUSION DETECTION: TRANSITIONING FROM CONVOLUTIONAL MODELS TO SPIKING AND LIQUID NEURAL NETWORKS

Deepak Jangid
Lecturer (Computer Engineering)
Govt. Polytechnic College, Jodhpur

ABSTRACT

Detection of occlusion in faces is still a big problem in face recognition systems. Real-world scenarios where individuals are partially obscured by masks, sunglasses, or objects passing by greatly reduce the effectiveness of traditional deep learning models in accurately identifying faces. Convolutional Neural Networks (CNNs) have been very successful in this area, but require significant CPU power and are not well suited to real-time use on edge devices. In this paper, advanced neural network architectures—such as Spiking Neural Networks (SNNs) and Liquid Neural Networks (LNNs)—are explored in detail and compared in the context of face occlusion detection. These models are compared with conventional networks, explained in relation to neuro-morphic event based vision.

Keywords: Face Occlusion Detection, Spiking Neural Networks (SNNs), Liquid Neural Networks (LNNs)

I. INTRODUCTION

Facial recognition has become a basic technology in biometric authentication systems, video surveillance and human-computer interaction (HCI). But occlusion, which refers to blocking the camera's line of sight with the face, greatly affects the reliability of the system. Occlusions hide important facial features (such as eyes, nose, lips) which are key to a correct identification and expression analysis.

The state-of-the-art method for occlusion detection is currently deep learning, mainly based on CNNs and Vision Transformers (ViTs). These models are based on large datasets and try to estimate and reconstruct occluded areas. But in resource-limited systems like autonomous vehicles or smart cameras with IoT, CNNs are not very practical due to their computational requirements [1].

Neuromorphic computing is a promising paradigm to overcome these limitations. Spiking Neural Networks (SNNs), which emulate the function of biological neural mechanisms, use discrete binary spikes for communication, are extremely energy-efficient and are well suited for power-constrained sensors [17]. On the other hand, Liquid Neural Networks (LNNs) have the exceptional feature that they can continuously update their parameters during inference, enabling them to be very robust to dynamic occlusions [13]. This paper aims to give an extensive overview of these two technologies in relation to face occlusion, analyzing their architectures, the computational complexity and their deployability in comparison to the traditional models.

II. BACKGROUND AND TRADITIONAL APPROACHES

Occlusion is a general problem in face recognition systems that can be divided into two categories:

1. **Static Occlusion:** Stationary objects in the frame, like medical mask, scarv or sunglasses.
2. **Dynamic Occlusion:** Temporary occlusion of video streams (e.g. hand covering face, object between the camera and the person being shot).

A. CNN-Based Approaches and Limitations: Occlusions have traditionally been a common problem addressed by CNNs. A study [1] showed that popular networks such as VGG16 and ResNet perform well on standard datasets, but significantly underperform in occlusion scenarios in the wild. To overcome this, sophisticated CNN frameworks have been created:

Adversarial Learning: Zhao et al. [2] introduced a model learning adversarial occlusions using backpropagation to get high accuracy on AR and LFW-Masked datasets.

Attention Mechanisms: To encourage the model to pay attention to the unoccluded part of the face, Chen et al. [3] introduced a Feature Fusion Residual Attention Network.

B. Vision Transformers (ViTs): In recent years, ViTs have started to outperform CNNs in facial occlusion tasks, owing to their strong global contextual information. Gao & Zhao [4] developed the Transformer Architecture for Occlusion (TFE) to achieve good reconstruction of hidden facial information. Likewise, Huang et al. [5] showed state-of-the-art performance on complex datasets, such as RAF-DB, using mid-level representation learning in their FERMixNet network.

Traditional Networks have some limitations: CNNs and ViTs are accurate, but they densely process every pixel of each frame. In the case of continuous dynamic occlusion in video streams, such models require vast amounts of computation power which is not suitable for low-power, real-time edge devices.

III. SPIKING NEURAL NETWORKS FOR OCCLUSION DETECTION

SNNs are considered as the third generation of neural networks [6]. Unlike CNNs, that process information using continuous values (floating points), SNNs process information in discrete time-series ‘spikes'.

A. Occlusion in Event-Based Vision

The occlusion is a kind of Event-Based Vision. The best results are obtained by SNNs in combination with neuromorphic Dynamic Vision Sensors (DVS) [7]. Event cameras do not record dense RGB frames, but rather only the pixels that are changing (e.g., motion) to record events. Neurological research has demonstrated that the frontotemporal network in the brain is heavily dependent on such spatiotemporal cues in processing occluded faces [8]. This biological process can be simulated computationally in SNNs by inhibiting static background and only initiating computations when the moving occluding object appears.

B. Existing SNN Research

There has been significant research that has confirmed the use of SNNs in facial and object occlusion:

1. **Object Detection:** Kim et al. [9] showed that SNNs can be used for deep object detection with a considerable energy saving over similar CNNs through their design of Spiking-YOLO.
2. **Object Tracking** Ji et al. [10] proposed the Spiking Convolutional Tracking Network (SCTN). The model is able to track moving objects in the presence of high level occlusions, while only processing the active pixels, thus saving energy and computational bandwidth.
3. **Gait Recognition:** Addressing biometric occlusion, Tao et al. [11] developed GaitSpike, an SNN model to recognize human gait during dynamic occlusion. The network uses Leaky Integrate-and-Fire (LIF) neurons to build up spikes over time to estimate motion behind obstacles.
4. **Training Challenges:** Traditional backpropagation cannot be directly used since the output are non-differentiable because they are spikes. To approximate derivatives and train deep SNNs on complex facial data, researchers use Surrogate Gradient Learning [12].

IV. LIQUID NEURAL NETWORKS FOR DYNAMIC OCCLUSION

A class of Continuous-Time Recurrent Neural Networks (CTRNNs) based on Ordinary Differential Equations (ODEs) is called Liquid Neural Networks (LNNs) [13].

A. Liquid Time-Constants and LSMs: Liquid State Machines (LSM) are the architectural basis of LNNs. Maass [14] describes an LSM as a 'liquid reservoir' to safely store temporal streams of information. The unique feature of an LNN is that its time-constants of its neurons change dynamically at run-time, even after it has been trained [13].

B. The Role of LNNs in Dynamic Occlusion

1. **Continuous Adaptation:** The input of the data stream becomes very noisy when dynamic facial occlusion happens (e.g., a person covers his mouth when talking). LNNs are extremely tolerant of this unexpected sound. They quickly adjust their internal time-constants to keep the computations to the currently seen facial features only, and thus avoid network collapse when data points disappear.
2. **Video Analysis via PLSM:** Das et al. [15] created a Parallelized Liquid State Machine (PLSM) that can interpret complex video dynamics and unintentional occlusions. Their experiments demonstrated that their approach is superior to traditional models in dynamic video content, and that they are able to capture the temporal correlations of occlusions.

V. COMPARATIVE ANALYSIS AND BENCHMARKING

To appreciate the paradigm shift of deep learning to alternative neuromorphic methods it is necessary to analyze in detail how the methods differ in many domains, such as their computational complexity, energy consumption and mathematical bases. The following subsections contain an in-depth analysis of this situation, whereas Table I gives a high-level summary.

Table I: Technical Comparison of Neural Network Architectures for Occlusion Detection

Evaluation Metric	Traditional Models (CNN / ViT)	Spiking Neural Networks (SNN)	Liquid Neural Networks (LNN)
Data Processing Type	Frame-by-frame (Dense RGB)	Event-driven, Binary Spikes (Sparse)	Continuous-time (Time-series / Video)
Energy Efficiency	Very Low (High GPU dependency)	Highest (On Neuromorphic Chips)	High (Due to compact sparse networks)
Dynamic Occlusion Robustness	Low (Relies on static patterns)	Moderate (Spatiotemporal tracking)	Highest (Runtime adaptation)
Core Mathematics	Matrix Multiplication (MACs)	Leaky Integrate-and-Fire (LIF)	Ordinary Differential Equations (ODEs)

A. Computational Complexity and Energy Efficiency

The key difference between CNNs and SNNs is their mathematical functions. CNNs are based on high-density Multiply-Accumulate (MAC) operations. The number of MACs required to process a single 1080p frame and detect an occluded face is billions – even if a face does not move. Yet, SNNs employ few and sparse Accumulate (AC) operations. Since spikes are binary (0 or 1), no multiplication is used. SNNs only consume energy when the state of a pixel changes, which means that they can save energy in static environments [9]. A drawback of LNNs compared to SNNs is

the memory consumption is much smaller than that of CNN because they are more computationally expensive, which is caused by the requirement of ODE solver but is compensated by being exponentially less in number of neurons to perform the same task.

B. Temporal Processing and Latency

Convolutional neural networks (CNNs) are spatially dominant, but temporally blind. Adding 3D-CNNs or recurrent layers (LSTMs) are required to detect dynamic occlusions in video, which drastically inflate latency. SNNs have inherent temporal dynamics – the membrane potential of a LIF neuron decays over time and, thus, has a natural short-term memory to follow the trajectory of an occluding object with sub-millisecond latency. LNNs go one step further and operate in continuous time, i.e. they can process inputs arriving at irregular intervals, which is a major advantage when dealing with asynchronous event-camera data [13].

C. Robustness to Distribution Shifts and Noise

If a CNN sees an occlusion pattern it has not trained on (e.g. a scarf with a unique pattern) it experiences a distribution shift and has catastrophic failures. LNNs solve this thanks to their “liquid” nature. An LNN does not have fixed weights and time constants after training. During inference, an occlusion that introduces noise to the visual feed causes the LNN to adapt its internal equations dynamically to stabilize the output, a robustness not seen before in unconstrained environments [15].

VI. CONCLUSION

Face occlusion detection is an important bottleneck for modern biometric and surveillance systems. While conventional CNN and ViT models have demonstrated high accuracy on static images, the computational overhead associated with them renders them unfit for real-time edge devices dealing with continuous video streams. This review highlights that the integration of Spiking Neural Networks (SNNs) with event-based vision can lead to improved energy savings by processing only pixel-level changes related to occlusions. In addition, Liquid Neural Networks (LNNs) are highly robust to dynamic and unpredictable occlusions thanks to their ability to adapt parameters at runtime.

REFERENCES

[1] A. Nemavhola, C. Chibaya, and S. Viriri, "A Systematic Review of CNN Architectures, Databases, Performance Metrics, and Applications in Face Recognition," *Information*, vol. 16, no. 2, p. 107, Jan. 2025.

[2] C. Zhao, Y. Qin, and B. Zhang, "Adversarially Learning Occlusions by Backpropagation for Face Recognition," *Sensors*, vol. 23, no. 20, p. 8559, Oct. 2023.

[3] Y. Chen, S. Liu, and D. Zhao, "Occlusion facial expression recognition based on feature fusion residual attention network," *Front. Neurorobot.*, vol. 17, p. 1250706, Sep. 2023.

[4] J. Gao and Y. Zhao, "TFE: A Transformer Architecture for Occlusion Aware Facial Expression Recognition," *Front. Neurorobot.*, vol. 15, p. 763100, Oct. 2021.

[5] Y. Huang, J. Peng, and W. Zhang, "FERMixNet: An Occlusion Robust Facial Expression Recognition Model With Facial Mixing Augmentation and Mid-Level Representation Learning," *IEEE Trans. Affect. Comput.*, pp. 1–17, 2024.

[6] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Netw.*, vol. 10, no. 9, pp. 1659–1671, Nov. 1997.

[7] G. Gallego et al., "Event-Based Vision: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, Jan. 2022.

[8] J. Noroozi, E. Rezayat, and M. A. Dehaqani, "Frontotemporal network contribution to occluded face processing," *Proc. Natl. Acad. Sci.*, vol. 121, no. 48, p. e2407457121, Nov. 2024.

[9] S. Kim, S. Park, B. Na, and S. Yoon, "Spiking-YOLO: Spiking Neural Network for Energy-Efficient Object Detection," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, pp. 11270–11277, Apr. 2020.

[10] M. Ji, Z. Wang, and R. Yan, "SCTN: Event-based object tracking with energy-efficient deep convolutional spiking neural networks," *Front. Neurosci.*, vol. 17, p. 1123698, Feb. 2023.

[11] Y. Tao, C. Chang, and S. Saïghi, "GaitSpike: Event-based Gait Recognition With Spiking Neural Network," in *Proc. IEEE 6th Int. Conf. AI Circuits Syst. (AICAS)*, Abu Dhabi, UAE, 2024, pp. 357–361.

[12] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate Gradient Learning in Spiking Neural Networks," *IEEE Signal Process. Mag.*, vol. 36, no. 6, pp. 51–63, Nov. 2019.

[13] R. Hasani, M. Lechner, A. Amini, D. Rus, and R. Grosu, "Liquid Time-constant Networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 9, pp. 7657–7666, May 2021.

[14] W. Maass, "Liquid State Machines: Motivation, Theory, and Applications," in *Computability in Context: Computation and Logic in the Real World*, London, U.K.: Imperial College Press, 2011, pp. 275–296.

[15] D. Das, S. Bhattacharya, and U. Pal, "PLSM: A Parallelized Liquid State Machine for Unintentional Action Detection," *arXiv preprint arXiv:2105.09909*, 2021.