Vision Based System for Fall Detection

Parth Shrivastava

12219466

Abstract

Falls are one of the leading causes of injury, especially among the elderly population, and timely detection is critical to minimize health risks and enable prompt assistance. This project presents a deep learning-based image classification approach for automatic fall detection using a Convolutional Neural Network (CNN) and transfer learning with VGG16. The goal is to develop a robust and efficient model capable of distinguishing between "fall" and "no fall" scenarios from still images. A publicly available fall detection dataset was used, comprising images labeled accordingly.

The project begins with data preprocessing, including image resizing and normalization, followed by data labeling based on filename patterns (e.g., images starting with 'f' indicating a fall). The dataset was split into training and validation subsets to train the models effectively. A baseline CNN was developed from scratch and trained to learn distinguishing features from the dataset. To enhance accuracy and reduce training time, transfer learning was applied using the VGG16 model, leveraging pre-trained weights from ImageNet. The final model architecture includes global average pooling and dense layers optimized with dropout regularization to prevent overfitting.

The trained model was evaluated on a validation dataset, achieving promising accuracy in classifying fall vs. non-fall scenarios. Furthermore, the system was extended to take any custom input image and predict whether it represents a fall or not, showcasing the model's real-world applicability. This solution demonstrates the potential of deep learning in enhancing health and safety monitoring systems, and it can be integrated into smart surveillance or wearable devices for real-time fall detection and alerting.

Keywords

Fall Detection, Convolutional Neural Networks, Transfer Learning, VGG16, Image Classification, Computer Vision, Healthcare Technology.

Literature Review

Overview of Fall Detection Methods

Fall detection has been extensively studied using various approaches such as wearable sensors, ambient sensors, and vision-based systems. Vision-based methods have gained prominence due to their non-intrusive nature and compatibility with smart environments. Below are notable studies in this domain:

1. **Vision-Based Fall Detection Using CNNs**
   Núñez-Marcos et al. (2017) proposed a vision-based solution using CNNs to detect falls from video sequences. Optical flow images were used as input to model motion dynamics effectively. Their method achieved state-of-the-art results on three public datasets by employing a three-step training phase[1].

2. **Gait Analysis for Fall Prediction**
   Aqueveque et al. (2022) analyzed spatio-temporal gait parameters using SVM and Random Forest classifiers to predict fall risk among elderly individuals. The study demonstrated promising results but highlighted limitations such as small sample size and reliance on external databases[2].

3. **Machine Learning Models for Fall Risk Assessment**
   Velusamy et al. (2023) explored wearable sensors combined with machine learning algorithms like SVMs, Random Forests, and neural networks for fall prediction. They achieved high accuracy scores but faced challenges related to sensor placement and dependability[2].

4. **Deep Learning for Fall Detection**
   Philip et al. (2023) introduced a CNN-RNN architecture to detect falls using sensor data from daily living activities (ADLs). Their model achieved an accuracy of 95%, demonstrating the effectiveness of combining spatial and temporal feature extraction[2].

5. **Comprehensive Review of Vision-Based Systems**
   A review by MDPI (2023) summarized recent advancements in vision-based fall detection systems, highlighting the use of depth maps, optical flow images, and CNN architectures[3][4].

Methods Used in This Study

This study employs two key methodologies for image-based fall detection: a custom Convolutional Neural Network (CNN) and transfer learning using the VGG16 model. These approaches aim to compare and evaluate performance differences between a from-scratch model and a fine-tuned pre-trained model for the binary classification of fall vs. no-fall scenarios :

1. **Custom CNN Architecture**
   The first method involves the development of a custom Convolutional Neural Network (CNN) architecture, inspired by the work of Núñez-Marcos et al. [1], which demonstrated the feasibility of detecting human falls using spatial-temporal CNNs. In our case, we focus on spatial information derived from single-frame RGB images. The architecture begins with several convolutional layers that apply filters to capture local patterns such as edges, textures, and human body contours. Each convolutional block is followed by a max-pooling layer, which down-samples the feature maps and helps in reducing computational complexity and overfitting. ReLU activation functions are used throughout the network to introduce non-linearity. After the convolutional layers, fully connected dense layers are used to interpret the extracted features and classify them into binary categories (fall or no fall). Dropout regularization is also employed to avoid overfitting by randomly disabling neurons during training. This architecture provides a solid baseline for evaluating the model's ability to learn features specific to fall incidents.[1].

2. **Transfer Learning with VGG16**
   To enhance accuracy and take advantage of transfer learning, we employed the VGG16 architecture, a deep CNN model pre-trained on the ImageNet dataset, which contains over 14 million annotated images across 1,000 categories. VGG16 has proven highly effective in various image classification tasks due to its uniform

structure and depth. In this study, we load the VGG16 model with pre-trained weights and exclude its top classification layers, using it solely as a feature extractor. The extracted features from the final convolutional blocks are flattened using Global Average Pooling, then passed through custom dense layers tailored for the binary fall detection task. Only the newly added dense layers are trainable, while the convolutional base remains frozen in most experiments to preserve the learned features. This approach significantly reduces training time and enhances performance, especially on relatively smaller datasets, by leveraging generalized visual patterns already learned by the VGG16 network.[1][3].

By combining both methods—training a custom CNN from scratch and leveraging transfer learning—the study compares the trade-offs in terms of model complexity, accuracy, generalization capability, and suitability for deployment in real-time fall detection systems.

Methodology

1. Dataset Preparation

- **Data Acquisition**: This is a custom fall detection dataset with two directories of images and labels. Images directories consist of two subdirectories train (374 images) which is used for training and Val (111 images) for validation. Labels directory consists of two subdirectories train and Val here in this directory we have text files with labels of that particular image.
- **Preprocessing Steps:**
    - *Resizing*: Images were resized to 150x150 pixels to standardize input dimensions and reduce computational complexity.
    - *Normalization*: Pixel values were scaled between 0 and 1 to stabilize gradient descent during training.
    - *Bounding Box Parsing*: Labels were decoded to extract object coordinates for visualization purposes.

2. Model Development

Custom CNN Architecture

- Layers: Three convolutional blocks with filters (16, 32, 64), ReLU activation functions, and max-pooling layers.
- Fully connected layers: A dense layer with 128 neurons followed by a sigmoid output layer for binary classification.
- Compilation: Adam optimizer with binary cross-entropy loss function.

Transfer Learning with VGG16

- Pre-trained Base Model: VGG16 trained on ImageNet was used as a feature extractor by freezing its convolutional layers.
- Custom Head: Added layers include flattening, dense (256 neurons), dropout (0.5), and sigmoid activation for binary classification.
- Fine-Tuning: The dense layers were trained for five epochs to adapt features to the fall detection task.

3. Training Process

- Training Data: The custom CNN was trained on the preprocessed dataset for 20 epochs.
- Transfer Learning: The VGG16 model was fine-tuned for five epochs using the same dataset.

4. Evaluation Metrics

Accuracy and loss values were used as primary metrics:

- *Accuracy*: Measures the proportion of correct predictions over total samples.
- *Loss*: Represents the error between predicted values and actual labels; lower loss indicates better model fit.

Results

Custom CNN Performance

The custom CNN achieved:

- Accuracy: 58.71%
- Loss: 2.5117

These results indicate that the custom CNN struggled to generalize well on the validation dataset due to limited feature extraction capabilities.

Transfer Learning with VGG16 Performance

The VGG16 model achieved:

- Accuracy: 86.73%
- Loss: 1.3581

These results highlight the effectiveness of transfer learning in improving classification accuracy by leveraging pre-trained features.

Interpretation of Metrics

1. *Accuracy*: Indicates how well the model differentiates between falls and non-falls; higher accuracy reflects better predictive performance.
2. *Loss*: Represents the model's error during prediction; lower values suggest improved fit between predictions and ground truth labels.

Conclusion

This research emphasizes the potential of deep learning models in identifying falls through image-based data analysis. By employing a custom-designed Convolutional Neural Network (CNN) alongside a transfer learning approach utilizing the VGG16 architecture, the study observed notable improvements in detection accuracy—particularly with the use of the pre-trained VGG16 model. These results underscore the viability of camera-based solutions for real-time fall monitoring, which could be implemented within smart home systems or healthcare institutions to enhance surveillance and ensure safety.

Major Findings

- **Performance of Custom CNN:**
  The CNN model built from scratch served as a foundational benchmark, proving that fall detection is feasible without pre-trained networks. However, the model encountered limitations in capturing complex features, which impacted its overall accuracy.
- **Effectiveness of VGG16 Transfer Learning:**
  The VGG16 model, fine-tuned for the task, achieved a validation accuracy of **86.73%**, significantly outperforming the custom CNN. This highlights the advantage of using deep, pre-trained architectures, particularly when working with datasets that lack extensive diversity or volume.
- **Real-World Relevance:**
  Vision-based fall detection systems provide a non-invasive alternative to wearable sensors. This is especially useful for elderly users who may find wearables uncomfortable or intrusive. Moreover, integration with existing camera infrastructure makes this solution both scalable and economically efficient.

Future Work

- **Dataset Enhancement:**
  Increasing the dataset's diversity—by including various lighting conditions, camera angles, and environmental settings—can help improve the model's robustness and generalization across real-life scenarios.
- **Adoption of More Sophisticated Models:**
  Future experiments could incorporate other state-of-the-art models like ResNet, EfficientNet, or MobileNet to further boost performance and reduce inference time.
- **Deployment for Real-Time Detection:**
  Implementing the trained models on edge devices such as smart surveillance cameras or embedded systems could enable real-time detection and rapid alerting, thereby enhancing emergency response times in critical situations.

Summary

In conclusion, this work adds meaningful contributions to the field of fall detection through computer vision, showing that deep learning—especially when combined with transfer learning—offers a reliable and scalable approach. The study's insights into model selection, practical deployment, and the advantages of non-intrusive monitoring pave the way for the integration of

AI-powered safety systems in homes and healthcare settings. These findings also open avenues for future improvements that could push the boundaries of accuracy, speed, and real-world applicability.

References

1. Núñez-Marcos, A., Azkune, G., & Arganda-Carreras, I. (2017).
Vision-based fall detection with convolutional neural networks. *International Journal of Distributed Sensor Networks, 13*(12), Article ID 9474806.
https://onlinelibrary.wiley.com/doi/10.1155/2017/9474806

2. Mubashir, M., Shao, L., & Seed, L. (2013).
A survey on fall detection: Principles and approaches. *Neurocomputing, 100*, 144–152.
https://pubmed.ncbi.nlm.nih.gov/24945676/

3. Brodowski, J., Huber, S., Bollheimer, L. C., & Eskofier, B. M. (2023).
Comparison of machine learning approaches for near-fall detection in older people based on wearable sensor data. *Frontiers in Digital Health, 5*, Article 1223845.
https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2023.1223845/full

4. 4. Casilari, E., Oviedo-Jiménez, M. A., & García-Lagos, F. J. (2021).
Comprehensive review of vision-based fall detection systems. *Sensors, 21*(3), 947.
https://pmc.ncbi.nlm.nih.gov/articles/PMC7866979/