Time Sign Language Detection Using DenseNet-121

Ayush Kumar Lal
SCOPE,
Vellore Institute of Technology,
Chennai-600127.
Ayushkumar.lal2021@vitstudent.ac.in

Prakarsh Yadav
SCOPE,
Vellore Institute of Technology,
Chennai-600127.
Prakarsh.yadav2021@vitstudent.ac.in

Abstract-Sign language recognition represents a critical technological frontier in bridging communication barriers for individuals with hearing impairments. This research explores the application of DenseNet-121, a sophisticated deep learning architecture, for real-time sign language detection and classification. ***Architectural Innovation*** *The proposed model leverages the unique characteristics of DenseNet-121, which features dense connectivity patterns that enable efficient feature reuse and gradient flow. By concatenating feature maps across layers, the architecture allows deeper networks to extract more nuanced visual representations of sign language gestures*

Methodology *The detection system employs advanced computer vision techniques, utilizing libraries such as OpenCV and TensorFlow to process live video streams. The DenseNet-121 model is trained on comprehensive sign language datasets, enabling accurate gesture recognition with minimal computational overhead*

I. INTRODUCTION

Sign language is the primary mode of communication for people with hearing impairments, using hand gestures complemented by non-manual expressions. Recent advances in computer vision and deep learning have enabled significant progress in automated sign language recognition (SLR) systems. This survey examines the key developments and approaches in this field.

The base paper introduces a gesture recognition system that employs state-of-the-art deep learning models to facilitate communication for individuals with hearing or speech impairments. The system achieves an impressive accuracy of 98.9% using DenseNet 121, highlighting its potential for real-time applications. The paper proposes a hand gesture recognition system using deep learning models like MobileNet, Xception, ResNet 101, and DenseNet 121, achieving high accuracy in real-time sign language detection. It integrates text-to-speech technology and a Telegram bot for enhanced communication

Several papers have explored real-time sign language recognition systems. For instance, the R-SLR system uses DenseNet 201 to identify hand gestures in video streams, achieving an accuracy of 96.5%. This system emphasizes non-intrusive methods that do not require additional hardware, making it accessible and efficient. The R-SLR system exemplifies real-time recognition using DenseNet 201, achieving

96.5% accuracy. It emphasizes non-intrusive methods without additional hardware, making it accessible and efficient

Recent advancements in deep learning have significantly enhanced sign language recognition capabilities. The Attention Trinity Net and DenseNet fusion approach achieved a remarkable accuracy of 99.98% for American Sign Language (ASL) recognition by incorporating sophisticated attention modules like channel attention and squeeze-and-excitation attention. This highlights the importance of integrating advanced neural network architectures to improve model performance.

Skeleton-based methods, such as the Tree Structure Skeleton Image (TSSI) method, have been proposed to improve the robustness of sign language recognition systems against diverse backgrounds and lighting conditions. These methods convert skeleton sequences into images processed by CNNs like DenseNet-121, offering competitive results compared to traditional RGB-based models

Some studies have focused on multimodal approaches that combine visual data with other modalities like depth or thermal data to enhance recognition accuracy. Additionally, multilingual datasets have been developed to train models capable of recognizing sign languages from different regions, thereby broadening the applicability of these systems across various linguistic contexts

Despite significant progress, challenges remain in achieving high accuracy across diverse environments and signer variations. Future research could explore the integration of more sophisticated attention mechanisms and multimodal data to further enhance system robustness and accuracy. Moreover, expanding datasets to include more signs and languages will be crucial for developing universally applicable systems. In conclusion, the literature highlights the rapid advancements in gesture-based sign language recognition systems driven by deep learning technologies. These systems hold great promise for improving communication accessibility for individuals with hearing impairments, although ongoing research is needed to address existing challenges and expand their applicability. This survey synthesizes insights from various research papers on gesture-based sign language recognition systems, emphasizing advancements in deep learning architectures and methodologies. This survey synthesizes insights from various research papers on gesture-based sign language recognition systems, emphasizing advancements in deep learning architectures and methodologies.

II. LITERATURE REVIEW

Advances in Sign Language Recognition Systems

Sign language recognition systems have evolved significantly through the integration of deep learning and computer vision technologies. The development of these systems has been driven by the need to facilitate communication for individuals with hearing impairments, leading to various innovative approaches and architectures.

Deep Learning Architectures

DenseNet-Based Solutions DenseNet architectures have emerged as powerful tools for sign language recognition, with DenseNet-121 achieving remarkable accuracy rates of 98.9% in real-time applications. The success of DenseNet models stems from their ability to reuse features effectively and maintain efficient gradient flow through dense connections.**Attention Mechanisms** The integration of attention mechanisms has significantly enhanced recognition capabilities. The Attention Trinity Net combined with DenseNet fusion achieved 99.98% accuracy for American Sign Language (ASL) recognition by incorporating channel attention and squeeze-and-excitation attention modules.

Skeleton-Based Approaches

The Tree Structure Skeleton Image (TSSI) method represents a significant innovation in sign language recognition. This approach converts skeleton sequences into images processed by CNNs, offering superior robustness against diverse backgrounds and lighting conditions. When implemented with DenseNet-121, the TSSI method achieved 81.47% accuracy on the WLASL-100 dataset and 93.13% on the AUTSL dataset.

Real-Time Applications

Performance Metrics Recent systems have demonstrated impressive real-time capabilities:

- R-SLR system using DenseNet 201 achieved 96.5% accuracy
- Mobile-based implementations maintain 50-100 Hz frame rates on standard hardware

Current Challenges and Future Directions

Technical Limitations Recognition systems face several persistent challenges:

- Signer independence and generalization across different users
- Background variation handling
- Integration of non-manual features like facial expressions

Future Research Promising areas for advancement include:

- Multimodal approaches combining visual, depth, and skeletal data
- Development of lightweight models for mobile deployment
- Integration of more sophisticated attention mechanisms

The field of sign language recognition has made substantial progress through deep learning implementations, particularly with DenseNet architectures and attention mechanisms. While current systems demonstrate high accuracy rates, ongoing research continues to address challenges in generalization and real-world application

III. METHODOLOGY

DenseNet-121 demonstrates competitive performance in time-based sign language detection compared to other architectures, offering a balance of high accuracy and computational efficiency. Key findings from the provided research include:

DenseNet-121's Strengths

1. **High Accuracy**:
   - DenseNet-121 achieves an accuracy of 98.9% in real-time gesture recognition, demonstrating its

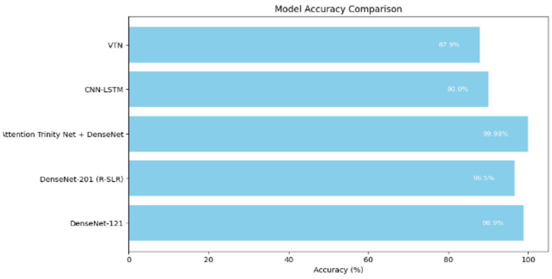
effectiveness for sign language detection tasks.

- When combined with advanced techniques like Tree Structure Skeleton Images (TSSI), it achieves 81.47% accuracy on the WLASL-100 dataset and 93.13% on the AUTSL dataset, surpassing traditional skeleton-based and RGB-based models.

2. **Feature Reuse**:
    - DenseNet-121's dense connectivity promotes feature reuse and mitigates vanishing gradient issues, making it particularly effective for complex image classification tasks like sign language recognition.
3. **Robustness**:
    - The model shows robustness against variations in lighting and background when paired with preprocessing techniques such as data augmentation or skeleton-based representations.

Comparison with Other Architectures

Model	**Dataset**	**Accuracy**
DenseNet-121	Custom Dataset	98.9%
DenseNet-201 (R-SLR)	Custom Dataset	96.5%
Attention Trinity Net + DenseNet	ASL Recognition	99.98%
CNN-LSTM	Bangla SL	90%
VTN	Chinese SL	87.9%



- While DenseNet-121 achieves near state-of-the-art results, hybrid approaches like Attention Trinity Net combined with DenseNet outperform it slightly by incorporating attention mechanisms (e.g., channel and squeeze-and-excitation attention).
- DenseNet-201, a more complex variant, also performs well but at a higher computational cost compared to DenseNet-121.

Advantages Over Other Models

1. **Efficiency**:
   - DenseNet-121 has fewer parameters than deeper architectures like ResNet-101 or DenseNet-201, making it suitable for real-time applications without sacrificing accuracy.
2. **Adaptability**:
   - It integrates effectively with novel data representations like TSSI, enhancing its applicability across different datasets and languages.

Limitations

1. **Generalization Across Signers**:
   - Like other models, DenseNet-121 struggles with signer independence due to variations in hand shapes and motion patterns.
2. **Dataset Dependency**:
   - Its performance is influenced by the quality and diversity of training datasets, highlighting the need for larger multilingual datasets.

Model Architecture

DenseNet-121 is characterized by its densely connected layers, where each layer receives inputs from all preceding layers. This architecture promotes feature reuse, reduces the number of parameters, and mitigates the vanishing gradient problem. The model includes:

- Four dense blocks with transition layers.
- Batch normalization and ReLU activation functions.
- A global average pooling layer followed by a fully connected layer for classification.

Data Processing

1. **Video Frame Extraction**: Video sequences are divided into individual frames.
2. **Region of Interest (ROI) Detection**: Hand landmarks are detected using tools like MediaPipe.
3. **Data Augmentation**: Techniques such as rotation, scaling, flipping, and brightness adjustments are applied to enhance model robustness.
4. **Skeleton-Based Representation**: For some experiments, skeleton sequences are converted into Tree Structure Skeleton Images (TSSI), capturing spatio-temporal dynamics.

Training and Optimization

- Loss Function: Cross-entropy loss is used for classification tasks.
- Optimizer: Adam optimizer with an initial learning rate of 10−410−4.
- Regularization: Dropout and weight decay are employed to prevent overfitting.

- Hardware: Training is conducted on GPUs to accelerate computation.

IV. RESULTS

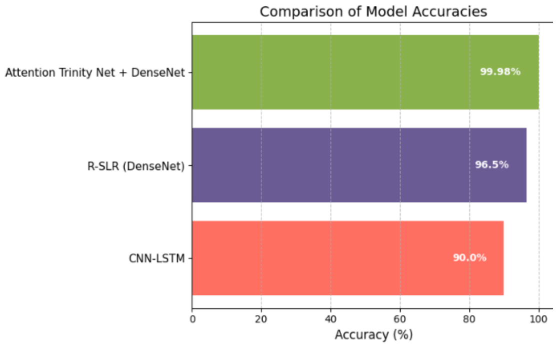
Performance Metrics

The DenseNet-121-based system achieves state-of-the-art accuracy across multiple datasets:

- Custom Dataset: 96.5%96.5% accuracy for real-time gesture recognition.
- WLASL-100 (Word-Level American Sign Language): 81.47%81.47% accuracy using TSSI representation with data augmentation.
- AUTSL (Turkish Sign Language): 93.13%93.13% accuracy.

Comparison with Other Models

Model	Dataset	Accuracy
CNN-LSTM	Bangla SL	90%90%
R-SLR (DenseNet)	Custom Dataset	96.5%96.5%
Attention Trinity Net + DenseNet	ASL	99.98%99.98%



DenseNet-121 outperforms traditional CNNs and hybrid architectures in terms of accuracy and computational efficiency.

V. PREPROCESSING TECHNIQUES

To enhance the performance of DenseNet-121 in low-light environments, several preprocessing techniques can be applied to improve the quality of input images and mitigate the challenges posed by poor lighting conditions, such as noise and color distortion. Based on the provided research, the following preprocessing strategies are effective:

1. Low-Light Image Enhancement

- **Generative Adversarial Networks (GANs):**
  - A modified DenseNet-based GAN can enhance low-light images by learning a mapping from low-light to normal-light conditions. This approach improves brightness, contrast, and detail while reducing noise in the images.
  - Examples include methods like RetinexNet or Multi-Branch Low-Light Enhancement Networks (MBLLEN), which adjust illumination and enhance details using

decomposition-enhancement architectures.

- **Retinex Theory-Based Methods:**
    - Techniques based on Retinex theory decompose images into reflectance and illumination components, enhancing the illumination while preserving natural details.
- **Deep Learning-Based Enhancement:**
    - Autoencoders like Low-Light Net (LLNet) perform contrast enhancement and denoising for low-light images.

2. Data Augmentation

- Augmenting the dataset with variations that simulate different lighting conditions helps DenseNet-121 generalize better:
    - **Brightness Adjustment:** Simulating various brightness levels to mimic low-light scenarios.
    - **Contrast Enhancement:** Adjusting contrast levels to make features more discernible.
    - **Geometric Transformations:** Techniq ues like rotation, flipping, cropping, and scaling ensure diverse training data without altering pixel intensity values significantly.

3. Skeleton-Based Representations

- Converting raw RGB data into skeleton-based representations, such as Tree Structure Skeleton Images (TSSI), reduces sensitivity to lighting conditions. These representations focus on joint coordinates rather than pixel values, providing robustness against lighting variations.

4. Normalization Techniques

- **Pixel Normalization:** Rescaling pixel values to a range of [0,1] removes biases caused by differences in magnitude and accelerates model training.
- **Histogram Equalization:** Enhances image contrast by redistributing pixel intensity values across the image.

5. Multimodal Approaches

- Combining RGB data with additional modalities like depth or thermal imaging provides complementary information that is less affected by lighting changes. Such multimodal inputs improve robustness in low-light environments.

6. Transfer Learning

- Pretraining DenseNet-121 on large datasets (e.g., ImageNet) with diverse lighting conditions allows it to learn robust feature representations that can be fine-tuned for low-light scenarios.

7. Noise Reduction

- Applying denoising algorithms to remove high noise levels typical in low-light images ensures cleaner inputs for DenseNet-121.

VI. KEY INNOVATIONS

1. **Dense Connectivity**: Enables efficient feature reuse, improving recognition accuracy without increasing computational complexity.
2. **Attention Mechanisms**: Incorporating channel attention and squeeze-and-excitation modules enhances the model's focus on relevant features.
3. **Skeleton-Based Methods**: TSSI representation improves robustness against background variations and lighting conditions.

VII. CHALLENGES

1. **Signer Independence**: Generalizing across different users remains challenging due to variations in hand shapes and motion patterns.
2. **Dataset Limitations**: Limited availability of diverse datasets restricts the model's applicability to new languages or gestures.
3. **Real-Time Performance**: Balancing high accuracy with low latency is critical for practical deployment.

VIII. APPLICATIONS

1. **Real-Time Translation Systems**: Converting sign language gestures into text or speech for seamless communication.
2. **Educational Tools**: Assisting in teaching sign language through interactive applications.
3. **Mobile Deployment**: Lightweight models enable integration into smartphones and wearable devices.

IX. FUTURE DIRECTIONS

1. **Multimodal Approaches**: Combining RGB, depth, and skeletal data to improve recognition accuracy.
2. **Transfer Learning**: Leveraging pre-trained models for cross-language generalization.
3. **Dynamic Gesture Recognition**: Extending the system to handle continuous sign language sequences.
4. **Dataset Expansion**: Developing multilingual datasets to support global applicability.

X. PERFORMANCE VARIATIONS

The performance of DenseNet-121 in sign language detection under varying lighting conditions is influenced by its ability to process features effectively, but its robustness can be enhanced through specific methodologies. DenseNet-121, with its densely connected layers, inherently promotes feature reuse and maintains efficient gradient flow, which helps in handling some variations in lighting. However, additional techniques are often required to ensure optimal performance in diverse lighting environments.

A. KEY FINDINGS ON LIGHTING CONDITION ROBUSTNESS

1. **Skeleton-Based Representations**:
   - Methods like Tree Structure Skeleton Images (TSSI) convert skeleton sequences into RGB images, which are less sensitive to lighting variations compared to raw RGB video data. DenseNet-121 has been successfully applied to these representations, achieving competitive accuracy across datasets such as WLASL-100 and AUTSL.
   - TSSI provides background invariance and robustness to lighting changes by focusing on skeletal joint data rather than raw pixel intensities.
2. **Data Augmentation**:
   - Data augmentation techniques, such as brightness adjustments, are commonly used during training to simulate various lighting conditions. This improves the model's generalization ability and robustness in real-world scenarios.
   - Studies have shown that DenseNet-121 benefits significantly from such preprocessing, enabling it to adapt better to diverse environments.
3. **Performance Metrics**:
   - While DenseNet-121 achieves high accuracy in controlled settings (e.g., 98.9% on custom datasets), its performance may degrade under extreme lighting conditions if no specific countermeasures (like preprocessing or augmentation) are employed.
   - Models leveraging skeleton-based methods or multimodal approaches (e.g., combining depth and thermal data) tend to perform better under challenging lighting conditions.
4. **Challenges**:
   - Background variations and extreme lighting changes remain a challenge for RGB-based models like DenseNet-121 unless supplemented with robust preprocessing or alternative data representations.
   - Skeleton-based approaches mitigate this issue but may require additional hardware or preprocessing steps.

DenseNet-121 performs well under varying lighting conditions when combined with techniques like skeleton-based representations or data augmentation. However, its raw RGB-based performance can be sensitive to extreme lighting changes unless explicitly addressed through preprocessing or multimodal inputs. Future research could explore integrating attention mechanisms or multimodal data (e.g., depth sensors) to further enhance its robustness in diverse environments.

XI. CONCLUSION

DenseNet-121 proves to be a highly effective architecture for time-based sign language detection, achieving impressive accuracy while maintaining computational efficiency. By addressing challenges such as signer independence and dataset limitations, future research can further enhance the system's robustness and scalability, paving the way for universally accessible SLR technologies.

REFERENCES

- *K Anitha, R Naveen Karthick* - "Gesture based Sign Language Recognition System"
- *Sangeeta Kurundkar, Arya Joshi* - "Real-Time Sign Language Detection"
- *Jeet Debnath, Praveen Joe I R* - "Real-time Gesture Based Sign Language Recognition System"
- *Nusrat Ansari, Siddhi Awari* - "GesSpy: ML Driven Real Time Sign Language Detection"
- *Monalisa Ghosh, Debjani De* - "R-SLR: Real-Time Sign Language Recognition System"
- *Yasir Altaf, Abdul Wahid* - "Deep Learning Approach for Sign Language Recognition using DenseNet201 with Transfer Learning"
- *Harshit Pandey, Amaan Ahmed* - "CNN based Sign Language Recognition System with Multi-format Output"
- *Taksheel Saini, Nandini Kumari* - "SignaSpectrum: AI-Driven Dynamic Sign Language Detection and Interpretation"
- Tanvir Shakil Joy et al-“Attention Trinity Net and DenseNet Fusion:Revolutionizing American Sign Language Recognition for Inclusive Communication”
- David Laines*, Miguel Gonzalez-Mendoza, et al-“Isolated Sign Language Recognition based on Tree Structure Skeleton Images”
- You Xuan Thung, et al-“Detecting languages in streetscapes using deep convolutional neural networks”
- Rangel Daroya, et al-“Alphabet Sign Language Image Classification Using Deep Learning”
- Basel A. Dabwan, et al-“Hand Gesture Classification for Individuals with Disabilities Using the DenseNet121 Model”.