

# Performance Analysis of InceptionV3-RCNN and CGAN Models for Image Forgery Detection

Babita Rimal<sup>1</sup>, Balkrishna Subedi<sup>2</sup>, Sagar Timalsena<sup>3</sup>

<sup>1</sup>Faculty Member, Divya Gyan College, Tribhuvan University

<sup>2</sup>Asst. Professor, Central Department of Computer Science and Information Technology, Tribhuvan University

<sup>3</sup>Department of Computer Science, Banaras Hindu University

## Article History

Received: 08/12/2025

Revised: 11/01/2026

Accepted: 05/01/2026

Published: 15/01/2026

## Corresponding Author

Name: Babita Rimal

Email: babitarimal@divyagyan.edu.np

ORCID: <https://orcid.org/0009-0007-4885-1137>

## Abstract

*Due to the usage of advanced editing software and rising of AI-generated content, there seems to be the great problem with the forgery detection. This paper is a comparison of two deep learning models InceptionV3-RCNN and a Conditional GAN (CGAN) using CASIA v2.0 dataset which is benchmark for forgery detection. The result concluded that InceptionV3-RCNN network was highly effective on the augmented dataset with an accuracy of 98.61 while CGAN model had an accuracy of 76.52%. The training time for InceptionV3-RCNN is less and inferred time is more, whereas CGAN has taken more training time and less inferred time. On the whole, the two models bring out different strengths. The combined approach of InceptionV3-RCNN and CGAN be useful to improve the effectiveness of image forgery detection which could be explored in future work. As the results suggest, by integrating the strong classification properties of InceptionV3-RCNN and localization fine-grained properties of CGAN the combined features can be utilized.*

**Keywords:** *Image Forgery Detection, InceptionV3-RCNN, Conditional GAN, Deep Learning, Digital Forensics.*

## **Introduction**

The use of advanced digital image editing tools, which are powered by AI has mostly used in the manipulation of visual content. In most cases, the manipulations can be highly challenging to get identified by the detection system regardless of whether the manipulation is splicing, copy-move, or deepfakes. Significant issues can be seen due to Forged images in the field of journalism, the law enforcement, medicine, scientific studies, and even social media. The traditional detection methods are insufficient to detect the subtle or highly sophisticated forgeries that make use of metadata or manual features. Due to this, a reliable system for detection of forgery has become an need in digital image forensics(Ross et al., 2020).

The recent advancement in deep learning have provided new in opportunities to overcome the challenges pose in the digital forensics. Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) models have proven to be the powerful tools. They can directly acquire meaningful patterns and spatial inconsistencies as dictated by the image data. Two of these models, namely 1) InceptionV3-RCNN, a multi-scale feature extraction-based architecture which utilizes adversarial learning to produce tampering masks and classify images whether it is authentic or manipulated; and (2) Conditional GAN, a model that employs adversarial learning to produce tampering masks and classify image as authentic or manipulated. InceptionV3-RCNN is a multiscale feature extraction-based system that integrates region-based detection and this allows it to classify and localize the areas of manipulation with great efficiency. Conditional GANs, by comparison, follow adversarial learning, which designs pixel-level tampering masks, but at the expense of more intricate training. These complementary features encourage a comparative analysis of the two architectures to detect forgery of images.

This work is aimed at comparing these two models by their classification accuracy and computational efficiency and the overall capability to help in the image forgery detection.

## **Literature Review**

Image forgeries detection methods has developed widely due to the use of deep learning approaches as opposed to the traditional feature-based technique. The initial copy moves and splicing detection systems were based on block-based extracting features or key point-based. They performed well on simple manipulations. When subsequent post-processing functions such as compression, rotation, and noise were used then these methods were not so effective.

So, using convolutional neural networks (CNNs), helped learning discriminative features directly from image data (Abidin et al., 2019).

In the forgery image detection, CNN based copy move / splicing manipulation got a decent accuracy in CASIA data but the results were worse with complicated manipulations. (Mallick et al., 2022). Further stability and accuracy of transformation are achieved by using models based on the dense Nets that convoluted by counting on the dense feature (Alzahrani, 2024), (Hingrajiya and Patel, 2023). When applied together with the Error Level Analysis (ELA), the preprocessing Pretrained networks, e.g., VGG-19, Inception-V3, ResNet, and EfficientNet, have also proven to be effective (Joshi et al., 2022).

The region-based approaches tend to have higher computational complexity in inference. Inception combined with Faster R-CNN that has multi-scale feature extraction to detect and localize the forged region. The CNN models based on region are the extensions of classification that involve localization to improve forgery localization accuracy (Vijiyakumar et al., 2024). Another approach which has become prominent in pixel-level forgery localization is Generative Adversarial Networks (GANs). The CGAN model that was proposed by (Abdalla et al., 2019) has the ability to detect the minor copy move based image forgeries. Recent GAN-based models are more successful in mask generation. But they have poor training stability and invariance due to data augmentation which results in hidden distortions (Pham and Park, 2023). More recent studies investigate the concept of robustness and generalization. (Wu et al., 2023) presented contrastive clustering to enhance resilience to post-processing operations while (Shi et al., 2023) presented discrepancy-guided reconstruction learning to incorporate semantic and forgery-specific cues. Regardless of these developments, there is still an evident trade-off between the robustness in classifications (CNN-based models) and fine-grained localization (GAN-based models).

However, this study offers a direct comparative study of a region-based CNN (InceptionV3-RCNN) and a Conditional GAN under the similar experimental conditions using CASIA v2.0 dataset unlike previous researchers who usually are the part of one. Through combined assessment of classification and localization and computational efficiency, the research paper raises real-world image forgery detection applicability trade-offs and contributes to the successful choice of appropriate models for image forgery detection.

## **Methodology**

### **A. Dataset**

The CASIA v2.0 Image Tampering Detection Dataset (Dong et al., 2013), is the dataset used for this study which consists of 3 folders namely authentic, tampered and ground-truth. It includes mainly tampered image with splicing, copy-move, and region editing. All the images are in JPEG format and with different sizes.

Experiments were done on an augmented version. Flips, rotations, and introducing noise were among the augmented data transformations with which the models trained on a more diverse set of distortions. For the both models, data set is split in 80:20 ratio for train and test split.

### **B. Data Preprocessing**

Both the approaches required different preprocessing setup. In the case of InceptionV3-RCNN, the images were reduced to the size of 299x299x3. They were then normalized, and augmentation is performed on the dataset. For assessing the generalization capability of the model, the augmentation played an important role. It helps in changing lighting, orientation, and texture.

Similarly, In the case of the Conditional GAN (CGAN), image size 256x256x3 was taken which matched with their ground-truth masks. The spatial structure remains unchanged. It helps keeping both the image and the corresponding mask to transform identically since mask-aligned augmentation was used. As CGAN generator, generates to pixel level tampering masks.

### **C. InceptionV3-RCNN**

InceptionV3-RCNN model combines the features of multi-level feature extraction with regions-based approach. It helps to identify regions of images they are manipulated. InceptionV3 backbone, as Shown in Figure 1 below, initially works on the input image i.e., augmented data and obtains multi scale feature maps. Parallel convolutional filters of different sizes are used in each layer. Convolutions filters helps the model to detect minor as well as high-level semantic manipulations that are often used to tamper the image.

The feature that are obtained then given fed to region-based convolutional neural network (RCNN) detection head. RCNN module primarily produces region-based area of interest that

has potential regions of manipulation. RCNN head the performs binary classification of each region proposal as tampered or authentic. Also bound-box regression is performed to localize forged regions. Multi-task loss is used in training that is a combination of classification loss and localization loss to enable the network to trade-off detection accuracy and spatial precision.

This combined architecture of InceptionV3 and RCNN helps locating global contextual cues but are also sensitive to localized discrepancies. So, it is highly suitable to localized various manipulation in detecting robust image forgery detection.

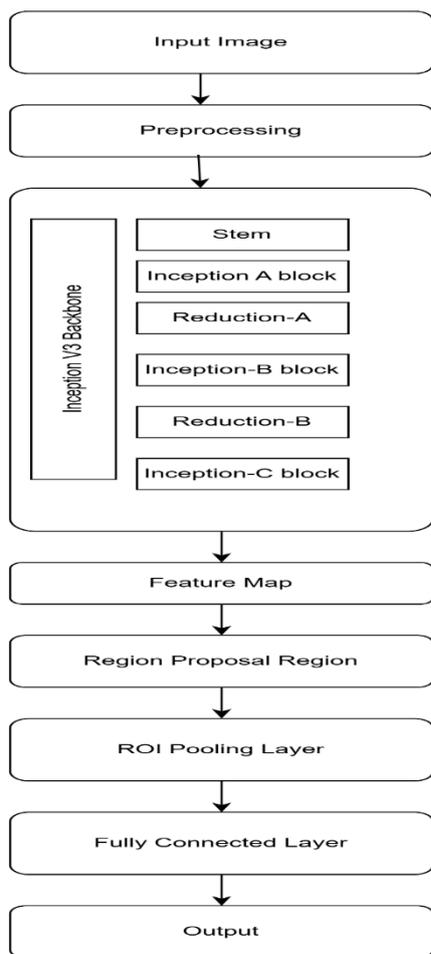


Figure 1: Inception V3 RCNN

#### D. Conditional GAN

The Conditional Generative Adversarial Network (CGAN) that will be used in the given research paper is aimed at pixel-wise forgery localization and authenticity classification. As shown in Figure 2 below, the CGAN is made up of both a generator and a discriminator both

of which are conditioned to the input image. Conditioning enables the model to gain manipulation patterns that are directly related to image content.

The generator has a U-Net architecture encoder-decoder model. The encoder gradually generates multi-scale feature representations that reflect not only global structure but also local artifacts and the decoder loads back a tampering mask that identifies manipulated areas. Jump connections between similar encoder and decoder layers retain spatial information and enhance boundary precision of the produced masks.

The image-mask pair is sent to the discriminator which compares two goals: adversarial discrimination of the real and generated masks, and binary classification of image authenticity. Adversarial loss, Dice loss and L1 loss are used to guide training, which together promote accuracy of mask generation, stable convergence and accurate localization.

The CGAN architecture has a more localized localization compared to region-based detection, but has a more complicated training compared to the adversarial optimization.

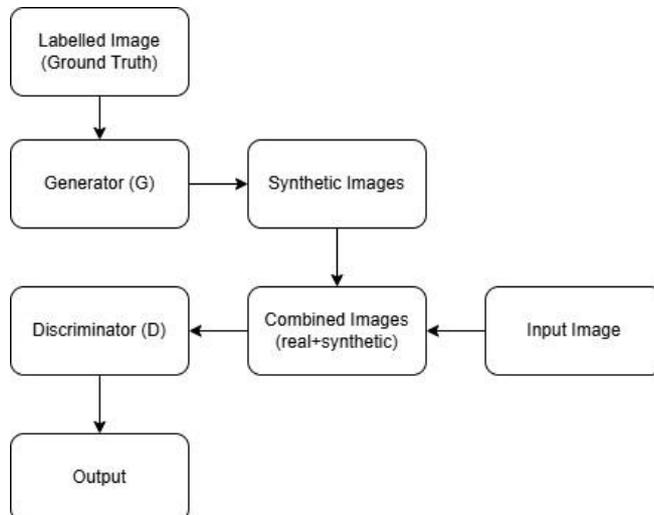


Figure 2: CGAN Architecture

## E. Evaluation Metrics

To make certain, that the two models were fairly and comprehensively compared, set of conventional classification and computational measures were used to assess the performance of each under augmented and non-augmented dataset conditions.

For **classification assessment**, the following measures were used:

- **Accuracy:** It represents the overall ratio of correctly identified authentic and tampered images.
- **Precision:** It reflects how reliably the model identifies tampered images
- **Recall:** It measures the model's ability to detect actual tampered images.
- **F1-score:** It is harmonic mean of precision and recall.
- **AUC-ROC:** It evaluates the model's capability to distinguish between authentic and tampered images

Besides classification measures, the performance of computation using training time and inference time was also evaluated in order to get the clear idea of various models.

### **F. Training Configuration and Hyperparameter**

In case of the InceptionV3-RCNN model, transfer learning was used which included the pretrained ImageNet versions of the InceptionV3 backbone. This model was trained end-to-end by means of the Adam optimizer with a starting learning rate of 0.0001 and a batch size of 16. 30 epochs were trained using early stopping which relied on a validation loss to avoid overfitting. In binary classification, categorical cross-entropy loss was utilized; in the RCNN detection head, Smooth L1 loss was utilized with bounding-box regression.

In the case of the Conditional GAN (CGAN), the discriminator and generator networks were both trained via the Adam optimizer with a learning rate of 0.0002 and  $\beta_1 = 0.5$ . Adversarial training has memory limits that led to the use of a batch size of 8. 50 epochs were used to train to get convergence between the discriminator and the generator. The adversarial learning binary cross-entropy loss combined with generator loss till the generator focuses on minimizing its loss, Dice Loss of coverage with ground-truth tampering mask, and L1 Loss of reconstruction of the model. The discriminator minimized binary cross-entropy loss on real/fake discrimination as well as authenticity classification.

### **G. Proposed Hybrid Approach (Future Work)**

Although this study assesses InceptionV3-RCNN and CGAN separately, findings of the experiments suggest that they complement each other. InceptionV3-RCNN shows simple distinguish performance and stable training performance, and CGAN captures the fine pixel-level location information with faster calculation speed for inference.

A possible solution could be a hybrid approach by using InceptionV3-RCNN as the main detector to classify images and spot suspicious regions, and then use a CGAN-based module to obtain pixel-level localization in the detected suspicious regions. Such a pipeline would have the advantage of minimizing false positives while increasing the localization precision. This combined framework was not used in the current study and suggested as a future research direction. The comparative results presented in this work provide empirical justification of such integration and serve as a basement for future development of such hybrid models.

## Result And Discussion

### A. Classification Report

The following table shows accuracy, precision, recall, f1 score for each class of the respective models:

*Table 1: Table showing Classification report of Inception V3-RCNN and CGAN*

Classification Metrics	Accuracy	Precision	Recall	F1 Score	AUC-ROC
Inception V3-RCNN	98.61%	0.9640	0.9721	0.9680	0.8449
CGAN	76.59%	0.4686	0.6244	0.5354	0.7926

Table 1 represents the results of both models using the CASIA v2.0 dataset. The data presented in the table above is for precision, recall and F1-score are macro-averaged values between the classes 'authentic' and 'tampered'.

The InceptionV3-RCNN model performed with the accuracy of 98.61%, strong precision (0.9640), recall (0.9721) and F1-score (0.9680). With a AUC-ROC value of 0.8449, it has a reliable discriminative ability between true and fraudulent images under different thresholds. These results also indicate a good generalization of the model in spite of the variations introduced by data augmentation.

In contrast, the CGAN obtained a lower accuracy 76.59%, with lower precision value (0.4686) and moderate recall (0.6244). The lower F1-score indicates difficulties in ensuring that

classification is consistent under the conditions of augmentation. Although the CGAN shows reasonable performance of AUC-ROC (0.7926), it still has an inferior result of classification reliability than the region-based CNN method.

### B. Accuracy Graph

The following graph shows the training and validation accuracy

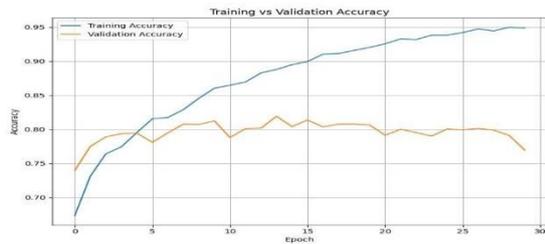


Figure 3: Training Vs Validation Accuracy for Inception V3RCNN

Figure 3 give insight for the training and validation accuracy plots for the InceptionV3-RCNN model. The curves suggest that there is stable convergence and no overfitting.

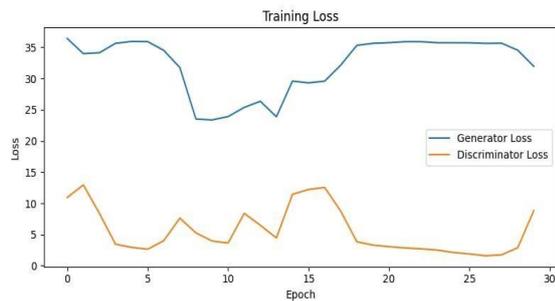


Figure 4: Generator and Discriminator Loss for CGAN

Figure 4 provides the insight of the generator and discriminator loss trends for CGA. It represents that adversarial stabilization is gradually achieved during the training epoch with higher variance than the CNN-based model.

### C. AUC-ROC Graph

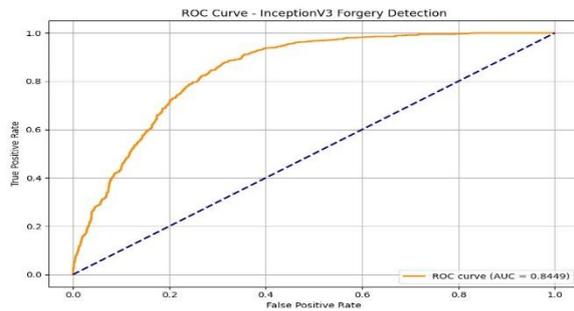


Figure 5: AUC-ROC For InceptionV3-RCNN

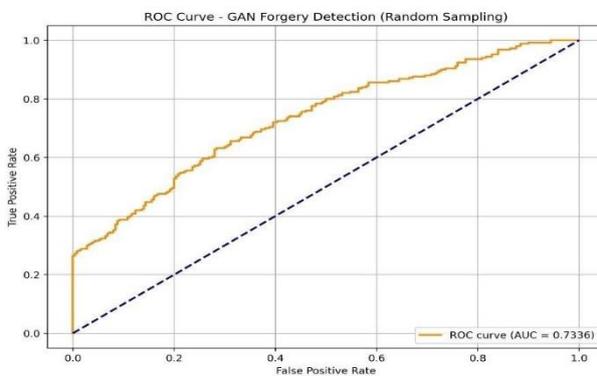


Figure 6: AUC-ROC For CGAN

Figures 5 and 6 show the Area under the Curve ROC (AUC) for InceptionV3-RCNN and CGAN respectively. Here InceptionV3-RCNN achieves a better AUC-ROC value than CGAN. The outcome of show that its value is satisfactory. Since it has no perfect separability which leaves a room for improvement by using hybrid or ensemble methods.

### D. Computational Performance

Regarding the efficiency of the computations, InceptionV3-RCNN has less training time as it uses transfer learning. But it has more time to get the inference result. On the other hand, CGAN has more training time as a result of adversarial optimization. But it has a faster inference which may be useful in real-time applications of localization.

In general, the result indicates that a distinct trade-off between InceptionV3-RCN, having higher classification robustness and CGAN having finer-grained localization will be more efficient which can be explored in future work.

## Conclusion

Performance analysis of InceptionV3-RCNN and Conditional GAN models were conducted for detecting image forgery using the CASIA v2.0 dataset. Through experimentation, it is concluded that InceptionV3-RCNN has better classification accuracy, better ratings of precision and recall, and more robust training characteristics for augmented data whereas CGAN shows weaker in the classification aspect, but has better result in inference speed and pixel-wise localization.

These results suggest that Inception V3RCNN models are beneficial where the classification is in priority, and the use of GAN-based models as beneficial in the tasks where localization is in priority. One area of future research could involve the studies of hybrid architectures that would unite the strength of CNN-based detectors and the localization capability of GAN-based schemes.

## References

- Abdalla, Y., Iqbal, M. T., & Shehata, M. (2019). Copy-Move Forgery Detection and Localization Using a Generative Adversarial Network and Convolutional Neural-Network. In *Information* (Vol. 10, Issue 9). <https://doi.org/10.3390/info10090286>
- Abidin, A., Majid, H., A Samah, A., & Hashim, H. (2019). *Copy-Move Image Forgery Detection Using Deep Learning Methods: A Review*. <https://doi.org/10.1109/ICRIIS48246.2019.9073569>
- Alzahrani, A. (2024). Digital Image Forensics: An Improved DenseNet Architecture for Forged Image Detection. *Engineering, Technology & Applied Science Research*, 14, 13671–13680. <https://doi.org/10.48084/etasr.7029>
- Dong, J., Wang, W., & Tan, T. (2013). *CASIA-2.0 Image Tampering Detection Dataset*. <https://www.kaggle.com/datasets/divg07/casia-20-image-tampering-detection-dataset>
- Hingrajiya, K. H., & Patel, C. (2023). An Approach for Copy-Move and Image Splicing Forgery Detection using Automated Deep Learning. *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 1–5. <https://doi.org/10.1109/ESCI56872.2023.10100202>
- Joshi, R., Gupta, A., Kanvinde, N., & Ghonge, P. (2022). Forged Image Detection using SOTA Image Classification Deep Learning Methods for Image Forensics with Error Level Analysis.

2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1–6. <https://doi.org/10.1109/ICCCNT54827.2022.9984489>

Mallick, D., Shaikh, M., Gulhane, A., & Maktum, T. (2022). Copy Move and Splicing Image Forgery Detection using CNN. *ITM Web Conf.*, 44. <https://doi.org/10.1051/itmconf/20224403052>

Pham, N., & Park, C.-S. (2023). Toward Deep-Learning-Based Methods in Image Forgery Detection: A Survey. *IEEE Access*, PP, 1. <https://doi.org/10.1109/ACCESS.2023.3241837>

Ross, A., Banerjee, S., & Chowdhury, A. (2020). Security in smart cities: A brief review of digital forensic schemes for biometric data. *Pattern Recognition Letters*, 138, 346–354. <https://doi.org/https://doi.org/10.1016/j.patrec.2020.07.009>

Shi, Z., Chen, H., Chen, L., & Zhang, D. (2023). Discrepancy-guided reconstruction learning for image forgery detection. *ArXiv Preprint ArXiv:2304.13349*.

Vijiyakumar, K., Govindasamy, V., & Akila, V. (2024). An effective object detection and tracking using automated image annotation with inception based faster R-CNN model. *International Journal of Cognitive Computing in Engineering*, 5, 343–356. <https://doi.org/https://doi.org/10.1016/j.ijcce.2024.07.006>

Wu, H., Chen, Y., & Zhou, J. (2023). Rethinking image forgery detection via contrastive learning and unsupervised clustering. *ArXiv Preprint ArXiv:2308.09307*.