Light-Weight Deep Learning Model for Human

Action Recognition in Videos

Human Action Recognition (HAR) from a visual stream has recently attained much researcher consideration in the domain of computer vision. Due to its large applications like monitoring of health, home automation, and teleimmersion, among others. However, it still faces human variances, occlusion, lighting changes, and complicated backgrounds. The evaluation criteria rely on the features collection approach as well as learning data being performed correctly. The success of Deep Learning (DL) has resulted in a variety of impressive outcomes, including neural networks. Nonetheless, a robust features vector is required for an efficient classifier to give the class label. Features serve as the essential component of any data set. Indeed, feature extraction may affect the algorithm's performance and computational cost. For this research framework, we used pre-trained deep learning models VGG19, Dense Net and Efficient Net for feature extraction from the sequence of images and classified each action with the help of the SoftMax layer. UCF50 action dataset used, which contains 50 sections and evaluates performance with the help of precision, recall, f1-score and AUC score. Testing accuracy from models achieved VGG19- 90.11, DenseNet-92,57 and EfiicientNet-94.25

**Existing Systems** :

**DenseNet**: A convolutional neural network known for its extensive interconnectedness, allowing each layer to receive inputs from all previous layers and broadcast its feature maps to all subsequent layers. This model helps address vanishing or growing gradients and allows feature reuse​(Light-Weight\_Deep\_Learn…)​.

**VGG (Visual Geometry Group Network)**: This CNN architecture is characterized by its very deep architecture, using small 3x3 filters. The paper specifically mentions VGG19, which includes convolutional layers, pooling layers, and dense layers with full connections followed by a SoftMax layer for classification​(Light-Weight\_Deep\_Learn…)​.

**EfficientNet**: Known for its efficient scaling method, which balances depth, width, and resolution of the network using a compound scaling coefficient. EfficientNet is included in the paper to evaluate the performance of the HAR framework​(Light-Weight\_Deep\_Learn…)​.

**Machine Learning (ML) Methods**: The paper contrasts traditional ML methods such as random forests, Bayesian networks, Markov models, and support vector machines with deep learning methods for HAR. These traditional methods require extensive data preprocessing and are less effective with large datasets compared to deep learning models​(Light-Weight\_Deep\_Learn…)​.

**Transfer Learning Models**: The paper utilizes transfer learning by employing pre-trained models on large datasets such as ImageNet. This approach leverages previously acquired knowledge to train new models in a more efficient manner​(Light-Weight\_Deep\_Learn…)​.

**Advantages** :

**Reduced Computational Complexity**: The proposed model is designed to be light-weight, meaning it requires less computational power and memory compared to traditional deep learning models. This makes it more suitable for deployment on devices with limited resources, such as mobile devices and embedded systems​(Light-Weight\_Deep\_Learn…)​.

**High Accuracy**: The model achieves high accuracy in recognizing human actions. In particular, the EfficientNet model achieves an accuracy of 94.25% on the UCF50 dataset, outperforming other models like DenseNet and VGG19​(Light-Weight\_Deep\_Learn…)​.

**Efficiency in Training**: By utilizing transfer learning, the proposed system can leverage pre-trained models on large datasets such as ImageNet. This approach reduces the training time and the amount of labeled data needed to achieve high performance​(Light-Weight\_Deep\_Learn…)​.

**Robustness to Variations**: The model demonstrates robustness to variations in video data, such as changes in background, lighting, and viewpoint. This is achieved through the use of advanced neural network architectures and techniques like batch normalization and dense connections​(Light-Weight\_Deep\_Learn…)​.

**Feature Reuse and Gradient Handling**: The DenseNet architecture employed in the proposed model allows for extensive feature reuse and effective handling of vanishing or growing gradients, which are common issues in deep learning​(Light-Weight\_Deep\_Learn…)​.

**Scalability**: The model's design ensures scalability, making it capable of handling larger datasets and more complex action recognition tasks without a significant increase in computational requirements​(Light-Weight\_Deep\_Learn…)​.

**Disadvantages:**

**Computational Requirements**: Although the proposed model is lighter compared to traditional deep learning models, training deep learning models from scratch still requires significant computational resources. This includes high-performance GPUs and considerable training time, especially when dealing with large datasets​(Light-Weight\_Deep\_Learn…)​.

**Data Dependency**: The effectiveness of the model relies heavily on the availability of large labeled datasets. Transfer learning mitigates this to some extent by using pre-trained models, but achieving the highest accuracy still depends on the quality and quantity of the data available for fine-tuning the model​(Light-Weight\_Deep\_Learn…)​.

**Generalization Issues**: While the model achieves high accuracy on specific datasets like UCF50, its performance might not generalize well to other datasets or real-world scenarios where the conditions differ significantly. Variations in lighting, background, and camera angles can affect the model's performance​(Light-Weight\_Deep\_Learn…)​.

**Complexity of Implementation**: Implementing and fine-tuning deep learning models can be complex and require expertise in machine learning and neural network design. This includes knowledge in selecting appropriate architectures, managing overfitting, and optimizing hyperparameters​(Light-Weight\_Deep\_Learn…)​.

**Maintenance and Updates**: Keeping the model updated with new data and techniques is necessary for maintaining its performance over time. This requires ongoing effort and resources, which can be challenging for applications with limited budgets or technical support​(Light-Weight\_Deep\_Learn…)​.

**Algorithms:**

3D Convolutional Neural Networks (3D CNNs)

Stacked Denoising Auto-Encoders

**Proposed Systems:**

**VGG19**: A convolutional neural network (CNN) model known for its deep architecture and use of small filters (3x3 pixels). VGG19 is enhanced with batch normalization layers to control gradient explosions.

**DenseNet**: A type of CNN that interconnects each subsequent layer in a feed-forward fashion, facilitating feature reuse and addressing issues related to vanishing or growing gradients.

**EfficientNet**: A CNN architecture that uses a compound scaling method to adjust the network's depth, width, and resolution uniformly, improving parameter efficiency and speed.

**Advantages:**

**Feature Reuse and Gradient Handling (DenseNet)**:

* **DenseNet** addresses the issues of vanishing or growing gradients through extensive interconnections between layers, allowing feature reuse which enhances learning efficiency and performance.

**Efficient Training and Generalization (VGG19)**:

* **VGG19** benefits from a deep architecture using small filters, which improves the model’s ability to learn intricate patterns. The incorporation of batch normalization helps to control gradient explosions, leading to more stable and efficient training.

**Scalable and Parameter Efficient (EfficientNet)**:

* **EfficientNet** employs a compound scaling method that uniformly scales network depth, width, and resolution. This approach significantly enhances parameter efficiency and speed, making it more suitable for real-world applications where computational resources may be limited.

**High Accuracy**:

* The use of pre-trained models such as VGG19, DenseNet, and EfficientNet leads to high accuracy in human action recognition tasks. EfficientNet, in particular, achieved a testing accuracy of 94.25%, demonstrating its effectiveness in classifying human actions from video data.

**Transfer Learning**:

* By leveraging pre-trained models, the proposed systems benefit from prior training on large datasets like ImageNet. This reduces the need for extensive computational resources and time to train models from scratch, enabling faster deployment and better generalization to new tasks.

**Robust Performance in Varied Conditions**:

* The models show robustness in handling common challenges in human action recognition, such as occlusion, lighting changes, and complex backgrounds, leading to reliable performance in diverse environments.

**Application Versatility**:

* The high performance of these models makes them suitable for various applications, including health monitoring, home automation, and tele-immersion, where accurate human action recognition is critical.

**Algorithms:**

Convolutional Neural Network (CNN) for VGG19

Dense Convolutional Network (DenseNet) for DenseNet

Efficient Convolutional Network (EfficientNet) for EfficientNet

SoftMax Function for classification

**SYSTEM SPECIFICATION:**

**HARDWARE REQUIREMENTS:**

* **System :** Intel i7
* **Hard Disk :** 1 TB.
* **Monitor** : 14’ Colour Monitor.
* **Mouse :** Optical Mouse.
* **Ram :** 8GB.

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 10.
* **Coding Language :** Python.
* **Front-End :** Html. CSS
* **Designing :** Html,css,javascript.
* **Data Base :** SQLite.

**REFERENCES**

[1] P. Pareek and A. Thakkar, “A survey on video-based Human Action

Recognition: recent updates, datasets, challenges, and applications,”

Artif Intell Rev, vol. 54, no. 3, pp. 2259–2322, Mar. 2021, doi:

10.1007/s10462-020-09904-8.

[2] P. K. Singh, S. Kundu, T. Adhikary, R. Sarkar, and D. Bhattacharjee,

“Progress of Human Action Recognition Research in the Last Ten

Years: A Comprehensive Survey,” Archives of Computational

Methods in Engineering, vol. 29, no. 4, pp. 2309–2349, Jun. 2022,

doi: 10.1007/s11831-021-09681-9.

[3] A. Ladjailia, I. Bouchrika, H. F. Merouani, N. Harrati, and Z.

Mahfouf, “Human activity recognition via optical flow: decomposing

activities into basic actions,” Neural Comput Appl, vol. 32, no. 21,

pp. 16387–16400, Nov. 2020, doi: 10.1007/s00521-018-3951-x.

[4] K. Simonyan and A. Zisserman, “Two-Stream Convolutional

Networks for Action Recognition in Videos.”

[5] S. Ji, W. Xu, M. Yang, and K. Yu, “3D Convolutional neural

networks for human action recognition,” IEEE Trans Pattern Anal

Mach Intell, vol. 35, no. 1, pp. 221–231, 2013, doi: 10.1109/TPAMI.2012.59.

[6] F. Gu, K. Khoshelham, and S. Valaee, “Locomotion activity

recognition: A deep learning approach,” in IEEE International

Symposium on Personal, Indoor and Mobile Radio Communications,

PIMRC, Feb. 2018, vol. 2017-October, pp. 1–5. doi:

10.1109/PIMRC.2017.8292444.