Comparative Analysis of Data Augmentation Techniques in Hand Gesture Recognition

Diego Souza and Valmir Macario

Federal Rural University of Pernambuco - UFRPE, Recife PE 52171-900, Brazil

Abstract. Hand gesture recognition has gained significant attention due to its widespread applications in human-computer interaction, virtual reality, and assistive technologies. However, the scarcity of large, labeled datasets poses challenges such as overfitting and limited model generalization. To address this, we systematically evaluate 13 classical and six state-of-the-art (SOTA) data augmentation techniques for hand gesture recognition, conducting experiments on HGR1, OUHANDS, LIBRAS-UEFS, and EgoHands using the HGR-Net CNN architecture. Our results show that contrast-based augmentations (e.g., Solarize, Invert) improved accuracy by up to 21.16%, while mixing-based methods (e.g., MixUp, CutMix) often reduced performance, likely due to excessive distortion of gesture structures. Additionally, combining the best-performing augmentations was critical for maximizing accuracy across all datasets. However, these combinations did not always produce additive improvements, underscoring the importance of dataset-specific augmentation strategies for achieving optimal model performance.

Keywords: Hand Gesture Recognition · Data Augmentation · Convolutional Neural Networks · Model Generalization · Deep Learning · Classification.

1 Introduction

Gestures serve as a natural form of human communication, often complementing speech to convey intentions, emotions, and commands [16]. This has driven the development of gesture recognition systems for applications in virtual reality, robotics, sign language interpretation, and touch-free interfaces [24]. Despite advances in deep learning-based hand gesture recognition (HGR), a major limitation remains: the lack of large, labeled datasets, which negatively impacts model generalization and robustness.

Convolutional Neural Networks (CNNs) have become the leading approach for computer vision tasks [19], excelling in feature extraction and pattern recognition. However, their performance is often limited by insufficient and imbalanced training data, leading to suboptimal generalization in real-world scenarios. To address this challenge, data augmentation techniques have been widely adopted to artificially expand training sets by introducing geometric transformations, contrast adjustments, and synthetic data generation [13, 27, 28]. In the field of hand gesture recognition, the lack of large publicly available datasets remains a major obstacle, with few exceeding 10000 labeled samples [25], creating a significant bottleneck for training deep learning models.

This study aims to provide a comprehensive analysis by evaluating the effectiveness of 13 classical and 6 state-of-the-art (SOTA) data augmentation techniques in enhancing hand gesture recognition (HGR) performance. Unlike previous works focusing on individual augmentation methods, we compare multiple approaches across four diverse datasets using HGR-Net [7], a state-of-the-art convolutional neural network architecture. Our contributions include:

- A comprehensive benchmark analyzing data augmentation strategies in hand gesture recognition.
- An evaluation across multiple datasets with varying backgrounds, lighting conditions, and occlusions.
- Insights into dataset-specific augmentation strategies for optimizing deep learning models in HGR.

By bridging the gap between augmentation theory and practical application, this work provides valuable guidance for researchers and practitioners seeking to improve robustness and accuracy in HGR systems.

2 Related Work

Extensive research has demonstrated that data augmentation enhances dataset diversity and improves model generalization. Shorten and Khoshgoftaar [27] emphasized the benefits of geometric transformations, color space modifications, and adversarial training, which introduce meaningful variations and mitigate overfitting in CNNs. Pereira et al. [23] explored various augmentation techniques to stabilize training in models exposed to noisy labels, a challenge that is highly relevant to hand gesture recognition. Additionally, Fayaz et al. [9] validated the effectiveness of CutOut [8], CutMix [29], and MixUp [30] in enhancing model robustness, particularly in medical imaging applications. These findings suggest that optimizing augmentation strategies for specific domains, including HGR, is crucial for achieving improved performance and reliability.

Custom image generation has emerged as a promising approach to addressing the limited availability of hand gesture datasets. Limonchik et al. [20] utilized virtual 3D hand models to augment training data, improving classification accuracy with custom CNN architectures. Similarly, GAN-based augmentation has been applied to dynamic skeleton-based HGR [26], demonstrating the effectiveness of synthetic data in expanding training sets. Gomaa et al. [10] introduced SynthoGestures, a framework leveraging Unreal Engine to generate realistic hand gesture data, reducing overfitting risks in driving scenarios.

Hybrid augmentation strategies have been proposed to enhance dataset diversity by combining traditional weak data augmentation with green-screen image capture, generating synthetic background variations to improve ground truth annotation, as demonstrated by [1] and [4]. While these approaches focus primarily on background replacement, our work extends the exploration to a broader range of augmentation techniques, incorporating contrast-based transformations and state-of-the-art (SOTA) methods such as AutoAug and AugMix to assess their impact on hand gesture recognition models.

Spatial augmentations such as random cropping, rotation, shear, and zoom have been widely applied in HGR research. Dadashzadeh et al. [7] utilized these techniques to expand the training set four-fold, achieving state-of-the-art results. Other studies have also demonstrated the effectiveness of these augmentations in enhancing gesture recognition performance [14, 18].

However, pixel-based and advanced augmentation techniques remain largely underexplored in HGR models and datasets. This study addresses that gap by analyzing the impact of both spatial and pixel-based augmentations, along with their combinations, across multiple HGR datasets to identify techniques that consistently improve recognition accuracy. By benchmarking a diverse set of augmentations using a consistent CNN architecture, this research provides a comprehensive evaluation of their effectiveness. These insights can guide the development of custom datasets and optimize augmentation strategies for synthetic data generation and real-world applications.

3 Data Augmentation for HGR

This section provides an overview of the data augmentation techniques evaluated in this study, detailing their relevance to Hand Gesture Recognition (HGR) and their expected impact on model performance. Additionally, we introduce the HGR-Net architecture, explaining why it was selected as the benchmark for this analysis.

3.1 Data augmentation:

The data augmentation techniques evaluated in this study are categorized into basic augmentations and state-of-the-art (SOTA) techniques. These techniques were selected to analyze their impact on HGR models, assessing their effectiveness in improving recognition accuracy and model robustness. Below, we provide a brief overview of each method and its potential influence on hand gesture recognition.

Basic Augmentations: We applied spatial-level and pixel-level transformations, often referred to as weak augmentations, to introduce controlled variability in the dataset. These transformations can help improve generalization by making the model robust to different hand positions, lighting conditions, and background variations.

– **Random Crop** – Randomly selects a portion of the image and resizes it to its original dimensions. *It can enhance robustness to variations in hand positioning and prevent overfitting to specific image regions.*

- **Horizontal Flip** – Flips the image along its vertical axis. *It might help the model generalize across left- and right-hand gestures.*
- **Rotation** – Rotates the image within a specified range. *It can simulate natural variations in hand orientation, potentially improving recognition across different angles.*
- **Translation** – Moves the image horizontally or vertically. *It can help ensure the model remains invariant to slight shifts in hand position within the frame.*
- **Shear** – Tilts the image along one axis, distorting its shape. *It might be useful in helping the model recognize gestures from different perspectives and viewpoints.*
- **Invert** – Reverses pixel intensity values (e.g., black becomes white, and vice versa). *It improves adaptation to extreme lighting conditions by forcing the model to rely on shape rather than color.*
- **Equalize** – Adjusts the image's brightness and contrast by redistributing pixel intensity. *It normalizes lighting differences, potentially making gesture details more distinguishable.*
- **Posterize** – Reduces the number of color levels in an image. *It simplifies visual details, this could help the model focus on essential gesture features rather than fine textures.*
- **Contrast** – Increases or decreases the difference between light and dark areas. *It can help with the visibility of hand features against various backgrounds, improving gesture clarity.*
- **Autocontrast** – Automatically scales contrast to maximize the image's dynamic range. *It could help enhance feature distinctiveness while preserving natural color balance.*
- **Brightness** – Adjusts the overall luminance of the image. *It ensures the model remains robust to varying lighting conditions which is common in real-world hand gesture images.*
- **Sharpness** – Enhances or reduces edge definition. *It can help make hand contours more distinct, improving recognition of finer gesture details.*
- **Solarize** – Inverts pixel values above a certain brightness threshold. *It helps introduce controlled contrast variations, helping the model adapt to different illumination settings.*

SOTA Augmentations: We also evaluated six advanced augmentation techniques, including automated and adaptive methods that enhance dataset diversity and improve generalization.

- **CutOut [8]** – Randomly removes a square region of pixels from the image, replacing it with a solid color (usually black or mean pixel value). *It can help the model become invariant to missing or occluded hand regions, improving robustness to partial hand visibility.*
- **CutMix [29]** – Replaces a randomly selected region of an image with a patch from another image, blending their labels proportionally. *It encourages the model to learn features from mixed samples, improving generalization, but may negatively impact gesture recognition where shape details are critical.*

- **MixUp [30]** – Creates a new training image by blending two images with a weighted sum, generating a mixed-label output. *It aims to improve model regularization and decision boundaries but may degrade performance in hand gesture recognition by distorting gesture shapes.*
- **AugMix [12]** – Applies multiple random augmentations to an image, then blends them together using a weighted sum. *It can enhance robustness to distribution shifts and improve generalization but requires careful tuning to avoid excessive distortion in hand gestures.*
- **AutoAug [6]** – Learns an optimal augmentation policy from a predefined set of transformations using reinforcement learning. *It adapts augmentation strategies for specific datasets, leading to improved recognition accuracy but requiring dataset-specific optimization.*
- **RandAug [5]** – Simplifies the process proposed by AutoAug by randomly applying a set number of augmentations from a predefined list with varying magnitudes. *It reduces computational overhead compared to AutoAug while maintaining strong performance, but requires hyperparameter tuning for best results.*

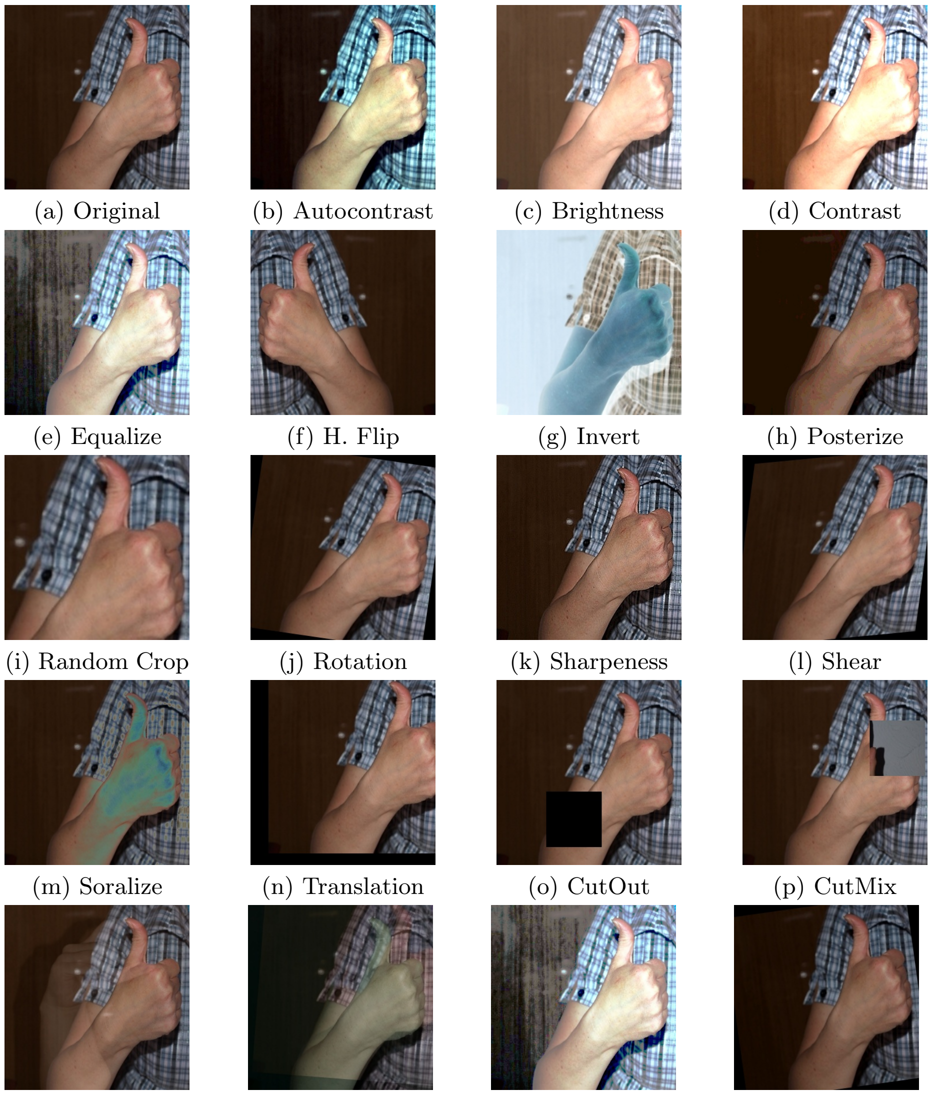
Figure 1 illustrates the effects of all methods on the original image, highlighting their impact on image transformation.

3.2 Convolutional Neural Network (CNN) - HGR-Net

HGR-Net [7] is a two-stage convolutional neural network (CNN) architecture designed for robust hand gesture recognition, as illustrated in Figure 2. It integrates hand segmentation and gesture recognition into a single pipeline, enhancing accuracy and generalization.

- **Hand Segmentation (Stage 1):** The first stage utilizes a fully convolutional residual network (FCRN) combined with Atrous Spatial Pyramid Pooling (ASPP) to segment hand regions from the background. This approach is particularly effective in eliminating background noise and does not require depth information, making it robust to illumination variations and complex environments.
- **Hand Gesture Recognition (Stage 2):** The second stage employs a *two-stream CNN*: One stream processes the original RGB image to extract appearance-based features. The other processes the segmented hand image, focusing on shape-based features. These feature representations are then fused in a fully connected layer before classification, ensuring that both color and structural information contribute to recognition accuracy.

Unlike traditional approaches that rely on handcrafted features, HGR-Net automatically learns meaningful representations from segmented hand images, leading to higher recognition accuracy and adaptability. Moreover, it operates without requiring depth information, making it well-suited for the datasets used in this study. Spatial augmentations have been shown to further enhance its performance, as demonstrated by Dadashzadeh et al. [7].



(a) Original (b) Autocontrast (c) Brightness (d) Contrast
(e) Equalize (f) H. Flip (g) Invert (h) Posterize
(i) Random Crop (j) Rotation (k) Sharpeness (l) Shear
(m) Soralize (n) Translation (o) CutOut (p) CutMix
(q) MixUp (r) AugMix (s) AutoAug (v) RandAug

Fig. 1. Illustration of the data augmentation techniques applied to the original image. (b)-(n) represent basic augmentations, while (o)-(v) depict state-of-the-art (SOTA) augmentations.

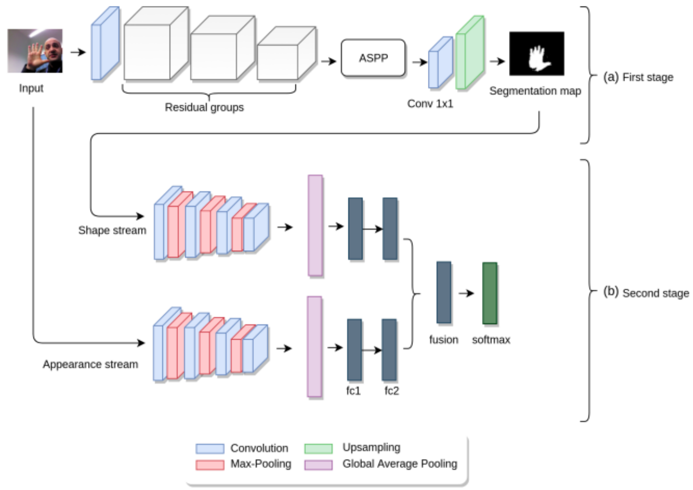


Fig. 2. Illustration of the HGR-Net architecture [7], highlighting its two-stage approach for segmentation and recognition.

Since this study focuses on evaluating the impact of diverse data augmentation techniques across varied hand gesture datasets, a consistent and robust model was essential. HGR-Net was selected as the benchmark architecture due to its state-of-the-art performance, computational efficiency, ease of implementation, and strong generalization capabilities enabled by its two-stream recognition architecture.

4 Methodology

This section provides an overview of the datasets used in this study and outlines the experimental setup, detailing the data augmentation strategies, CNN configuration, and training process, along with the specific adaptations made to optimize model performance.

4.1 Datasets

Four publicly available datasets with ground-truth hand masks were selected for this study:

HGR1 [11, 15, 22] - This is a hand gesture recognition dataset designed for gesture classification and shape analysis. It contains 899 RGB images from 12 individuals, featuring 27 distinct hand gestures. Collected in uncontrolled environments, the dataset ensures diverse backgrounds, lighting conditions, hand orientations, and user variations, while avoiding face-hand occlusions. Due to its small size and variability, HGR1 serves as a valuable benchmark for evaluating the impact of data augmentation on extremely limited datasets.

OUHANDS [21] - It consists of 3,000 images depicting 10 distinct hand gestures collected from 23 subjects, ensuring diversity. Each image is accompanied by a segmentation mask, and the dataset was also captured under challenging conditions, including variations in illumination, complex backgrounds, face-hand occlusions, and diverse hand shapes and sizes.

LIBRAS-UEFS [3] - This dataset comprises 9,600 images featuring 40 distinct hand signs from Brazilian Sign Language (LIBRAS), captured from six individuals. It includes letters, numbers, and words that can be recognized solely by hand configuration, excluding gestures requiring movement or body proximity. Each sign is represented by 240 images (50×50 resolution), with half being grayscale gesture images and the other half binary masks. The dataset was collected under controlled conditions, ensuring a standardized camera distance, simple white backgrounds, and minor variations in lighting, hand postures, and hand sizes.

EgoHands [2] - Unlike the other datasets in this study, EgoHands does not focus on static hand gestures. Instead, it consists of 48 Google Glass videos capturing first-person interactions between two individuals performing four activities—Jenga, Chess, Cards, and Puzzle—across different locations. For this study, 50 random frames were extracted from each video, generating 2,400 images with segmentation masks. The four activities were used as labels, ensuring a diverse representation of real-world hand movements for training and testing.

4.2 Experimental Setup

The experiment implementation was structured into three key steps to ensure a systematic evaluation of augmentation techniques on the HGR-Net model.

Step 1: Data Augmentation Implementation First, we applied data augmentation techniques using predefined parameters and transformation magnitudes based on AutoAug [6]. To ensure that spatial augmentations enhanced model generalization without introducing excessive distortions, each transformation was fine-tuned for hand gesture recognition, preserving critical gesture details.

- Random Crop: The minimum crop size was set to 85% of the image length to reduce the risk of excluding important hand details.
- Shear and Rotation: The distortion angle was limited to 15 degrees to prevent unnatural transformations that could alter gesture structure.
- Translation: The displacement was restricted to 15% of the image size, ensuring that gestures remained within recognizable boundaries.

These refinements allowed the model to leverage augmentation benefits while maintaining the integrity of hand shapes, preventing excessive deformations that could hinder recognition accuracy.

For state-of-the-art (SOTA) augmentations, additional refinements were made. CutOut and CutMix were modified to limit the cut size to a maximum of 25%, minimizing the risk of removing critical gesture components such as fingers and hand shapes. AugMix, AutoAug, and RandAug were primarily used with default configurations from [6], but some transformation magnitudes were adapted to align better with dataset characteristics. MixUp was applied with minimal intensity to prevent excessive blending, ensuring that key gesture features remained intact and recognition accuracy was not compromised.

We opted for offline data augmentation, applying all transformations before training to ensure reproducibility and consistency across multiple test iterations. To ensure consistency and preserve gesture integrity, spatial augmentations were applied to both RGB images and hand masks.

This approach introduced variability without compromising critical hand features. Given the small dataset sizes and the risk of overfitting, we designed the experiments to closely resemble real-world conditions. To achieve this, we selected the most challenging subjects and the most diverse image settings to construct the test set, ensuring a rigorous evaluation of model performance. Additionally, all images were resized to 320×320 to maintain uniformity and streamline the training process.

Step 2: CNN Model Training Setup The second step involved configuring the CNN model for training. HGR-Net operates in two stages, and we followed its default setup, utilizing the cost functions and Adam optimization algorithm as described in [7]. The learning rate was initialized at 0.001, with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We used the default batch size of 8 for the segmentation stage and 2 for the recognition stage, training all datasets for 200 epochs. To minimize overfitting in datasets with rapid convergence, we applied learning rate decay and an early stopping callback (patience: 20 epochs). While both network stages were trained with augmented data, this study focuses exclusively on evaluating the performance of the recognition stage.

Step 3: Dataset Splitting and Training For the final step, we split the datasets to begin training. Initially, we trained the baseline models without augmentation, allocating 70% of the dataset for training, 10% for validation, and 20% for testing. For experiments incorporating data augmentation, we expanded the training set by adding augmented samples equivalent to 50% of the original dataset size, ensuring that class distribution was maintained and preventing imbalance.

When combining multiple augmentation techniques, additional augmented samples were stacked on top of the original dataset, substantially increasing the training set size. No augmentations were applied to the validation set, ensuring an unbiased evaluation of the model's performance.

5 Results

The augmentation techniques were first evaluated individually on each dataset to assess their impact on model performance. The results for all augmentations, highlighting the best-performing ones, are presented in Table 1.

Table 1 shows that most augmentation techniques had a significant impact on HGR-Net. Across the four datasets (HGR1, LIBRAS, EGOHANDS, and OUHANDS), augmentation increased accuracy by an average of +7.74%, though its effectiveness varied depending on image characteristics, background complexity, and gesture diversity.

Among all techniques, contrast-based augmentations (Solarize, Invert, Autocontrast) yielded the highest accuracy gains, with Solarize performing best overall. This highlights the importance of contrast enhancement in gesture recognition, as it improves feature distinction and makes hand shapes more distinguishable across diverse backgrounds.

In the LIBRAS dataset, a grayscale set with minimal background variation, Autocontrast and Equalize significantly improved recognition by normalizing brightness and skin tone variations.

Spatial augmentations (Translation, Rotation, Shear) were particularly effective in datasets with diverse hand positions (HGR1, OUHANDS). By introducing positional variability, these transformations enhanced model robustness, ensuring that minor shifts in hand placement did not degrade recognition accuracy, thus improving adaptability to real-world conditions.

In contrast, MixUp reduced the accuracy across all datasets, likely due to gesture blending that creates ambiguous representations, making classification more difficult. Horizontal Flip had little effect on most datasets, suggesting that mirroring gestures do not introduce meaningful variation. However, it was highly effective in EgoHands, where first-person interactions inherently involve diverse hand orientations, making flipping a valuable transformation.

Finally, CutOut, CutMix, and AugMix produced mixed results, suggesting that aggressive transformations that modify hand shapes may not be well-suited for gesture recognition. These techniques risk removing or distorting critical gesture-defining features, potentially hindering model learning rather than improving generalization.

The chart in Figure 3 illustrates the accuracy improvement trends for all augmentation techniques applied to the HGR1, LIBRAS, EGOHANDS, and OUHANDS datasets. Each bar represents the relative accuracy increase compared to the original dataset, highlighting augmentation effectiveness based on dataset characteristics.

The best-performing augmentation techniques were selected for a combination analysis, yielding higher accuracy across all four datasets (HGR1, LIBRAS, EGOHANDS, and OUHANDS) compared to using individual augmentations, as shown in Table 2.

For HGR1, combining the top two contrast-based techniques (Solarize and Invert) with the best spatial transformations (Translation and Shear) resulted in an impressive accuracy of 92.78%, marking a 24.96% improvement over the

Augmentation	HGR1 (%)	LIBRAS (%)	EGOHANDS (%)	OUHANDS (%)
Basic Augmentations				
Original	67.82	88.44	67.85	74.69
Random Crop	79.23	85.37	71.23	77.00
Horizontal Flip	67.19	86.25	80.00	73.50
Rotation	77.31	90.31	73.69	74.17
Translation	81.01	88.47	77.85	79.17
Shear	79.77	89.00	76.31	**80.50**
Invert	84.29	88.69	**82.92**	74.33
Equalize	81.98	90.81	71.54	77.67
Posterize	81.55	89.31	66.92	75.67
Contrast	76.11	87.44	66.92	74.50
Autocontrast	79.07	**91.81**	69.23	77.17
Brightness	77.93	85.75	73.23	78.17
Sharpness	82.10	88.13	71.54	73.50
Solarize	**88.98**	90.75	70.15	77.50
SOTA Augmentations				
CutOut	81.49	87.12	72.77	72.67
CutMix	78.77	87.75	64.00	72.83
MixUp	61.85	86.50	63.69	68.00
AugMix	73.29	88.00	69.38	72.17
AutoAug	76.81	91.25	**82.92**	78.00
RandAug	80.92	90.50	71.38	77.17

Table 1. Accuracy results for all datasets with basic and state-of-the-art (SOTA) Augmentations. **Best results for each dataset are highlighted.**

baseline (no augmentation). The LIBRAS dataset achieved the highest accuracy overall (93.25%) when combining spatial augmentations and AutoAug, demonstrating the effectiveness of adaptive augmentation policies.

OUHANDS exhibited moderate performance compared to the other datasets, even when applying the best-performing augmentation techniques and combinations. This suggests that further fine-tuning and dataset-specific optimization are required to maximize its performance.

For EgoHands, where hand exposure is significantly lower compared to the background, the dataset benefited greatly from contrast-based augmentations, leading to substantial accuracy improvements compared to other augmentation combinations.

Most of the techniques individually demonstrated strong performance, and their strategic combination can further boost accuracy by expanding the training set and improving model generalization, confirming the impact of carefully curated augmentation strategies on hand gesture recognition models. Based on these results, we propose the following practical recommendations to optimize augmentation strategies:

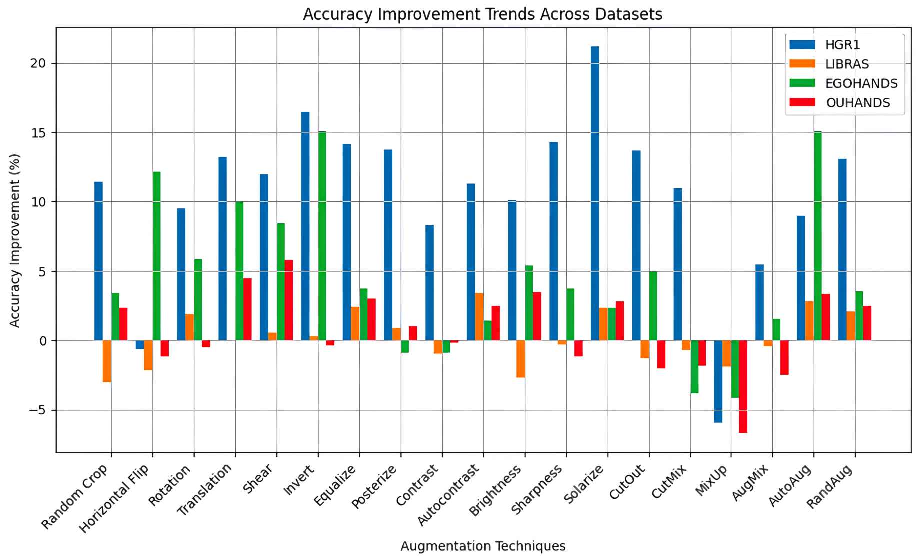


Fig. 3. Impact of Data Augmentation on Accuracy Across Different Hand Gesture Recognition Datasets.

- Leverage contrast-based augmentations (Solarize, Invert) and spatial transformations (Translation, Shear) to enhance feature distinction and improve generalization.
- For datasets with minimal background variation, prioritize Autocontrast and Equalize to normalize brightness and mitigate skin tone variability.
- For datasets where hand pose variation is not critical, apply Horizontal Flip to improve recognition of orientation-diverse hand movements.
- Avoid CutOut, CutMix, MixUp, and AugMix, as they distort gesture structures, making classification more challenging.
- AutoAug and RandAug are highly effective but require dataset-specific optimization. When fine-tuned, they can significantly enhance model generalization, especially when combined with classic augmentations such as Translation, Shear, or contrast adjustments, further improving robustness in hand gesture recognition tasks.
- Spatial-based augmentations must be carefully tuned to prevent the removal of critical gesture features, ensuring that transformations do not hinder recognition accuracy.

6 Conclusion and future work

In this study, we conducted a comprehensive and systematic analysis of classical and state-of-the-art (SOTA) data augmentation techniques and their impact on Hand Gesture Recognition (HGR) datasets.

Augmentation	HGR1 (%)	LIBRAS (%)	EGOHANDS (%)	OUHANDS (%)
Original	67.82	88.44	67.85	74.69
Classic Augmentation Combinations				
Invert + Shear	87.22%	92.94%	88.92%	70.17%
Invert + Translation	88.89%	92.56%	77.85%	74.83%
Solarize + Translation	88.33%	90.56%	79.08%	75.67%
Solarize + Shear	85.56%	88.88%	56.62%	77.50%
Solarize + Invert	86.11%	91.81%	**90.46%**	80.50%
Shear + Translation	86.67%	90.13%	79.85%	76.50%
Solarize + Invert + Shear + Translation	**92.78%**	89.25%	82.31%	74.50%
Classic + SOTA Augmentation Combinations				
Shear + AugMix	83.89%	88.19%	70.92%	77.33%
Shear + AutoAug	87.78%	91.25%	70.62%	76.33%
Shear + RandAug	86.67%	90.31%	74.92%	78.00%
Solarize + AugMix	86.11%	89.63%	75.38%	76.67%
Solarize + AutoAug	80.56%	90.94%	75.69%	75.17%
Solarize + RandAug	86.67%	89.81%	82.00%	78.00%
Solarize + Translation + AutoAug	88.33%	91.81%	75.69%	**81.83%**
Shear + Translation + AutoAug	89.44%	**93.25%**	68.31%	77.50%
Autocontrast + Equalize + AutoAug	91.67%	89.88%	68.46%	76.67%
Rotation + Contrast + AugMix	86.11%	87.44%	64.15%	78.33%

Table 2. Accuracy of Combinations Formed from the Best-Performing Classic and SOTA Augmentations. **Best results for each dataset are highlighted.**

Our findings reveal that contrast-based augmentations (Solarize, Invert) outperform commonly used spatial transformations (Rotation, Translation, Shear). Additionally, combining multiple augmentations is essential for achieving optimal results, as it enhances feature diversity, expands the training set, and strengthens model generalization.

While data augmentation requires dataset-specific fine-tuning, certain techniques consistently enhance performance in HGR applications. However, advanced methods such as CutOut, CutMix, MixUp, and AugMix should be approached with caution, as they often overdistort hand features, potentially degrading recognition accuracy.

Future research should extend augmentation techniques to dynamic gestures, as this study focused on static gesture recognition. Incorporating online augmentation strategies using adaptive state-of-the-art techniques such as AutoAug and RandAug could further enhance model adaptability by dynamically applying transformations during training. Additionally, this analysis can be expanded by integrating newly proposed augmentation methods, such as the three novel techniques introduced by Kumar et al. [17], to assess their impact on gesture recognition performance.

References

1. Awaluddin, B.A., Chao, C.T., Chiou, J.S.: A hybrid image augmentation technique for user- and environment-independent hand gesture recognition based on

deep learning. Mathematics **12**(9) (2024). https://doi.org/10.3390/math12091393, https://www.mdpi.com/2227-7390/12/9/1393
2. Bambach, S., Lee, S., Crandall, D.J., Yu, C.: Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions. In: The IEEE International Conference on Computer Vision (ICCV) (December 2015)
3. Bastos, I.L., Angelo, M.F., Loula, A.C.: Recognition of static gestures applied to brazilian sign language (libras). In: 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images. pp. 305–312. IEEE (2015)
4. Chalasani, T., Ondrej, J., Smolic, A.: Egocentric gesture recognition for head-mounted AR devices. CoRR **abs/1808.05380** (2018), http://arxiv.org/abs/1808.05380
5. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical data augmentation with no separate search. CoRR **abs/1909.13719** (2019), http://arxiv.org/abs/1909.13719
6. Cubuk, E.D., Zoph, B., Mané, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation policies from data. CoRR **abs/1805.09501** (2018), http://arxiv.org/abs/1805.09501
7. Dadashzadeh, A., Targhi, A.T., Tahmasbi, M., Mirmehdi, M.: Hgr-net: a fusion network for hand gesture segmentation and recognition. IET Computer Vision **13**(8), 700–707 (2019). https://doi.org/https://doi.org/10.1049/iet-cvi.2018.5796, https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-cvi.2018.5796
8. Devries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. CoRR **abs/1708.04552** (2017), http://arxiv.org/abs/1708.04552
9. Fayaz, S., Shah, S.Z., Assad, A.: Revolutionizing deep learning: A comprehensive exploration and comparison of cutting-edge data augmentation techniques. Recent Patents on Engineering **19** (10 2024). https://doi.org/10.2174/0118722121339343240910064504
10. Gomaa, A., Zitt, R., Reyes, G., Krüger, A.: Synthogestures: A novel framework for synthetic dynamic hand gesture generation for driving scenarios. In: Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. UIST '23 Adjunct, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3586182.3616635, https://doi.org/10.1145/3586182.3616635
11. Grzejszczak, T., Kawulok, M., Galuszka, A.: Hand landmarks detection and localization in color images. Multimedia Tools and Applications **75**(23), 16363–16387 (2016). https://doi.org/10.1007/s11042-015-2934-5
12. Hendrycks, D., Mu, N., Cubuk, E.D., Zoph, B., Gilmer, J., Lakshminarayanan, B.: Augmix: A simple data processing method to improve robustness and uncertainty (2020), https://arxiv.org/abs/1912.02781
13. Hernández-García, A., König, P.: Data augmentation instead of explicit regularization. CoRR **abs/1806.03852** (2018), http://arxiv.org/abs/1806.03852
14. Islam, M.Z., Hossain, M.S., ul Islam, R., Andersson, K.: Static hand gesture recognition using convolutional neural network with data augmentation. In: 2019 Joint 8th International Conference on Informatics, Electronics Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision Pattern Recognition (icIVPR). pp. 324–329 (2019). https://doi.org/10.1109/ICIEV.2019.8858563
15. Kawulok, M., Kawulok, J., Nalepa, J., Smolka, B.: Self-adaptive algorithm for segmenting skin regions. EURASIP Journal on Advances in Signal Processing **2014**(170), 1–22 (2014). https://doi.org/10.1186/1687-6180-2014-170, http://asp.eurasipjournals.com/content/2014/1/170

16. Kendon, A.: Gesture: Visible action as utterance. Cambridge University Press (2004)
17. Kumar, S., Asiamah, P., Jolaoso, O., Esiowu, U.: Enhancing image classification with augmentation: Data augmentation techniques for improved image classification (2025), https://arxiv.org/abs/2502.18691
18. Köpüklü, O., Rong, Y., Rigoll, G.: Talking with your hands: Scaling hand gestures and recognition with cnns. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 2836–2845 (2019). https://doi.org/10.1109/ICCVW.2019.00345
19. Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J.: A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems **33**(12), 6999–7019 (2022). https://doi.org/10.1109/TNNLS.2021.3084827
20. Limonchik, B.: 3 d model-based data augmentation for hand gesture recognition (2017), https://api.semanticscholar.org/CorpusID:22246903
21. Matilainen, M., Sangi, P., Holappa, J., Silvén, O.: Ouhands database for hand detection and pose recognition. In: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA). pp. 1–5 (2016). https://doi.org/10.1109/IPTA.2016.7821025
22. Nalepa, J., Kawulok, M.: Fast and accurate hand shape classification. In: Kozielski, S., Mrozek, D., Kasprowski, P., Malysiak-Mrozek, B., Kostrzewa, D. (eds.) Beyond Databases, Architectures, and Structures, Communications in Computer and Information Science, vol. 424, pp. 364–373. Springer (2014). https://doi.org/10.1007/978-3-319-06932-6_35, http://dx.doi.org/10.1007/978-3-319-06932-6_35
23. Pereira, E., Carneiro, G., Cordeiro, F.R.: A study on the impact of data augmentation for training convolutional neural networks in the presence of noisy labels. In: 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). vol. 1, pp. 25–30. IEEE (2022)
24. Rubin Bose, S., Sathiesh Kumar, V.: Hand gesture recognition using faster r-cnn inception v2 model. In: AIR. pp. 19–1 (2019)
25. Schak, M., Gepperth, A.: Gesture recognition on a new multi-modal hand gesture dataset. In: Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM,. pp. 122–131. INSTICC, SciTePress (2022). https://doi.org/10.5220/0010982200003122
26. Shen, J., Dudley, J.J., Kristensson, P.O.: The imaginative generative adversarial network: Automatic data augmentation for dynamic skeleton-based hand gesture and human action recognition. CoRR **abs/2105.13061** (2021), https://arxiv.org/abs/2105.13061
27. Shorten, C., Khoshgoftaar, T.: A survey on image data augmentation for deep learning. Journal of Big Data **6** (07 2019). https://doi.org/10.1186/s40537-019-0197-0
28. Taylor, L., Nitschke, G.: Improving deep learning using generic data augmentation. CoRR **abs/1708.06020** (2017), http://arxiv.org/abs/1708.06020
29. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. CoRR **abs/1905.04899** (2019), http://arxiv.org/abs/1905.04899
30. Zhang, H., Cissé, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. CoRR **abs/1710.09412** (2017), http://arxiv.org/abs/1710.09412

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Auxiliadora Cunha – CRB-4 1134

S719c Souza, Diego Rafael Ferreira de.

Comparative analysis of data augmentation techniques in hand gesture recognition / Diego Rafael Ferreira de Souza. – Recife, 2025.

15 f.; il.

Orientador(a): Valmir Macario Filho.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal Rural de Pernambuco, Bacharelado em Ciência da Computação, Recife, BR-PE, 2025.

Inclui referências.

1. Hand Gesture Recognition. 2. Data Augmentation. 3. Convolutional Neural Networks. 4. Deep Learning 5. Classification. I. Macario Filho, Valmir, orient. II. Título

CDD 004