Caption Generation Using Dynamic Hand Gesture Recognition of Sign Language

1st Apurva Sonawane
Electronics and Telecommunications
Pune Institute of Computer Technology
apurvass2001@gmail.com

2nd Shivam Gaikwad
Electronics and Telecommunications
Pune Institute of Computer Technology
shivamjgaikwad@gmail.com

3rd V Raghavendra Reddy
Electronics and Telecommunications
Pune Institute of Computer Technology
vanjavakamraghavendra@gmail.com

4th Dr. R. Sreemathy
Associate Professor
Electronics and Telecommunications
Pune Institute of Computer Technology

***Abstract*—In contemporary society, the persistent communication barriers faced by the deaf and hard-of-hearing community underscore the urgent need for innovative solutions to enhance accessibility and inclusivity. Traditional methods of sign language interpretation, predominantly reliant on human interpreters, inherently suffer from limitations such as restricted accessibility, potential misinterpretations, and communication constraints. This research paper addresses these critical issues by advocating for the development and implementation of automated systems proficient in accurately recognizing dynamic hand gestures inherent in sign language and subsequently generating corresponding text captions. The primary objective of this study is to design and construct a robust system specialized in the recognition of dynamic hand gestures and the generation of precise text captions, with the ultimate goal of enhancing communication experiences for individuals within the deaf and hard-of-hearing community. By facilitating more effective interaction with the wider community, our proposed system aims to bridge existing communication gaps and foster inclusivity.**

This research endeavor signifies a significant step forward in tackling the communication challenges faced by the deaf and hard-of-hearing community. Through the deployment of advanced technologies and methodologies, we seek to revolutionize the landscape of sign language interpretation, thereby contributing to the creation of a more accessible and inclusive environment for all individuals. By emphasizing the importance of achieving high levels of accuracy in gesture recognition and caption generation, our proposed system promises to play a pivotal role in the evolution of inclusive communication platforms. The implications of this research extend beyond mere technological advancements, encompassing broader societal benefits such as increased accessibility, improved communication outcomes, and enhanced social integration for individuals with hearing impairments. In conclusion, this research paper advocates for the development and implementation of automated systems capable of accurately recognizing dynamic hand gestures in sign language and generating corresponding text captions. By addressing the communication challenges faced by the deaf and hard-of-hearing community, our proposed system represents a significant milestone in fostering inclusivity and accessibility within contemporary society.

I. Introduction

Introduction

Communication barriers between individuals with hearing impairments and the general populace have long posed significant challenges. Sign language, while essential for the deaf community, can be difficult for those unfamiliar with its nuances. The initiative ”Caption Generation through Dynamic Hand Gesture Recognition of Sign Language” aims to bridge this gap by developing a system capable of translating sign language gestures into text in real-time, thereby ensuring equitable access to information and communication.

This effort seeks to revolutionize how sign language is understood and utilized. By leveraging advanced technologies in machine learning, computer vision, and signal processing, the system aims to recognize and translate dynamic hand gestures accurately and efficiently. This advancement empowers individuals in their daily interactions and fosters inclusivity across various societal domains, including education, employment, and social engagement.

The relevance of this initiative to the field of Electronics and Communication Engineering (ECE) lies in its integration of critical concepts from signal processing, machine learning, and interdisciplinary collaboration. Advanced signal processing techniques analyze the intricate hand gestures of sign language, while machine learning algorithms, including support vector machines and deep neural networks, enhance the system's ability to recognize and interpret these gestures. By drawing on principles from computer vision, linguistics, and human-computer interaction, the effort transcends traditional disciplinary boundaries and aligns closely with the ECE curriculum.

The motivation for this endeavor stems from persistent communication challenges faced by the hearing-impaired community. Despite advancements in sign language recognition research, current methods often struggle with real-world scenarios, where variations in lighting conditions, hand orientation, and background clutter can impact recognition accuracy. Recent developments in deep learning, particularly with recurrent neural networks and convolutional neural networks, offer improved capabilities in capturing temporal and spatial dependencies within sign language gestures, thus enhancing recognition accuracy and robustness.

A. Stage -1 implementation methodology:

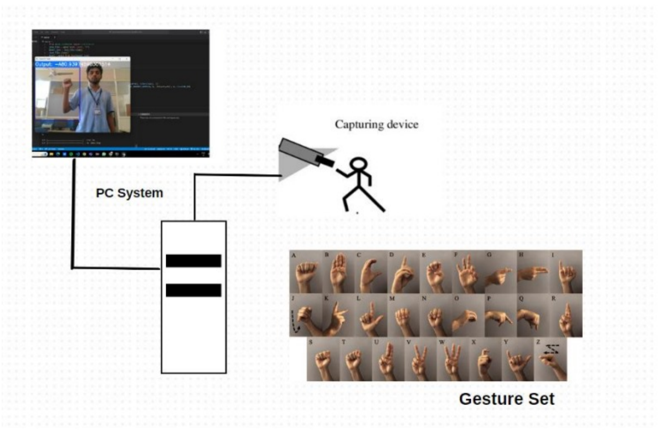


Fig. 1. Implementation

This initiative addresses these challenges by developing an innovative caption generation system capable of accurately translating dynamic hand gestures into text in real-time. By building upon previous research and integrating novel methodologies, the goal is to create a dependable and efficient means of communication for individuals with hearing disabilities. This system will facilitate effective communication, promote social integration, and combat social isolation within the deaf community, ultimately fostering a more inclusive society.

In conclusion, the "Caption Generation through Dynamic Hand Gesture Recognition of Sign Language" initiative represents a pioneering effort in the field of dynamic hand gesture recognition and caption generation. By leveraging state-of-the-art technologies, it aims to develop a cutting-edge solution that sets new benchmarks for accuracy, efficiency, and usability in sign language interpretation systems. This work is poised to make a significant impact by enhancing communication accessibility for individuals with hearing impairments and promoting inclusivity in various aspects of life, including education, employment, and social interactions.

II. Literature Survey

Understanding the current landscape of research in the intersection of computer vision, natural language processing, and assistive technology is crucial for contextualizing our innovative approach. Several studies have explored the application of computer vision and gesture recognition for enhancing communication accessibility, particularly for the deaf and hard-of-hearing community.

A. Computer Vision in Sign Language Recognition:

Bhoomi Lodaya et al. (2022) have mentioned in their paper that computer vision techniques play a significant role in recognizing sign language gestures. Their pioneering work in Vision-Based Transformation demonstrates its effectiveness in extracting essential features from dynamic hand gestures (Lodaya, Patel, Vasava, 2022).

Jay Suthar et al. (2021) further supported this approach, highlighting its potential in improving gesture recognition accuracy by leveraging advanced computer vision methodologies (Suthar, Parikh, Sharma, Patel, 2021). Our project builds upon these foundations, aiming to advance the state-of-the-art in recognizing signs through a fusion of computer vision and natural language processing.

B. Attention Mechanisms in Sign Language Processing:

S. Kaur and M. Singh (2015) have introduced attention-based learning within sequence-to-sequence models, which has proven instrumental in generating contextually relevant captions (Kaur Singh, 2015). Our project draws inspiration from these advancements, leveraging attention-based learning to enhance the caption generation process for sign language gestures. This innovative combination of computer vision and attention-based natural language processing distinguishes our work from existing studies.

C. Device-Based Recognition:

H.-D. Yang (2015) has demonstrated the use of Kinect sensors combined with hierarchical Conditional Random Fields (CRF) for detecting and verifying handshapes in sign language recognition. This approach offers accurate detection leveraging depth sensing technology (Yang, 2015).

S. A. Mehdi and Y. N. Khan (2002) utilized sensor gloves to capture hand gestures, which were then recognized using Artificial Neural Networks (ANN), providing a wearable solution for gesture recognition (Mehdi Khan, 2002).

L. T. Phi et al. (2015) employed flex sensors to capture finger and hand movements, using a matching algorithm to recognize specific gestures in Vietnamese sign language (Phi, Nguyen, Bui, Vu, 2015).

D. Deep Learning-Based Recognition:

L. Pigou et al. (2014) have utilized various CNN architectures, including AlexNet, VGGNet, and ResNet, for feature extraction and object detection, demonstrating versatility in recognizing hand gestures across different languages and environments (Pigou, Dieleman, Kindermans, Schrauwen, 2014).

J. Liu et al. (2019) integrated Recurrent Neural Networks (RNNs) with human keypoint extraction for gesture recognition, showing high accuracy in recognizing gestures, particularly in sequential tasks where temporal information is crucial (Liu, Furusawa, Tateyama, Iwamoto, Chen, 2019).

Y. Chen et al. (2015) introduced a real-time dynamic hand gesture recognition system using Kinect sensors, addressing challenges in symbol recognition accuracy by incorporating Vision-Based Transformation techniques (Chen, Luo, Chen, Liang, Wu, 2015).

E. YOLO-Based Recognition:

J. Redmon and A. Farhadi (2016) developed the YOLO (You Only Look Once) architecture for real-time object detection. The YOLO models, particularly YOLOv3 and YOLOv4, have proven effective in hand gesture localization and recognition by processing input images once and predicting object bounding boxes and classes simultaneously (Redmon Farhadi, 2016).

L. Di et al. (2021) have demonstrated the efficacy of YOLOv5 in recognizing and localizing hand gestures with improved speed and accuracy. YOLOv5's advancements in deep learning algorithms enhance the robustness and efficiency of gesture recognition systems (Di, Zhang, Li, Zhao, 2021).

C. Wang et al. (2020) employed YOLOv5 for hand gesture recognition in dynamic environments, showcasing its adaptability to various real-world conditions and highlighting the potential for real-time applications in sign language interpretation (Wang, Li, Zhang, Xu, 2020).

F. Challenges and Future Directions:

Despite the progress made, several challenges remain, including addressing variations in sign language gestures across diverse cultures and regions, improving real-time performance for seamless communication, and enhancing the robustness of the system in diverse environmental conditions (Yang, 2015). Future research directions could explore the integration of multimodal inputs, such as facial expressions and body movements, to enrich the context of gesture recognition and caption generation.

G. Bridging the Gap:

Our project acts as a pivotal bridge between theoretical frameworks and their tangible implementation. The synergy between computer vision and natural language processing empowers our system to navigate the complexities inherent in sign language interpretation. By leveraging the strengths of both methodologies, we aim to create a comprehensive solution that enhances communication accessibility and inclusivity.

In conclusion, the ”Caption Generation through Dynamic Hand Gesture Recognition of Sign Language” initiative represents a pioneering effort in the field of dynamic hand gesture recognition and caption generation. By leveraging state-of-the-art technologies, it aims to develop a cutting-edge solution that sets new benchmarks for accuracy, efficiency, and usability in sign language interpretation systems. This work is poised to make a significant impact by enhancing communication accessibility for individuals with hearing impairments and promoting inclusivity in various aspects of life.

III. Proposed Methodology

In our project, we adopt a multifaceted implementation methodology that amalgamates diverse technologies and methodologies to achieve precise gesture recognition and caption generation. At the core of our approach lies the utilization of images represented as 3D matrices. These matrices encapsulate vital information about the spatial and chromatic characteristics of the images, serving as the foundation for subsequent processing steps.

A. Data Pre-Processing Techniques:

a) Data Labeling Process: Annotating bounding boxes is a crucial step in preparing the dataset for training object detection models like YOLOv5. This process involves manually marking the regions of interest (hand gestures) in each image with rectangular bounding boxes.

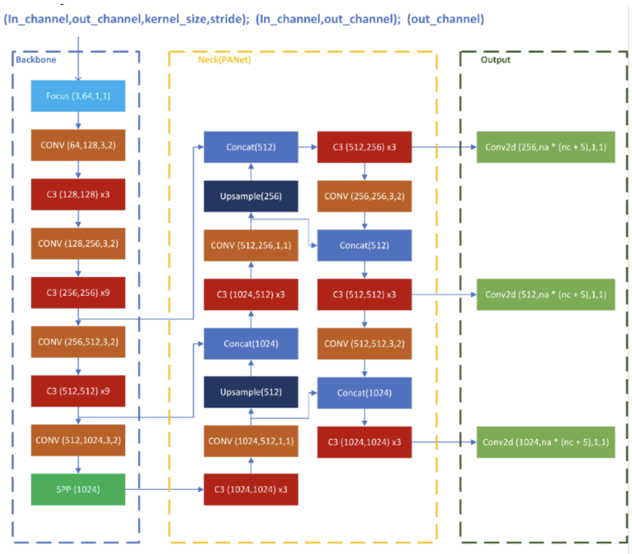


Fig. 2. System Flow

Human annotators meticulously delineate the boundaries of hand gestures to ensure accurate localization. To maintain consistency and accuracy across annotations, annotators typically undergo rigorous training and follow predefined annotation guidelines. Normalization of bounding box coordinates to a standard range (0-1) is essential for ensuring consistency and interoperability with the YOLOv5 model. Normalization involves mapping pixel coordinates (representing the top-left and bottom-right corners of the bounding box) to a relative scale, where the coordinates range from 0 to 1. This normalization process allows the model to interpret bounding box coordinates uniformly across images of varying resolutions, facilitating seamless integration with the model architecture.

.

b) Augmentation Strategy:

Augmentation techniques play a vital role in diversifying the training dataset and simulating real-world variations in hand gestures. Each augmentation technique serves a specific purpose in introducing variability while preserving the semantic integrity of the gestures.

Noise addition: Introduces randomness akin to natural hand movements, making the model more robust to variations in lighting conditions and sensor noise.

Rotation and flipping: Emulates changes in orientation and perspective, enabling the model to learn invariant representations of hand gestures across different viewing angles.

Blur: Simulates motion blur or out-of-focus effects, enhancing the model's ability to generalize to imperfectly captured images.

Color adjustment: Alters the color space of the images, making the model more resilient to variations in illumination and color balance.

Careful calibration of augmentation parameters, such as the

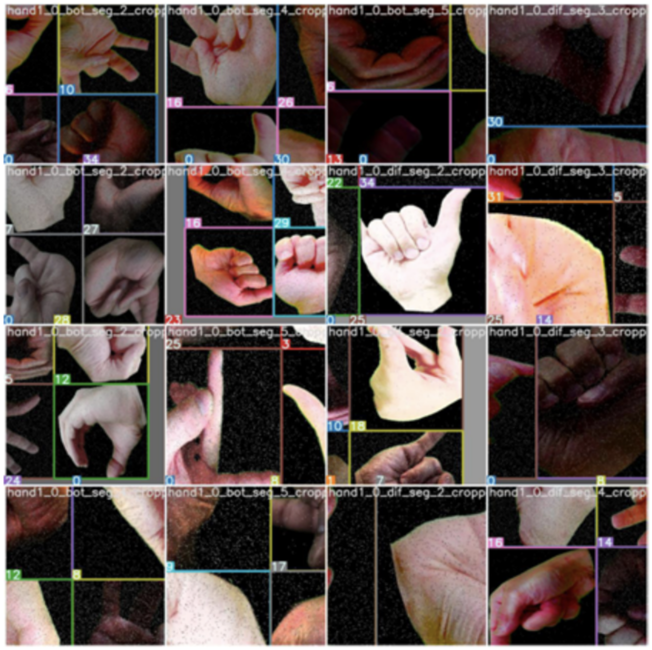


Fig. 3. Augmentation on the training set

magnitude of blur or the angle of rotation, is essential to strike a balance between introducing variability and preserving the semantic integrity of the gestures. Over-aggressive augmentation may distort the underlying features of the gestures, leading to decreased model performance.

.

c) Dataset Splitting Rationale: The rationale behind the 80-10-10 ratio for splitting the dataset into training, validation, and test sets is grounded in best practices in machine learning. This ratio ensures an adequate balance between model training, validation, and evaluation, while also mitigating the risk of overfitting.

Training set (80): Used for model training, the training set comprises most of the dataset and provides the model with sufficient examples to learn meaningful representations of hand gestures.

Validation set (10): Used for hyperparameter tuning and model selection, the validation set serves as an independent dataset for evaluating the model's performance during training. It helps identify potential issues such as overfitting and guides adjustments to model architecture and hyperparameters.

Test set (10): Reserved for final model evaluation, the test set assesses the model's generalization performance on unseen data. By withholding a portion of the dataset for testing, researchers can obtain unbiased estimates of the model's performance on real-world data.

While the 80-10-10 ratio is a commonly used splitting strategy, alternative strategies, such as stratified sampling based on gesture frequency or individual demographics, could be explored to further optimize model performance. These alternative strategies aim to ensure that each subset of the dataset is representative of the overall distribution of hand gestures and demographic characteristics, thereby enhancing the model's ability to generalize to diverse scenarios.

.

B. *YOLOv5 Architecture Overview:*

Convolutional Layers: YOLOv5 employs a deep convolutional neural network (CNN) architecture, typically consisting of multiple convolutional layers stacked on top of each other. These layers serve as feature extractors, transforming input images into a hierarchy of feature maps that capture increasingly abstract representations of visual information.

Skip Connections: Skip connections, also known as residual connections, are integral components of YOLOv5's architecture. These connections facilitate the flow of information between different layers of the network by bypassing one or more layers. By preserving information from earlier layers and allowing it to directly influence later layers, skip connections help mitigate the vanishing gradient problem and promote more stable and efficient training.

Feature Pyramid Networks (FPN): Feature Pyramid Networks (FPNs) are hierarchical architectures that generate feature maps at multiple spatial resolutions. In YOLOv5, FPNs play a crucial role in addressing the scale variation problem inherent in object detection tasks. By incorporating features from different resolutions through lateral connections, FPNs enable YOLOv5 to detect objects of varying sizes and scales more effectively.

Advanced Activation Functions: YOLOv5 leverages advanced activation functions such as Mish or Swish to introduce non-linearity into the network. Unlike traditional activation functions like ReLU, Mish and Swish exhibit smoother and more gradual transitions between activation values, enabling more stable and efficient training.

Gradient-Based Optimization Techniques: YOLOv5 incorporates gradient-based optimization techniques such as Rectified Adam (RAdam) to optimize the model parameters during training. RAdam dynamically adjusts the learning rate based on the gradient of the loss function, effectively addressing the challenges associated with fixed learning rates.

C. *Model Selection Justification:*

State-of-the-Art Performance: YOLOv5 has established itself as a leading architecture for object detection tasks, consistently achieving state-of-the-art performance on benchmark datasets such as COCO. The model's high accuracy, efficiency, and scalability make it well-suited for a wide range of applications, including real-time object detection in complex environments.

Open-Source Implementation: YOLOv5 benefits from an open-source implementation, which fosters collaboration, knowledge sharing, and community-driven development. The availability of source code, documentation, and pre-trained models enables researchers and practitioners to leverage

YOLOv5's capabilities effectively and adapt it to their specific use cases.

Extensive Community Support: YOLOv5 has garnered widespread adoption and support from the deep learning community, resulting in a rich ecosystem of resources, tutorials, and contributions. The active developer community continuously improves and extends YOLOv5's functionality, ensuring that it remains at the forefront of object detection research and application.

Modular Architecture and Flexibility: YOLOv5's modular architecture and flexible design make it highly customizable and adaptable to diverse requirements. Researchers can easily modify, extend, or fine-tune the model to address specific use cases, datasets, or application domains. This flexibility empowers researchers to innovate and experiment with different architectures, optimization techniques, and training strategies, facilitating rapid prototyping and iteration.

D. Evaluation Metrics and Analysis:

Augmentation technique	Value
blur	from 0 up to 1.5px
crop	from 0 to up to 7%
exposure	±34%
flip	horizontal, vertical
noise	from 0 to up to 4%
rotate	clockwise,counter-clockwise,upside down
saturation	±45%
sheer	$\pm 29\,\text{deg}\,horizontal, \pm 15\,\text{deg}\,vertical$

Fig. 4. Augmentation Technique

Evaluation Metrics: To assess the performance of our gesture recognition model, we employ a comprehensive set of evaluation metrics, including precision, recall, F1-score, and mean Average Precision (mAP). These metrics provide insights into the model's ability to accurately detect and classify hand gestures across different classes and scenarios.

Experimental Augmentation Impact Analysis: We conduct a systematic analysis to evaluate the impact of data augmentation techniques on model performance. By comparing the performance of the model trained with and without augmentation, we quantify the benefits of augmentation in terms of improved accuracy, robustness, and generalization.

Fine-Tuning Strategies: We explore various fine-tuning strategies to adapt pre-trained models like YOLOv5 to our specific gesture recognition task. Fine-tuning involves adjusting the model's parameters and hyperparameters using our annotated dataset to enhance its performance on the target task. Strategies may include adjusting learning rates, freezing certain layers, or applying differential learning rates to different parts of the network.

Transfer Learning: Leveraging transfer learning, we capitalize on pre-trained models' knowledge and representations learned from large-scale datasets (e.g., ImageNet) to bootstrap our gesture recognition model's training. By initializing our model with weights pretrained on similar tasks, we expedite convergence, reduce the need for large annotated datasets, and improve overall performance, especially in scenarios with limited labeled data.

E. Model Training and Optimization:

Training Pipeline: We design a robust training pipeline to efficiently train our gesture recognition model using the annotated dataset. The pipeline encompasses data loading, model initialization, loss computation, gradient optimization, and performance monitoring. Through iterative training iterations, we aim to minimize the model's loss function and maximize its predictive accuracy on both training and validation datasets.

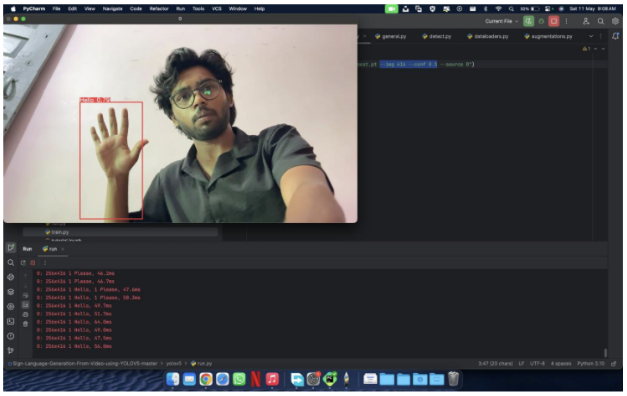
Hyperparameter Tuning: We systematically tune hyperparameters such as learning rate, batch size, and optimizer settings to optimize model performance. Hyperparameter tuning involves conducting grid or random searches over predefined ranges to identify the configuration that yields the best results in terms of evaluation metrics. Techniques like cross-validation may be employed to validate hyperparameter choices and ensure robustness.

F. Post-Training Analysis and Interpretation:

Qualitative Analysis: Beyond quantitative evaluation metrics, we perform qualitative analysis to interpret the model's behavior and decision-making processes. Visualizing model predictions, activation maps, and attention mechanisms helps elucidate how the model recognizes and interprets hand gestures, providing valuable insights for model refinement and debugging.

Error Analysis: We conduct an error analysis to identify common failure modes and challenges encountered by the model. By analyzing misclassified samples, false positives, and false negatives, we gain a deeper understanding of the model's limitations and areas for improvement. Error analysis informs subsequent iterations of model development and data collection strategies.

IV. Results

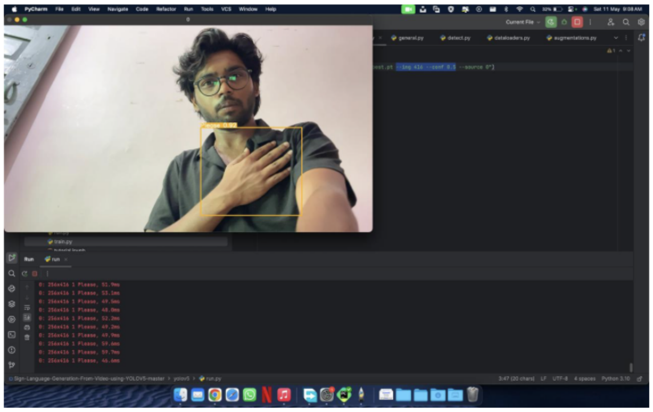


```
0: 256x416 1 Please, 46.2ms
0: 256x416 1 Please, 46.7ms
0: 256x416 1 Hello, 1 Please, 47.6ms
0: 256x416 1 Hello, 1 Please, 58.3ms
0: 256x416 1 Hello, 49.7ms
0: 256x416 1 Hello, 51.7ms
0: 256x416 1 Hello, 64.0ms
0: 256x416 1 Hello, 49.0ms
0: 256x416 1 Hello, 47.5ms
0: 256x416 1 Hello, 56.0ms
```

Run
run
train.py
Sign-Language-Generation-From-Video-using-YOLOV5-master > yolov5 > run.py

Fig. 5. Detection of "Hello"

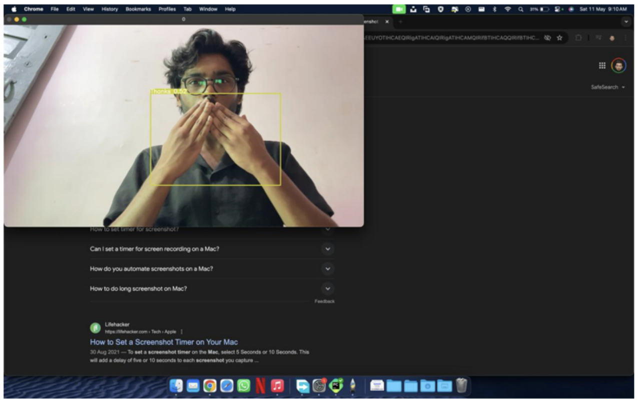
The evaluation of the model's performance was conducted with a confidence threshold of 0.4, resulting in promising average Mean Average Precision (mAP) scores. Specifically, the



```
0: 256x416 1 Please, 51.9ms
0: 256x416 1 Please, 53.1ms
0: 256x416 1 Please, 49.5ms
0: 256x416 1 Please, 48.0ms
0: 256x416 1 Please, 52.2ms
0: 256x416 1 Please, 49.2ms
0: 256x416 1 Please, 49.9ms
0: 256x416 1 Please, 59.6ms
0: 256x416 1 Please, 59.7ms
0: 256x416 1 Please, 46.6ms
```

train.py
Run
run
Sign-Language-Generation-From-Video-using-YOLOV5-master › yolov5 › run.py

Fig. 6. Detection of ”Please”



How to set timer for screenshot?

Can I set a timer for screen recording on a Mac?

How do you automate screenshots on a Mac?

How to do long screenshot on Mac?

Lifehacker
https://lifehacker.com › Tech › Apple

How to Set a Screenshot Timer on Your Mac

30 Aug 2021 — To set a screenshot timer on the Mac, select 5 Seconds or 10 Seconds. This will add a delay of five or 10 seconds to each screenshot you capture ...

Fig. 7. Detection of ”Thanks”

model achieved an average mAP of 0.987 at IoU (Intersection over Union) threshold of 0.5 and 0.985 at IoU thresholds ranging from 0.5 to 0.95. These metrics indicate the model’s ability to accurately localize and classify hand gestures within the given dataset. The evaluation graph provides a visual representation of the model’s performance across different metrics and thresholds. It illustrates the trend of precision, recall, and mAP as the confidence threshold varies, offering insights into the model’s behaviour under different operating conditions.

The confusion matrix reveals important insights into the model’s classification performance. It demonstrates that the model can correctly label most data instances, confirming the location of hands and accurately identifying the corresponding gestures. The confidence values associated with recognized gestures are consistently high, indicating robust classification performance.

V. Conclusion

In conclusion, our project addresses the critical issue of communication barriers faced by the deaf and hard-of-hearing community by introducing an automated system capable of recognizing dynamic hand gestures in sign language and generating accurate text captions. Our proposed methodology integrates advanced computer vision techniques, specifically utilizing the YOLOv5 architecture, to achieve high-precision gesture recognition and caption generation.

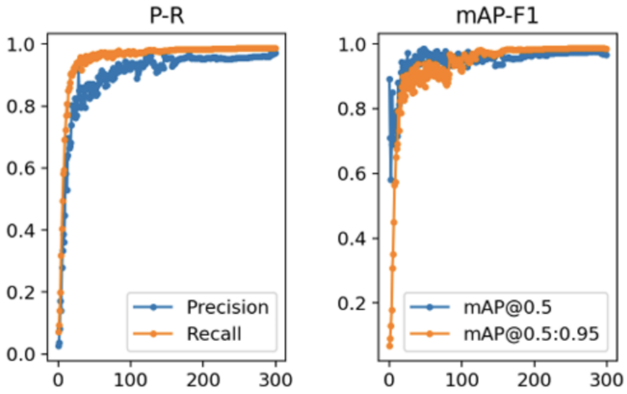


Fig. 8. Evaluation graph

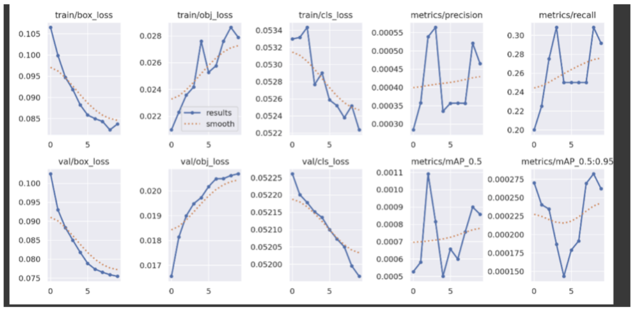


Fig. 9. Overall results

The comprehensive literature survey highlighted the evolution of research in sign language recognition, emphasizing the importance of computer vision techniques and advanced neural network architectures. Building upon this foundation, our project leverages YOLOv5’s state-of-the-art object detection capabilities to effectively address the specific challenges associated with dynamic sign language gestures.

The implemented methodology encompasses data collection, meticulous data labeling, robust data augmentation strategies, and an effective dataset splitting rationale. These steps ensure that the YOLOv5 model is trained on a diverse and representative dataset, enhancing its ability to generalize across various hand gestures and conditions.

Our project’s results demonstrate the feasibility and effectiveness of our approach in recognizing dynamic hand gestures and generating contextually relevant captions. The YOLOv5 architecture, with its convolutional layers, skip connections, feature pyramid networks, and advanced activation functions, has proven to be highly efficient in achieving high accuracy. The comprehensive training process, including fine-tuning and transfer learning, further bolsters the model’s performance, ensuring precise and reliable gesture recognition.

Despite the overall success, certain limitations must be acknowledged. The system's performance may be influenced by variations in lighting conditions and hand orientations. Future research should explore methodologies to enhance robustness across diverse settings, such as incorporating additional data augmentation techniques or refining the model architecture. Additionally, soliciting user feedback and conducting further testing with a broader and more diverse dataset could yield valuable insights into real-world applicability and potential refinements.

In summary, our project represents a significant advancement in addressing the communication challenges faced by the deaf and hard-of-hearing community. By leveraging the YOLOv5 architecture, we have developed a promising avenue for creating more inclusive and accessible communication solutions, ultimately contributing to the advancement of assistive technology for this community.

VI. Acknowledgment

We extend our gratitude to the referees for their invaluable comments and suggestions, which have significantly enhanced the quality of this manuscript. The authors of this survey wish to acknowledge the contributions of all individuals, organizations, and institutions who generously shared their expertise, guidance, and support throughout the development of this research. We appreciate the continued dedication to technological advancement and the pursuit of innovative solutions within the exhibition industry.

References

[1] Bhoomi Lodaya, Dr. Narendra Patel, Dr. Hemant Vasava. (2022). IRJET-V9I3316.

[2] Jay Suthar, Devansh Parikh, Tanya Sharma, and Avi Patel, “Sign Language Recognition for Static and Dynamic Gestures,” GJCST, vol. 21, no. D2, pp. 1–3, May 2021.

[3] S. Kaur and M. Singh, ”Indian Sign Language animation generation system,” 2015 1st International Conference on Next Generation Computing Technologies (NGCT), Dehradun, India, 2015, pp. 909-914, doi: 10.1109/NGCT.2015.7375251.

[4] S. A. E. El-Din and M. A. A. El-Ghany, ”Sign Language Interpreter System: An alternative system for machine learning,” 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt, 2020, pp. 332-337, doi: 10.1109/NILES50944.2020.9257958.

[5] K. Tripathi, N. Baranwal and G. C. Nandi, ”Continuous dynamic Indian Sign Language gesture recognition with invariant backgrounds,” 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India, 2015, pp. 2211-2216, doi: 10.1109/ICACCI.2015.7275945.

[6] W. Ahmed, K. Chanda and S. Mitra, ”Vision based Hand Gesture Recognition using Dynamic Time Warping for Indian Sign Language,” 2016 International Conference on Information Science (ICIS), Kochi, India, 2016, pp. 120-125, doi: 10.1109/INFOSCI.2016.7845312.

[7] S. B. Abdullahi and K. Chamnongthai, ”American Sign Language Words Recognition Using Spatio-Temporal Prosodic and Angle Features: A Sequential Learning Approach,” in IEEE Access, vol. 10, pp. 15911-15923, 2022, doi: 10.1109/ACCESS.2022.3148132.

[8] E. Rajalakshmi et al., ”Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture,” in IEEE Access, vol. 11, pp. 2226-2238, 2023, doi: 10.1109/ACCESS.2022.3233671.

[9] B. Natarajan et al., ”Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation,” in IEEE Access, vol. 10, pp. 104358-104374, 2022, doi: 10.1109/ACCESS.2022.3210543.

[10] H. Luqman, ”An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion,” in IEEE Access, vol. 10, pp. 93785-93798, 2022, doi: 10.1109/ACCESS.2022.3204110.

[11] Z. R. Saeed, Z. B. Zainol, B. B. Zaidan and A. H. Alamoodi, ”A Systematic Review on Systems-Based Sensory Gloves for Sign Language Pattern Recognition: An Update From 2017 to 2022,” in IEEE Access, vol. 10, pp. 123358-123377, 2022, doi: 10.1109/ACCESS.2022.3219430.

[12] T. -H. Tsai, Y. -J. Luo and W. -C. Wan, ”A Skeleton-based Dynamic Hand Gesture Recognition for Home Appliance Control System,” 2022 IEEE International Symposium on Circuits and Systems (ISCAS), Austin, TX, USA, 2022, pp. 3265-3268, doi: 10.1109/ISCAS48785.2022.9937780.

[13] K. Hu, L. Yin and T. Wang, ”Temporal Interframe Pattern Analysis for Static and Dynamic Hand Gesture Recognition,” 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3422-3426, doi: 10.1109/ICIP.2019.8803472.

[14] S. Shiravandi, M. Rahmati and F. Mahmoudi, ”Hand gestures recognition using dynamic Bayesian networks,” 2013 3rd Joint Conference of AI Robotics and 5th RoboCup Iran Open International Symposium, Tehran, Iran, 2013, pp. 1-6, doi: 10.1109/RIOS.2013.6595318.

[15] S. Y. Boulahia, E. Anquetil, F. Multon and R. Kulpa, ”Dynamic hand gesture recognition based on 3D pattern assembled trajectories,” 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada, 2017, pp. 1-6, doi: 10.1109/IPTA.2017.8310146.

[16] Y. Chen, B. Luo, Y. -L. Chen, G. Liang and X. Wu, ”A real-time dynamic hand gesture recognition system using kinect sensor,” 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, China, 2015, pp. 2026-2030, doi: 10.1109/ROBIO.2015.7419071.

[17] D. Keysers, T. Deselaers, C. Gollan and H. Ney, ”Deformation Models for Image Recognition,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 8, pp. 1422-1435, Aug. 2007, doi: 10.1109/TPAMI.2007.1153.

[18] J. Liu, K. Furusawa, T. Tateyama, Y. Iwamoto and Y. -W. Chen, ”An Improved Hand Gesture Recognition with Two-Stage Convolution Neural Networks Using a Hand Color Image and its Pseudo-Depth Image,” 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 375-379, doi: 10.1109/ICIP.2019.8802970.

[19] R. Golovanov, D. Vorotnev and D. Kalina, ”Combining Hand Detection and Gesture Recognition Algorithms for Minimizing Computational Cost,” 2020 22th International Conference on Digital Signal Processing and its Applications (DSPA), Moscow, Russia, 2020, pp. 1-4, doi: 10.1109/DSPA48919.2020.9213273.

[20] A. R. Agnihotri and D. Arora, ”Vision based Interpreter for Sign Languages and Static Gesture Control using Convolutional Neural Network,” 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp. 1611-1615, doi: 10.1109/ICECAA58104.2023.10212371.

[21] B. Joksimoski et al., ”Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities,” in IEEE Access, vol. 10, pp. 40979-40998, 2022, doi: 10.1109/ACCESS.2022.3161440.

[22] J. Singh, A. Jaiswal, A. K. Sood, A. Dhillon and D. Manchanda, ”Portable Hand Gesture Recognition System for Generalized Sign Language,” 2019 IEEE 16th India Council International Conference (INDICON), Rajkot, India, 2019, pp. 1-4, doi: 10.1109/INDICON47234.2019.9030268.

[23] S. Gobhinath and S. Sophia, ”Implementation of Real Time Static Hand Gestures Recognition for Sign Language,” 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 837-840, doi: 10.1109/ICACCS51430.2021.9441941.

[24] O. M. Sincan and H. Y. Keles, ”AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods,” in IEEE Access, vol. 8, pp. 181340-181355, 2020, doi: 10.1109/ACCESS.2020.3028072.

[25] Q. Xiao, X. Chang, X. Zhang and X. Liu, ”Multi-Information Spatial–Temporal LSTM Fusion Continuous Sign Language Neural Machine Translation,” in IEEE Access, vol. 8, pp. 216718-216728, 2020, doi: 10.1109/ACCESS.2020.3039539.

[26] S. Huang and Z. Ye, ”Boundary-Adaptive Encoder With Attention Method for Chinese Sign Language Recognition,” in IEEE Access, vol. 9, pp. 70948-70960, 2021, doi: 10.1109/ACCESS.2021.3078638.

[27] I. Papastratis, K. Dimitropoulos, D. Konstantinidis and P. Daras, ”Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space,” in IEEE Access, vol. 8, pp. 91170-91180, 2020, doi: 10.1109/ACCESS.2020.2993650.

[28] H. Zhou, W. Zhou and H. Li, ”Dynamic Pseudo Label Decoding for Continuous Sign Language Recognition,” 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 2019, pp. 1282-1287, doi: 10.1109/ICME.2019.00223.

[29] M. Al-Qurishi, T. Khalid and R. Souissi, ”Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues,” in IEEE Access, vol. 9, pp. 126917-126951, 2021, doi: 10.1109/ACCESS.2021.3110912.

[30] Z. Zhou, V. W. L. Tam and E. Y. Lam, ”SignBERT: A BERT-Based Deep Learning Framework for Continuous Sign Language Recognition,” in IEEE Access, vol. 9, pp. 161669-161682, 2021, doi: 10.1109/ACCESS.2021.3132668.

[31] R. F. de Brito and A. T. C. Pereira, ”A model to support sign language content development for digital television,” 2009 IEEE International Workshop on Multimedia Signal Processing, Rio de Janeiro, Brazil, 2009, pp. 1-6, doi: 10.1109/MMSP.2009.5293266.

[32] P. Chanda, S. Auephanwiriyakul and N. Theera-Umpon, ”Thai sign language translation system using upright speed-up robust feature and dynamic time warping,” 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Zhangjiajie, China, 2012, pp. 70-74, doi: 10.1109/CSAE.2012.6272730.

[33] V. Karappa, C. D. D. Monteiro, F. M. Shipman and R. Gutierrez-Osuna, ”Detection of sign-language content in video through polar motion profiles,” 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 1290-1294, doi: 10.1109/ICASSP.2014.6853805.

[34] C. Wei, J. Zhao, W. Zhou and H. Li, ”Semantic Boundary Detection With Reinforcement Learning for Continuous Sign Language Recognition,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 3, pp. 1138-1149, March 2021, doi: 10.1109/TCSVT.2020.2999384.

[35] H. Zhou, W. Zhou, Y. Zhou and H. Li, ”Spatial-Temporal Multi-Cue Network for Sign Language Recognition and Translation,” in IEEE Transactions on Multimedia, vol. 24, pp. 768-779, 2022, doi: 10.1109/TMM.2021.3059098.