IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **Volume 11, Sp.Iss 6 , 2022**

Fusion of RGB and Skeletal Data Using Gated Features for Human Action Recognition

Ch.Raghava Prasad

Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Deemed to be University, Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India chrp@kluniversity.in,

DOI : 10.48047/IJFANS/11/S6/001

Abstract

This paper offers a blended network of RGB, depth, and skeleton inputs fed into CNNs in both directions. In order to learn the combined temporal features of the action, CNNs are used to characterize the RGB and depth data, while LSTMs are used to encode the skeletal data in both directions. At last, the L2 distance metric is used to choose the probability distribution generated from the three inputs. Coupling the model with a mixed CNN BILSTM network and computing an L2 distance measure in place of score fusion improved performance to 94.73%. Finally, the proposed models were compared to both cutting-edge deep learning methods and classic machine learning models.

1.Introduction

We present a method for automating the merging of the most important skeleton-based features with color space information. In contrast, traditional fusion approaches just combined these multi-modal elements without acquiring the necessary knowledge during the fusion process to fully take advantage of the semantic relationship between them. Based on the temporal skeletal data as a proposed model, we offer a gated feature fusion (GFF) of multi modal feature data that allows for the introduction of attention into the appearance stream of RGB data. At first, CNN models are used to extract features from RGB and skeleton frames. In the next step, the gated fusion network merges the temporal data from many sources into a single latent subspace. Finally, a fully connected layer is constructed with the combined loss embeddings, and this layer is used to categorize the features in the latent subspace. Real-time action detection models have a significant challenge due to the presence of irrelevant context data. Self-attention models, depth-attention, and skeletal-attention are how most prior works have dealt with this difficulty. The lack of attention to the semantic connection between the many modes of information is a major flaw in these works.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 6 , 2022

In the past, focusing on multi modal data required geographic matching on data from two sources. Aligning multi-modal data into a single space is a common technique for drawing in viewers [1,2]. In this paper, we suggest using temporal convolution networks (TCN) for feature fusion in order to give the appearance models more consideration based on the poses they are given. However, these models were unable to properly categorize data that was presented in one modality but not the other.

2.Methodology

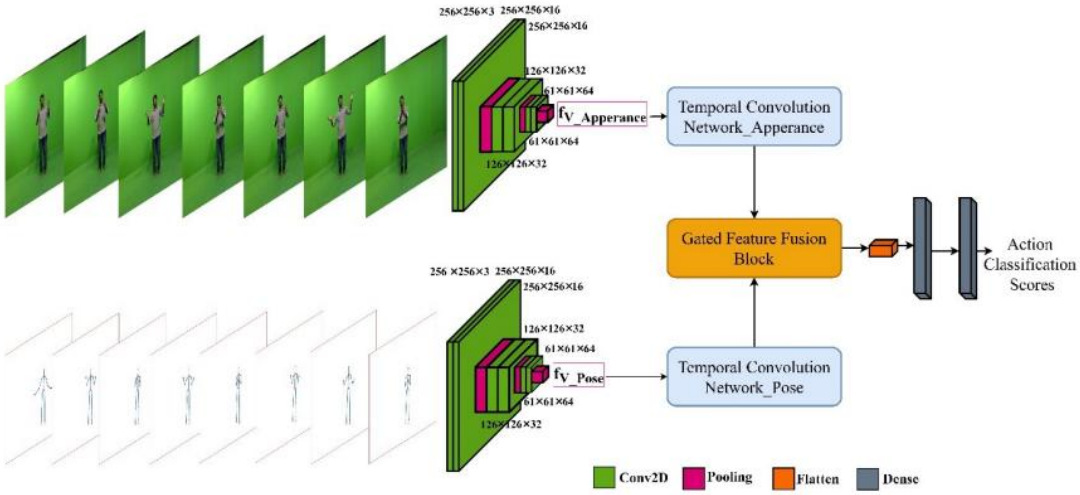

Fig. 1: Illustration of GaNet Framework for Action Classification.

Figure 1 depicts GaNet in its entirety. Two CNN streams are depicted in the figure.1: one for RGB video frames and the other for pose frames. Differences between the RGB appearance stream (SA) and the skeleton pose stream (SP) are described. At first, spatial features are extracted by the SA and SP streams at the conclusion of the final convolutional layers. The features are then fed into temporal convolution networks for modeling temporal relationships in appearance and pose features independently. After that, gated sigmoid functions are used in the GFF block to combine these spatial and temporal characteristics. In the end, the completely connected thick layers are tasked with learning the fused appearance-pose features to classify actions. The GFF component studies the pose distribution across the appearance characteristics and uses that knowledge to make informed feature selections for the purpose of action classification. In Figure 2 we see the GFF module in action.As shown in the picture.3, the TCN block receives the feature maps generated in the SA and SP streams.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper Volume 11, Sp.Iss 6 , 2022

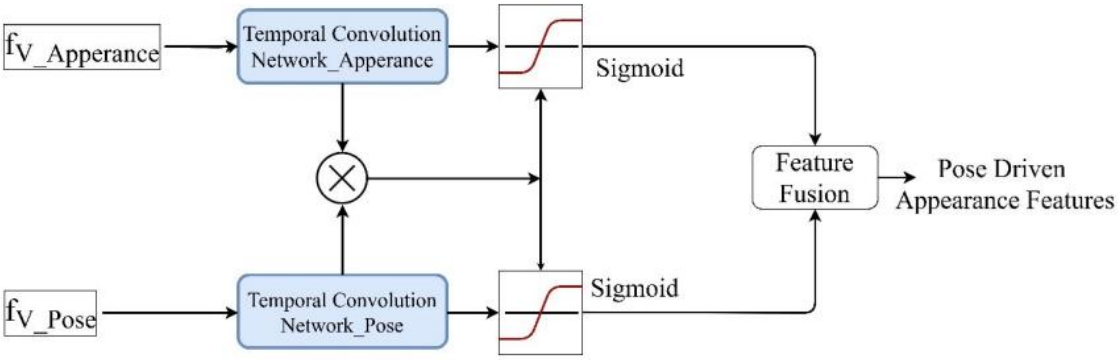

Fig. 2: The Gated Feature Fusion Module

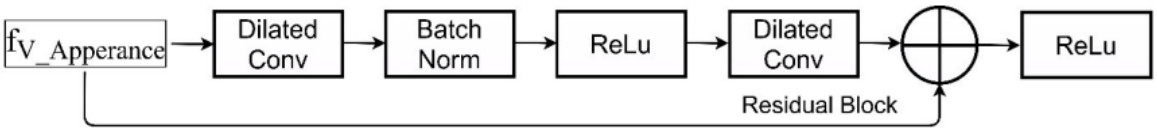

Fig. 3: Temporal Convolution Block

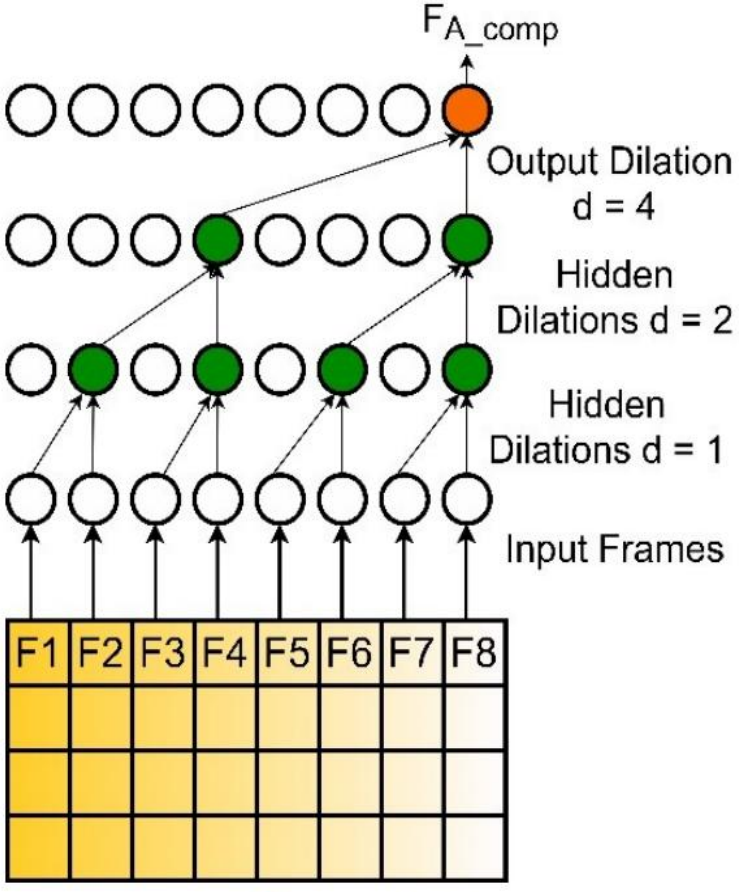

Fig. 4: Appearance TCN Process

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 6 , 2022

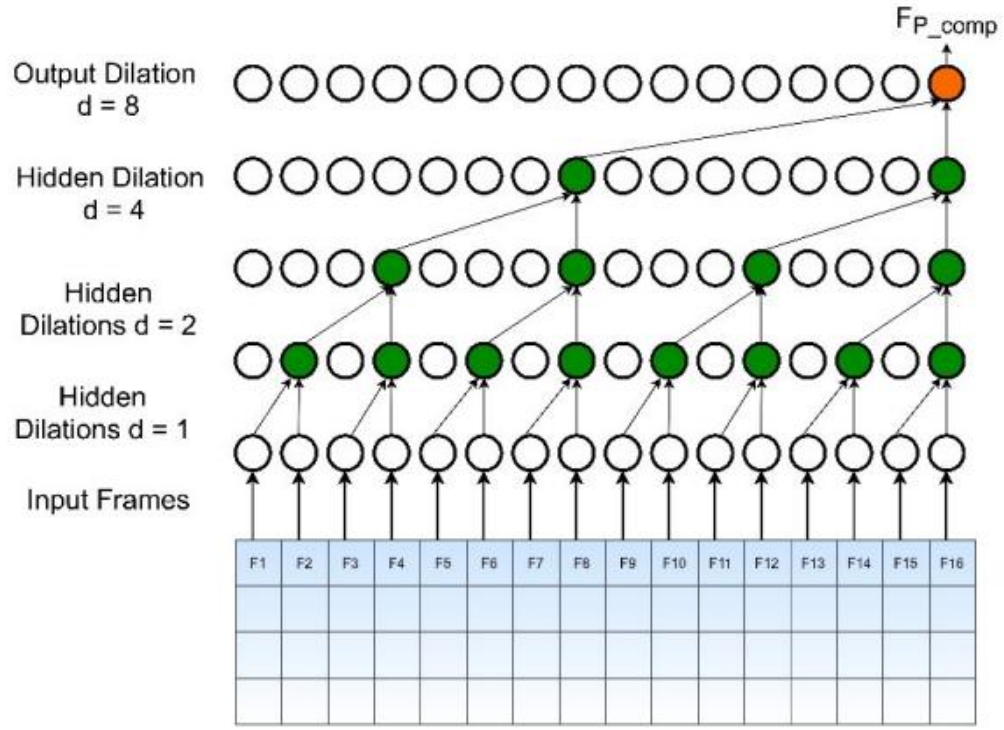

Fig. 5: Pose TCN Process

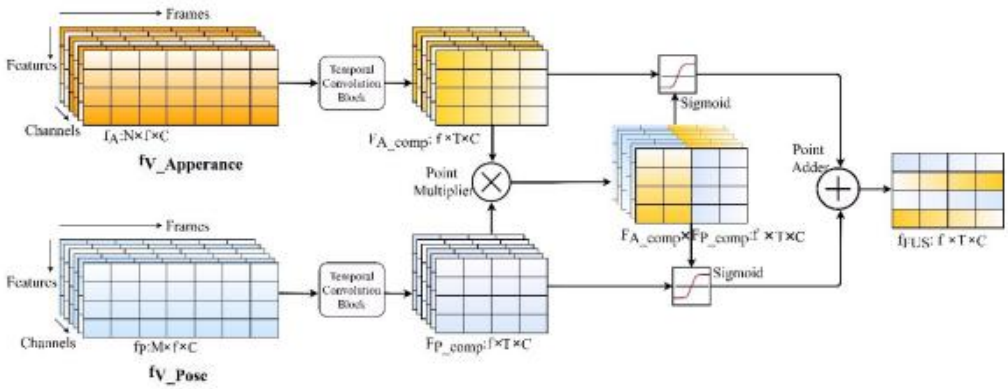

Fig. 6: Visual Illustration of GFF for action recognition

During the training procedure, we use two distinct TCNs with varying degrees of dilatation 'd' to perform temporal pooling and generate a relationship representative. The position TCN is depicted in figure.5 and the appearance TCN is depicted in figure.4.The qualities in both modalities can be highlighted by merely multiplying their elements together, without taking into account the affecting elements.To fix this issue, we used a sigmoid multiplier and gates to pick out relevant features.Figure 3 provides a graphical representation of the described procedure.6.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 6 , 2022**

3.Conclusions

To combine look and position features for human action identification, this paper presented a gated feature fusion with a temporal convolution network. Both the skeleton data and the RGB data are used to capture the appearance features and the stance, respectively. In the past efforts, the fusing procedure was applied in a very passive manner, without any constraints on the feature ensemble or consideration of the temporal relationships present in the video data. This work performed both the above operations.

References

1.T. Huynh-The, C.-H. Hua, T.-T. Ngo, and D.-S. Kim, "Image representation of pose-transition feature for 3d skeleton-based action recognition," Information Sciences, vol. 513, pp. 112–126, 2020.

2. H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "Attention branch network: Learning of attention mechanism for visual explanation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10 705–10 714.

3. M. A. Khan, M. Sharif, T. Akram, M. Raza, T. Saba, and A. Rehman, "Hand□ crafted and deep convolutional neural network features fusion and selection strat egy: an application to intelligent human action recognition," Applied Soft Comput□ ing, vol. 87, p. 105986, 2020.

4. J. Liu, A. Shahroudy, D. Xu, A. C. Kot, and G. Wang, "Skeleton-based action recognition using spatio-temporal lstm network with trust gates," IEEE transac tions on pattern analysis and machine intelligence, vol. 40, no. 12, pp. 3007–3021, 2017.

5. P. Zhang, C. Lan, J. Xing, W. Zeng, J. Xue, and N. Zheng, "View adaptive recur□ rent neural networks for high performance human action recognition from skeleton data," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2117–2126.

6. Y. Yoon, J. Yu, and M. Jeon, "Spatio-temporal representation matching-based open-set action recognition by joint learning of motion and appearance," IEEE Access, vol. 7, pp. 165 997–166 010, 2019.

7. Sundereshan G., Krishna K.V., Vineela R., Vyshnavi N. (2019), 'Dual band monopole antenna for UWB applications with added GSM and bluetooth bands', International Journal of Innovative Technology and Exploring Engineering, 8(6), PP.212-216.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 6 , 2022**

8. D.-O. Kim, N.-I. Jo, H.-A. Jang, and C.-Y. Kim, "Design of the Ultrawideband Antenna with a Quadruple-Band Rejection Characteristics Using a Combination of the Complementary Split Ring Resonators," Progress In Electromagnetics Research, Vol. 112, 93-107, 2011. doi:10.2528/PIER10111607.

9. Dimitris E. Anagnotou, Michael T. Chryssomallis, "Reconfigurable UWB antenna with RF-MEMs for on-demand WLAN Rejection", IEEE Transactions on Antennas and Propagation, Vol. 62, No. 2, Feb. 2014.

10. S. A. Aghdam, "Reconfigurable Antenna With a Diversity Filtering Band Feature Utilizing Active Devices for Communication Systems," in IEEE Transactions on Antennas and Propagation, vol. 61, no. 10, pp. 5223-5228, Oct. 2013.

11. N. Tasouji, J. Nourinia, C. Ghobadi and F. Tofigh, "A Novel Printed UWB Slot Antenna With Reconfigurable Band-Notch Characteristics," in IEEE Antennas and Wireless Propagation Letters, vol. 12, pp. 922-925, 2013.