



Handling Imbalance in Javanese Manuscript Character Dataset using Skeleton-based Balancing Generative Adversarial Networks

Muhammad 'Arif Faizin¹, Nanik Suciati^{2*}, Chastine Fatichah³

^{1,2,3}Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology,
Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia

¹6025231092@student.its.ac.id, ²nanik@its.ac.id, ³chastine@if.its.ac.id

Abstract

Javanese script is an important part of Indonesia's cultural heritage, representing cultural values from the past. However, recognizing and classifying Javanese characters within manuscripts is challenging due to the limited availability of data and uneven distribution of character classes. The decline in formal use of Javanese script has drastically reduced the pool of manuscript samples, causing certain characters to appear rarely and skewing class frequencies. Existing methods that utilize Generative Adversarial Networks (GANs) attempt to address this problem. However, they often struggle to generate characters that are both consistent and visually accurate in terms of structural details. To address these issues, this study introduces a skeleton-based balancing GAN (SkelBAGAN), which improves the structural details of the previous method for generating characters. The proposed method introduces three main enhancements: (i) a layer for extracting the character skeleton structure, (ii) an optimized pretrained network using an autoencoder for learning the skeleton distribution, and (iii) refinement of the evaluation function, preserving both the distribution and structural fidelity in the adversarial process. The performance of the proposed model is evaluated against previous methods using the Fréchet Inception Distance (FID) to assess distribution quality and the Structural Similarity Index Measure (SSIM) to evaluate structural fidelity. The results indicate that the proposed methods outperform previous methods in balancing the FID and SSIM metrics. The integration of all enhancements in SkelBAGAN achieves the lowest FID, indicating improved generative quality while maintaining competitive SSIM values. The qualitative study indicates that SkelBAGAN outperforms previous methods in character generation. These results highlight how the skeleton-based improvement of the quality of generated characters enhances the recognition performance for underrepresented Javanese characters in imbalanced datasets. Ultimately, this work contributes to the broader effort to preserve the Javanese script as a vital element of Indonesia's cultural identity.

Keywords: character classification; data imbalance; generative adversarial networks; javanese manuscript; skeleton-based generation

How to Cite: M. 'Arif Faizin, N. Suciati, and C. Fatichah, "Handling Imbalance in Javanese Manuscript Character Dataset using Skeleton-based Balancing Generative Adversarial Networks", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 4, pp. 820 - 831, Aug. 2025.
Permalink/DOI: <https://doi.org/10.29207/resti.v9i4.6572>

Received: April 23, 2025

Accepted: August 8, 2025

Available Online: August 18, 2025

This is an open-access article under the [CC BY 4.0 License](#)
Published by [Ikatan Ahli Informatika Indonesia](#)

1. Introduction

The Javanese language is a local language mainly used in Java, Indonesia, which holds cultural and social values [1]. In the past, the Javanese language was written using Javanese script. It was actively used in literature and daily writing. However, its use has declined over time. Today, although the Javanese language is still often used in daily life, most of its use has switched to the Latin script due to several factors. Some of these factors are the dominance of the national language, limited use in education, technological advancements, lack of cultural awareness, and shifting preferences and lifestyles [2], [3]. These factors have

caused a decrease in the use of Javanese script in manuscripts, resulting in limited availability of Javanese script data [4].

This limited availability results in uneven character usage in manuscripts, causing several characters to appear far less frequently than others, leading to imbalanced distributions between characters. This imbalance can bias recognition models toward frequent characters, making it difficult to recognize the underrepresented ones [5].

To address this issue, various traditional methods have been employed, particularly using data resampling by applying minority oversampling, majority

undersampling, or hybrid approaches. Standard methods, such as random sampling, SMOTE, and cluster sampling, are widely used in implementing data resampling. However, these techniques have limitations; for example, the undersampling may discard useful data, while the oversampling can lead to model bias in learning, especially when the dataset has few samples in minority classes [6].

A more recent strategy for addressing this issue involves the use of generative models. Generative Adversarial Networks (GANs) are deep learning models designed to generate synthetic data by learning the distribution of real samples through adversarial training. A generator network creates data, while a discriminator evaluates its authenticity, allowing the model to produce increasingly realistic results. Compared to traditional augmentation techniques (flip, rotate, shear, scale, etc.) GANs can create more realistic, diverse, and unambiguous data compared with the traditional augmentation method [7] - [9]. This ability makes GANs well-suited for handling data imbalance. By generating new samples for underrepresented classes, GANs can improve the performance of recognition models [10].

The initial variation of GANs can generate a sample with its classes, such as Conditional GAN (cGAN) [11] and Auxiliary Classifier GAN (ACGAN) [12]. Both utilize class properties to create single GANs with multi-class capabilities [13]. Another variation of GANs namely Wasserstein GAN (WGAN) [14] focuses on improving the stability of learning using Wasserstein distance, which provides valuable learning curves. The improved WGAN with gradient penalty, called WGAN-GP [15] is proposed to handle the weight clipping problem that limits the learning ability of the critic [16].

To address the real-world multi-class data imbalance, more specific generative models have been developed. The Balancing GAN (BAGAN) [17] proposed a method using an autoencoder's encoder and decoder as a discriminator and a generator to explore the minority class distribution. BAGAN is specifically designed to address class imbalance in training datasets. BAGAN successfully generates sample images from MNIST, CIFAR-10, Flowers, and GTSRB datasets even when the minority classes are reduced by as little as 2.5% of the original dataset. Further research [18] proposed the Enhanced BAGAN that handles the weakness of BAGAN, such as instability in similar images and inseparable latent vectors between classes. The Enhanced BAGAN adapts the gradient penalty mechanisms from WGAN-GP to limit the continuous latent and uses the adaptation of the cWGAN-GP's [19] discriminator to eliminate dependency on the autoencoder's encoder. These improvements allow Enhanced BAGAN to handle imbalanced datasets more effectively.

Moreover, the variety of handwriting styles in Javanese script characters increases the complexity of the script, making the recognition more complicated [20]. Some studies have used structural features to address the different styles [21]. The structural features are essential for extracting the character features and making characters easier to recognize. Some studies use skeletonization as the main method of structural extraction in character recognition [22]. The use of skeletons in character recognition shows better results by capturing fine-grained details. The adaptation of the skeleton on GAN using L1 loss in SkelGAN [23] shows good results for learning the structural features.

To address the imbalance distribution and improve its structural features, this study proposes a skeleton-based GAN method to enhance the structural shape details of Javanese characters while handling data imbalance. By extracting and enforcing the skeleton consistency through additional network and loss function, the proposed method aims to improve both structural fidelity and class-specific generation quality.

The main improvement in this research can be described in three points: first, it introduces the use of GAN-based augmentation for generating Javanese manuscript characters. As this research is written, no research uses generative-based data as the augmentation method in this case. Second, an improvement on the structural similarity of the previous works by increasing the structural similarity of the handwriting using skeletonization. Third, this study provides a detailed analysis by examining the data characteristics, quantitative results, and qualitative quality of the generated samples.

This research aims to contribute to the development of data augmentation methods that enhance the classification models of the Javanese script in manuscripts.

2. Methods

The research is divided into three focuses: the review of Enhanced Balancing GAN (Enhanced BAGAN) Architecture, which focuses on main points of the previous works, the proposed enhancement with SkelBAGAN, which explains the adaptation and use of skeleton on the existing architecture within refinement on the loss function, and the experimental environment and case for the evaluation.

2.1 Enhanced BAGAN Architecture

The previous version of Enhanced Balancing GAN (Enhanced BAGAN) [18] implements the two steps for training the GANs, named autoencoder initialization and balanced training in GANs.

The basic idea of the Balancing GAN (BAGAN) [17] is to learn the distribution of imbalanced data through the decoder and encoder in the autoencoder. The decoder and encoder are trained on the entire image to build

common knowledge about the dataset within each class, serving as the initialization steps.

The encoder uses convolutional layers to extract the features of the image into a latent vector, while the decoder uses the transposed convolutional layer combined with normalization to reconstruct the latent vector as a reconstructed image. Different from the original BAPAN, the Enhanced BAPAN uses a decoder to generate latent vectors from real images within the label to be passed to an embedding layer. The embedding layer generates a labeled latent vector as the decoder's input. These are the different parts from the

original BAPAN, which initialize the labeled latent vector before training the GAN. The decoder translates the labeled vector into a reconstructed image. The Enhanced BAPAN minimizes the L1 loss between real images and reproduced images for training the autoencoder networks. The L1 loss minimizes the pixel-by-pixel difference between the real images and reconstructed images, producing sharp images [24]. Unlike the original BAPAN, the label embedded in the layer provides information about the classes to the decoder, enabling the generator to generate supervised images. The illustration of initialization steps in Enhanced BAPAN is shown in Figure 1.

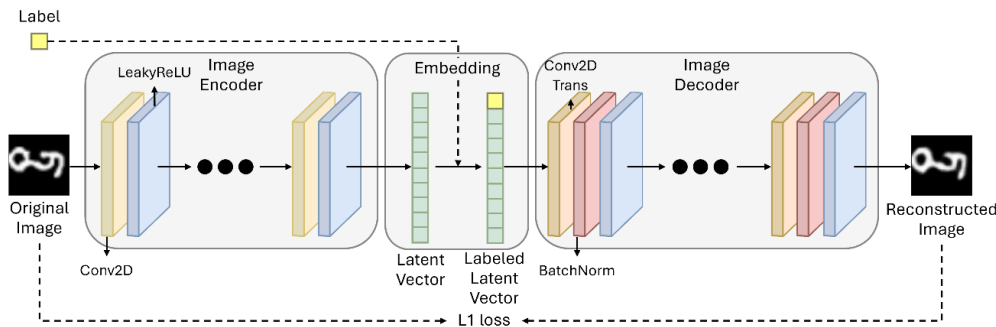


Figure 1. The initialization steps in Enhanced BAPAN

By inheriting the same structure and weights from the trained decoder, the enhanced GAN generator obtains information from the trained autoencoder. The generator is retrained with the discriminator using an adversarial loss function of real and fake images.

The label and random latent vector are used to produce a labeled latent vector using an embedding layer, which then generates fake images with the pre-trained image decoder. Unlike the original BAPAN, the discriminator tries to eliminate the dependency on inheriting the

trained encoder and uses the cWGAN-GP [19] discriminator architecture by using the validity score from the combination of features extracted from the embedded label and the simple convolutional encoder of the image. The Enhanced BAPAN employs gradient-penalty adaptation from [25] to constrain the gradient to a determined value. This implementation introduces a training ratio between the discriminator and generator bigger than one to train the discriminator more convergently. An illustration of the training steps in Enhanced BAPAN is shown in Figure 2.

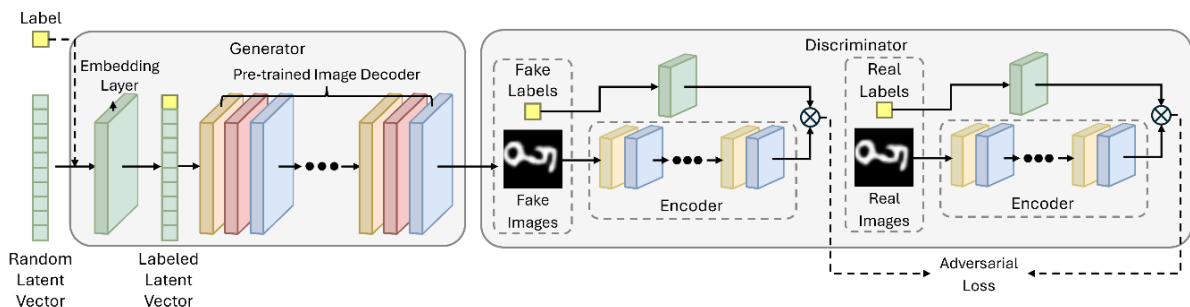


Figure 2. The training steps in Enhanced BAPAN

2.2 Proposed Enhancements with SkelBAPAN

This work proposes Skeleton-enhanced BAPAN (SkelBAPAN), which adapts the Enhanced BAPAN [18] as the baseline. SkelBAPAN proposes three main enhancements compared with the baseline, the skeleton extraction layer, a pre-trained skeleton autoencoder, and a modification of the skeleton-based refinement loss. The changes are primarily aimed at enhancing the structural similarity of the generated images.

Figure 3 shows the proposed method's autoencoder initialization. A specialized layer that extracts the real image or generated image into a skeleton is proposed using Zhang-Suen thinning algorithm [26]. These generated skeletons are used to gain both distribution and structural features. The extracted skeletons are used in the skeleton autoencoder, which is trained together with the image autoencoder. This specialized

autoencoder on skeletons learns the distributional features of the skeletons during the initialization steps.

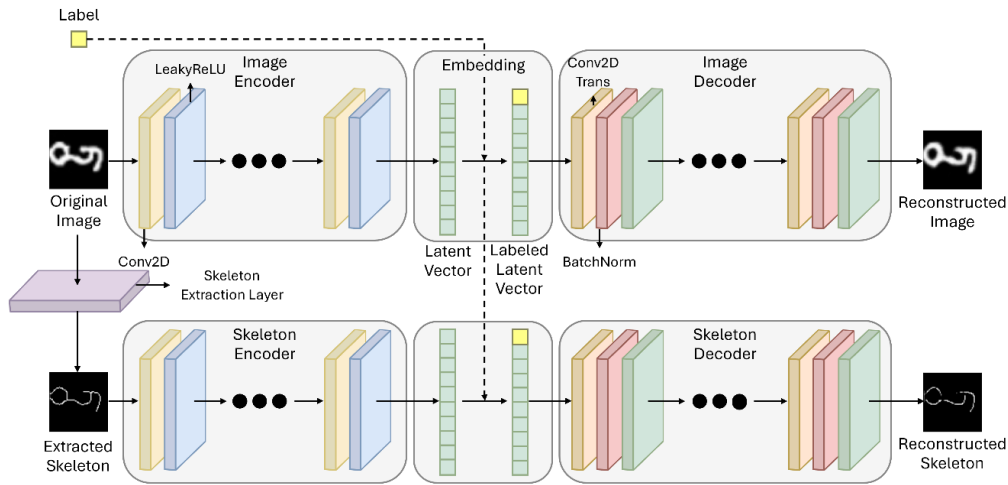


Figure 3. The autoencoder initialization steps for image and skeleton in proposed SkelBAGAN

Figure 4 illustrates the proposed method's adversarial training steps. There are the discriminator and the skeleton discriminator, both using a module using cWGAN-GP [19] discriminator architecture for adversarial purposes. This module uses features extracted from the embedded label and the convolutional encoder of the image. Adversarial loss is used in the discriminator, and the skeleton adversarial

loss is used in the skeleton discriminator. Meanwhile, the generator employs a skeleton adversarial loss, combined with an L1 loss, to preserve the distribution and structural quality of the generated images during adversarial training. Moreover, the detailed adversarial equation used in the training is explained in Section 2.2.3.

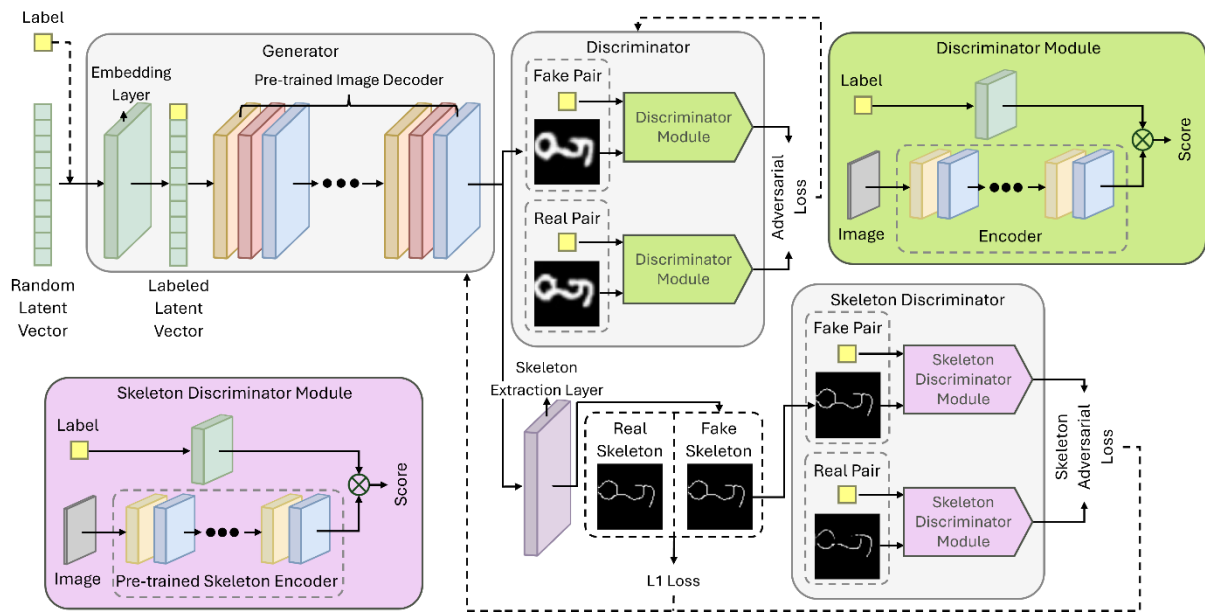


Figure 4. The training steps in proposed SkelBAGAN

2.2.1. Skeleton Extraction Layer

To preserve the skeleton information from the images, SkelBAGAN uses an extraction layer in the pipeline. The extraction layer is used in both autoencoder training and the adversarial learning process. This layer converts the pixel-based outputs of the generator into skeleton representations. The extracted skeletons are then used as input to the skeleton-specific discriminator, which evaluates the quality and fidelity of the generated

skeletons. This layer receives the pre-trained decoder layer, which outputs the generated image from the previous layers.

The skeleton extraction layer uses the Zhang-Suen thinning algorithm [26] for skeleton extraction. The Zhang-Suen thinning algorithm was chosen due to its computational efficiency, which utilizes simple Boolean operations, and its fast parallel processing capability that enhances speed during training,

compared to other methods [27]. The Zhang-Suen thinning algorithm could preserve the general shape and stroke continuity of binary character images, which is crucial for guiding skeleton-based generations. These additional layers are used to prevent the usage of skeletons in the autoencoder process that might interfere with the distribution learning process. The use of skeletons helps to provide small but useful features in recognizing characteristics, even on poor quality images [28].

2.2.2. Pre-trained Skeleton Autoencoder

SkelBAGAN proposes an additional autoencoder for learning the structural information from the image distribution. Inspired by Enhanced BAGAN [18], which uses a pre-trained autoencoder that learns the distribution between real images and generated images in an imbalanced manner. The Skeleton Enhanced GAN [26] utilizes a similar approach to enhance the structural qualities of the generated image, which is then separated by the image discriminator. The pre-trained skeleton autoencoder is attached to the autoencoder training and the discriminator.

Instead of using the mean absolute error loss when training the autoencoder, this paper proposed using the dice coefficient loss to measure the overlap between two binary skeleton lines. This loss is claimed to be more sensitive than the mean absolute error loss. For S_r as the real ground truth skeleton and S_g as the generated skeleton of the generated image, the dice coefficient is defined as in Equation 1.

$$\mathcal{L}_{dice}(S_r, S_g) = 1 - \frac{2 \times S_r \times S_g}{S_r + S_g} \quad (1)$$

The skeleton autoencoder also uses the embedding function to make different latent for each class, which will help the discriminator process to improve the loss. The proposed skeleton autoencoder is shown in Figure 3.

2.2.3. Skeleton-based Refinement Loss

SkelBAGAN proposes an enforcement of structural consistency in generating Javanese characters by explicitly guiding the generator to match the skeleton features of real samples. This can be achieved by integrating skeleton loss into adversarial processes to have a better effect on the generated characters.

SkelBAGAN still relies on the core of the adversarial framework of GANs, where a generator G and discriminator D compete in a min-max game. For X_r probability distribution of real images, X_g probability distribution of generated images, x_r sampled real images, and x_g sampled generated images, where $x_g = G(z)$ and z is a sample of random noise taken from normal distribution $z \sim N(0, I_{dim(z)})$, then the value function of GAN is defined as in Equation 2.

$$\min_G \max_D V(G, D) = \mathbb{E}_{x_r \sim X_r} [\log(D(x_r))] + \mathbb{E}_{x_g \sim X_g} [\log(1 - D(x_g))] \quad (2)$$

This function loss then builds on a series of GAN advancements. WGAN [14] introduces the Wasserstein distance loss for achieving more stability when training GAN. The WGAN approach then further improved training smoothness and 1-Lipschitz enforcement through gradient penalties in WGAN-GP [15]. The draGAN [29] then adapts the gradient penalty to the log-sigmoid loss, enhancing the “critic” robustness.

Meanwhile, the advancements in handling imbalanced data in GAN were introduced by Balancing GAN (BAGAN) [17] by adding autoencoder initialization to shape a class-aware latent shape. The Enhanced BAGAN [18] adapts this concept and combines the cWGAN-GP’s discriminator [19], gradient penalties, and auxiliary classifier loss for better stability and class separation.

Enhanced BAGAN proposed the additional cross-entropy loss of misclassified classes by selecting a random fake label from the balanced label Y_f for each fake image to strengthen the class information learning. For z is a sample of a random noise vector from a normal distribution where $z \sim \mathcal{Z}$, $z \sim N(0, I_{dim(z)})$, y_f is the fake labels $y_f \sim Y_f$, $y_f \sim U\{0, 1, 2, \dots\}$, and y_w is the fake wrong random sampling label where $y_w \sim Y_w$, $y_w \sim U\{0, 1, 2, \dots\}$, then the overall Enhanced BAGAN discriminator or critic loss is defined as in Equation 3. With the same approach, the definition of generator loss is defined as in Equation 4.

$$\begin{aligned} \max_C \mathcal{L}_C(X_r, Z, Y_r, Y_f, Y_w) &= -\mathbb{E}_{(x_r, y_r) \sim (X_r, Y_r)} [\log(C(x_r, y_r))] \\ &- \mathbb{E}_{(z, y_f) \sim (Z, Y_f)} [\log(1 \\ &- C(G(z, y_f), y_f))] \\ &- \mathbb{E}_{(x_r, y_w) \sim (X_r, Y_w)} [\log(1 \\ &- C(x_r, y_w))] + \mathcal{L}_{gp} \end{aligned} \quad (3)$$

$$\min_G \mathcal{L}_G(Z, Y_f) = -\mathbb{E}_{(z, y_f) \sim (Z, Y_f)} [\log(C(G(z, y_f)))] \quad (4)$$

Although Enhanced BAGAN is improving the former BAGAN, the approach is still limited to general problems [30]. Enhanced BAGAN uses the basic autoencoder as the initialization training. Better initialization may be focused on the problem of specific data, such as feature-based or statistical-based. In the Javanese character classification problem, the important main feature that might influence is structural features, which distinguish between the classes.

SkelGAN [23] and Skeleton Enhanced GAN [31] successfully transfer the style and content of the character using the skeleton-based feature to learn the structural features. SkelBAGAN proposed adapting the

L1 loss from SkelGAN to compare the real skeleton with the generated skeleton, which is more reliable for skeleton comparison compared to the L2 loss, which is blurrier. The L1 loss is applied in the generator instead of in the discriminator (referenced from [32]) to guide the generator to create an image that is structurally consistent with minimum distance from the real image. For λ_{L1} as the weight of L1 loss, s_r is the real or ground truth skeleton of the real image, and S is the function that extracts the generated image into skeleton and $H(z, y)$ is the extracted skeleton from the generated image $H(z, y) = S(G(z, y))$, then the skeleton loss is defined as in Equation 5. The generator loss is then defined as in Equation 6.

$$\mathcal{L}_{skel}(Z, Y_f) = \lambda_{L1} \mathbb{E}_{(s_r, z, y_f) \sim (s_r, Z, Y_f)} [\|s_r - H(z, y_f)\|_1] \quad (5)$$

$$\min_G \mathcal{L}_G(Z, Y_f) = -\mathbb{E}_{(z, y_f) \sim (Z, Y_f)} [\log(C(G(z, y_f)))] + \mathcal{L}_{skel} \quad (6)$$

To improve the understanding of the generator on the skeleton, this paper proposed a skeleton adversarial loss applied in the generator. The skeleton adversarial loss as defined in Equation 7 is formulated by introducing a separate discriminator trained to distinguish between real and generated skeletons. This discriminator is explained in the next section. The generator loss is updated as in Equation 8.

$$\mathcal{L}_{adv}^{sk}(Z, Y_f) = \lambda_{sk} (\mathbb{E}_{(s_r) \sim (s_r)} [\log(D_{sk}(s_r))] + \mathbb{E}_{(z, y_f) \sim (Z, Y_f)} [\log(1 - D_{sk}(H(z, y_f)))] \quad (7)$$

$$\min_G \mathcal{L}_G(Z, Y_f) = -\mathbb{E}_{(z, y_f) \sim (Z, Y_f)} [\log(C(G(z, y_f)))] + \mathcal{L}_{skel} + \mathcal{L}_{adv}^{sk} \quad (8)$$

2.3 Experiments

In this study, several key hyperparameters are configured for training the models. The latent dimensions of the encoder and decoder are set to 128. The training between the discriminator and the generator was conducted similarly to the training of the skeleton discriminator and the skeleton generator, with a ratio of 5:1. The image autoencoder was trained for 30 epochs. In comparison, the skeleton autoencoder is trained for 50 epochs. Both autoencoders are trained using a learning rate of 0.0002, with beta1 set to 0.5 and beta2 set to 0.9, and a batch size of 128. The gradient penalty loss weight is set to 1, and the skeleton L1 loss is fixed at 0.1, and the skeleton adversarial loss is set to 0.1. The overall adversarial training consists of 50 steps per batch, each for two epochs. The models are trained with consistent hyperparameters, and evaluation is carried out using the test image set. Almost all these

parameters were initially chosen based on the implementation of Enhanced BAKAN in [33], with the rest determined based on experiments.

The proposed method is evaluated against the Enhanced BAKAN as the baseline method. There are three variants of the proposed method, which utilize skeleton L1 loss, skeleton autoencoder, and skeleton adversarial loss. The purpose is to evaluate the advantages and the drawbacks of the different components in enhancing the quality of generated images. The evaluation is applied by comparing the test image set and a sample of 800 generated images.

These models are evaluated using two metrics: FID (Fréchet Inception Distance), which assesses distribution-based quality, and SSIM (Structural Similarity Index), which evaluates the structural similarity of the generated images. The FID is calculated by the mean and covariance of feature vectors from a pre-trained Inception V3 network for real and generated images. For μ_p, Σ_p and μ_q, Σ_q as the mean and covariance of the real and generated image features, the FID score is given as in Equation 9. The lower FID represents better quality and variety of the generated images.

$$FID(p, q) = \|\mu_p - \mu_q\|_2^2 + \text{Tr}(\Sigma_p + \Sigma_q - 2(\Sigma_p \Sigma_q)^{\frac{1}{2}}) \quad (9)$$

The SSIM score measures the perceptual similarity between two images in terms of luminance, contrast, and structural features. For μ_x, μ_y as the mean intensities, σ_x, σ_y as the standard deviations, and σ_{xy} is the covariance of the images x and y , the SSIM score is defined as in Equation 10. The higher SSIM values indicate better structural similarity between the two images [34].

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (10)$$

3. Results and Discussions

The results of the research are mainly directed to the problem of augmenting the dataset using SkelBAKAN to improve its structural similarity. This research aims to provide an in-depth analysis by examining the Javanese character dataset used, analyzing both the quantitative and qualitative results of the proposed method. A discussion of the proposed method is provided to answer the research questions.

3.1 Javanese Characters Dataset

This study utilizes the *Handwritten Javanese Character Sewaka Detection* (HJCS_DETC) dataset, a Javanese character dataset from our previous research [35], [36] that initially designed for object recognition tasks. This dataset was collected from a digitized manuscript called "Sewaka". It is part of the national library's public collection of manuscripts from Indonesia. The dataset

was labeled and annotated by Javanese experts. This dataset is available for limited access in Zenodo [37].

The dataset consists of YOLO-labeled annotations collected from 60 images, representing 74 classes of the Javanese script. This study focuses on a subset of 20 basic characters from the *wianjana* category, which are fundamental to the Javanese script. Figure 5 illustrates a sample of character positions and the number of characters in a sample of a page.

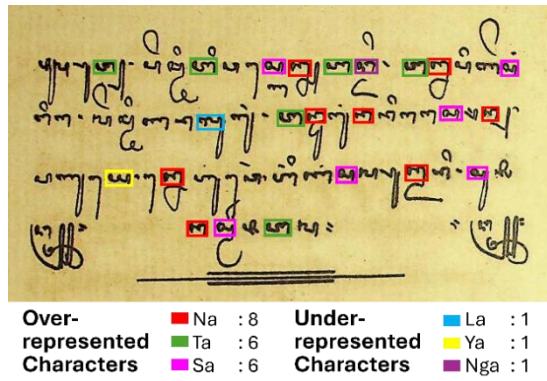


Figure 5. Illustration between over-represented and under-represented sample characters in HJCS_DET C

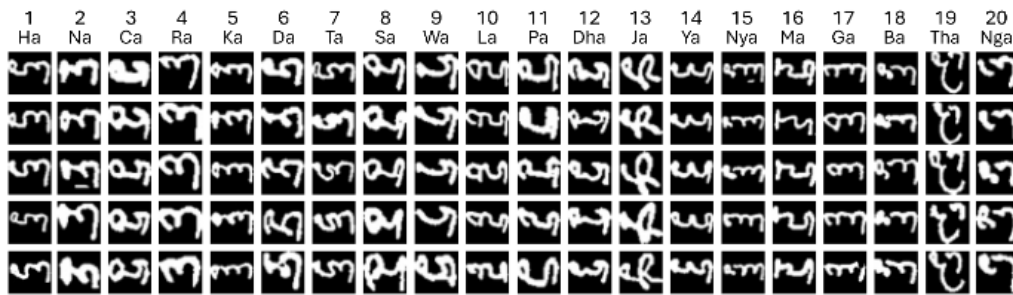


Figure 6. Sample real images of HJCS_DET C dataset

The distribution of the dataset is illustrated in Table 1, which shows the uneven distribution of the Javanese character datasets. Let N_{max} represent the sample number in the majority classes and N_i represent the sample number in the class i , then the Imbalance Ratio (IR) of each class is defined in Equation 11. The value of IR in minority classes shows that the dataset is imbalanced. Using the definition of Imbalance Degree (ID) in [39], for d_{emp} is the empirical Euclidean distance and d_{im} is the worst-case distance among the m minority classes, then ID is defined as in Equation 12. The calculated ID of train and test dataset are 9.33 and 10.33 respectively, showing that the dataset is mid-imbalanced.

$$IR(i) = \frac{N_{max}}{N_i} \quad (11)$$

$$ID = \frac{d_{emp}}{d_{im}} + (m - 1) \quad (12)$$

This dataset features variations in the handwriting styles of the characters, including differences in thickness,

Figure 5 also shows the different number of characters between over-represented characters and under-represented characters, indicating that the dataset is imbalanced.

Every character is taken as a separate image and then processed to ensure data consistency. Preprocessing steps involve converting images to grayscale, applying a Gaussian blur, and using binary thresholding to improve contrast. These steps reduce complexity and ensure consistency, as color information is not relevant for handwritten script. The process simplifies data distribution and guides the generator to learn the structures and the main shapes. The characters are resized to 32×32 pixels with added padding to center-align them within the frame. This process ensures uniformity and prepares the dataset for effective use in the generative framework. Figure 6 shows samples of preprocessed real images. The visual results show that the conversion of the color features from RGB into grayscale makes characters clearer and recognizable. The grayscale conversion helps simplify and reduce the complex computation in generating characters [38].

size, connectivity, and distortion. Figure 7 shows several samples of the different variations in the dataset.

Table 1. Class distribution of 20 characters in HJCS_DET C dataset

Class id-name	Train Data Num. of image	Imbalance Ratio	Test Data Num. of image	Imbalance Ratio
1-Ha	812	2.32	71	2.48
2-Na	1881	1.00	176	1.00
3-Ca	116	16.22	6	29.33
4-Ra	590	3.19	61	2.89
5-Ka	1119	1.68	105	1.68
6-Da	454	4.14	40	4.40
7-Ta	713	2.64	79	2.23
8-Sa	911	2.06	81	2.17
9-Wa	606	3.10	46	3.83
10-La	658	2.86	53	3.32
11-Pa	616	3.05	60	2.93
12-Dha	74	25.42	5	35.20
13-Ja	193	9.75	11	16.00
14-Ya	349	5.39	30	5.87
15-Nya	72	26.13	7	25.14
16-Ma	756	2.49	53	3.32
17-Ga	353	5.33	29	6.07
18-Ba	271	6.94	23	7.65
19-Tha	23	81.78	5	35.20
20-Nga	408	4.61	24	7.33

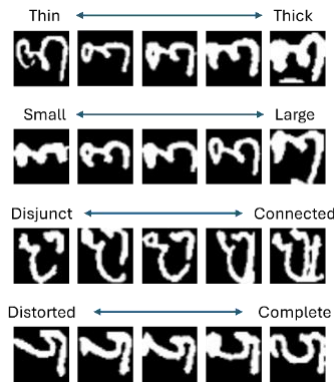


Figure 7. Sample of variety in historical styles of the characters

3.2 Quantitative Results

To perform a detailed quantitative evaluation comparing the baseline model (Enhanced BAKAN [18]) with the proposed model (SkelBAGAN), this comparison defines several versions of the SkelBAGAN: The first version (v1) implements the SkelBAGAN using a skeleton loss function (L1 loss); The second version (v2) builds on v1 by incorporating a pre-trained skeleton autoencoder; The third version (v3) further enhances v2 by adding a skeleton adversarial loss.

This comparison will also include an assessment of the real image score against the FID and SSIM target score, which represents the distribution of real and desired data. Table 2 presents a comparison of the average FID and SSIM scores between the models.

Table 2. Average FID and SSIM scores between the models

Method	FID ↓	SSIM ↑
Real images	121.26	0.546
Enhanced BAKAN [18]	189.50	0.511
SkelBAGAN v1	213.85	0.552
SkelBAGAN v2	184.24	0.549
SkelBAGAN v3	176.12	0.547

As shown in Table 2, the baseline Enhanced BAKAN achieves a moderately low FID score of 189.50. This

score indicates that the generated images are not sufficiently close to the real data distribution. The introduction of L1 loss in SkelBAGAN v1 results in a higher FID score of 213.85, meaning that the model focuses more on the structural alignment than the overall distribution. The addition of the skeleton autoencoder in SkelBAGAN v2 results in a significant reduction of the FID score to 184.24, illustrating that the autoencoder effectively enhances the alignment between the generated image and the real image distributions. The inclusion of the skeleton adversarial loss in SkelBAGAN v3 achieves the best FID of 176.12 among the tested methods, demonstrating that adversarial training on skeleton representations substantially improves the alignment of the generated image and real image distributions.

This highlights the importance of adversarial training on skeletons in refining generated sample distributions, bringing them more similar to real images. The detailed FID scores of each class are compared in Figure 8.

The detailed comparison of SSIM scores in each class is shown in Figure 9. The SSIM score provides a different perspective, focusing on the structural fidelity of the generated images. The baseline Enhanced BAKAN achieves relatively low SSIM with a score of 0.511, reflecting its limitations in preserving structural details. The introduction of skeleton L1 loss in SkelBAGAN v1 yields a higher SSIM score of 0.552, demonstrating the direct impact of skeleton L1 loss on maintaining character structure. However, the induction of a pre-trained skeleton autoencoder in SkelBAGAN v2 achieves a slightly lower SSIM of 0.549, likely due to the balance between structural accuracy and distribution alignment. Similarly, the use of skeleton adversarial loss in SkelBAGAN v3 slightly decreases SSIM to 0.547, indicating that while the adversarial loss enhances distribution alignment, it introduces a minor trade-off in structural fidelity.

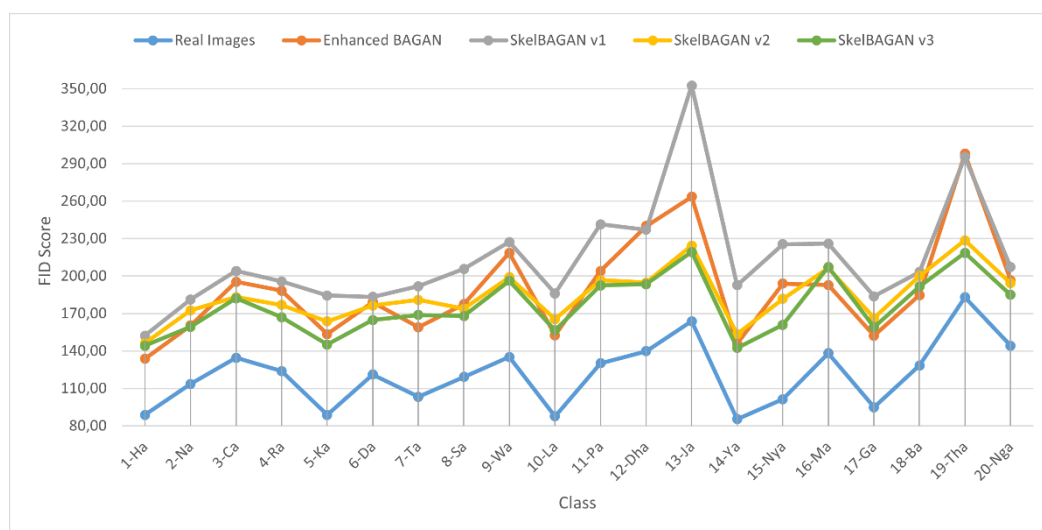


Figure 8. Class-based comparison of FID scores between the models

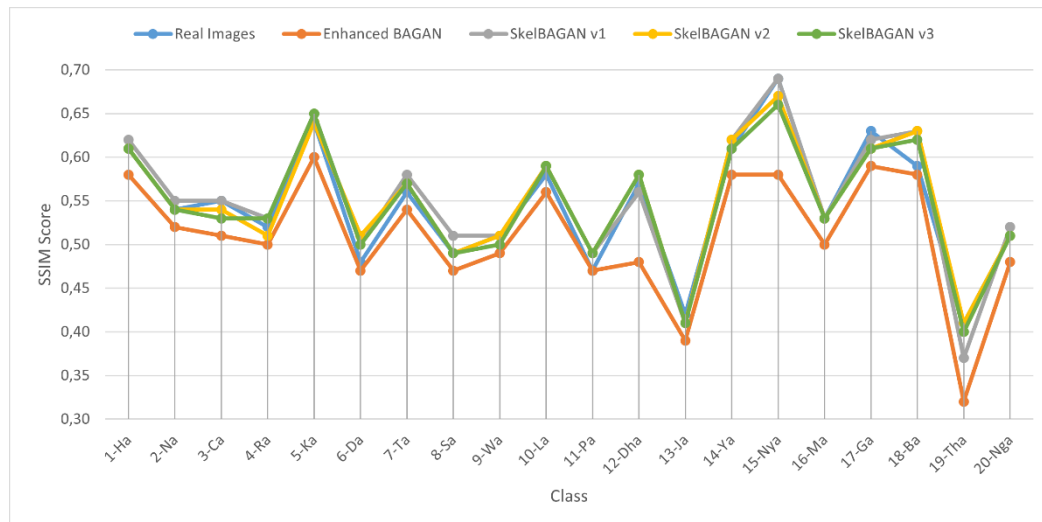


Figure 9. Class-based comparison of SSIM scores between the models

The results reveal a trade-off between FID and SSIM. While SkelBAGAN v1 resulting in the highest SSIM, SkelBAGAN v3 achieves improved FID with a slight reduction in SSIM. Since FID was sensitive to noise and SSIM was incorporating structural information [40], this trade-off highlights the need to balance distributional alignment and noise with structural fidelity, depending on the requirements. For tasks prioritizing structural preservation, SkelBAGAN v1 might be preferable, while SkelBAGAN v3 is more suitable for generating diverse and realistic datasets.

3.3 Qualitative Results

The comparison between the results of each model was presented in Table 3. A visual comparison between several models reveals slight visual differences, particularly in the “Da”, “Ja”, and “Tha” classes. This result validates the advantages of SkelBAGAN in enhancing structural features.

In addition to the quantitative evaluation, the qualitative assessment was conducted to focus on classes that exhibited lower SSIM scores in the quantitative results, specifically: 6-Da, 8-Sa, 11-Pa, 13-Ja, and 19-Tha as visualized in Figure 10. The red circle highlights the loss of the skeleton structures, while the yellow circles indicate the addition of incorrect structural elements in some classes.

Visual examination of the generated images between the models reveals the proposed SkelBAGAN models (v1, v2, and v3) consistently maintain the structural details of the characters. In contrast, the baseline Enhanced Bagan struggled to accurately reproduce the structure of these characters, often generating corrupt or misrepresented characters. The red circles in Figure 10 show where the baseline model lost critical skeletal features, failing to preserve essential structures. The yellow circles highlight the incorrect or extraneous elements, distorting the intended character shape. These examples demonstrate the baseline model's limitation in handling complex structural details effectively, showing the advantages of the proposed skeleton-based regularization methods. The addition of a skeleton autoencoder and skeleton adversarial loss in SkelBAGAN v3 enhances the generated image quality by refining both the global and local characteristics of the characters.

To complement FID and SSIM, a user study is conducted involving both expert and non-expert participants, tasked with ranking character images based on realism. For each character class, participants were shown five randomly ordered images: one real image, and one from each of Enhanced-BAGAN, SkelBAGAN v1, v2, and v3. Participants ranked the images from most to least realistic.

Table 3. Comparison between sample of real images, Enhanced Bagan, SkelBAGAN v1, v2, and v3

Method	Ha	Na	Ca	Ra	Ka	Da	Ta	Sa	Wa	La	Pa	Dha	Ja	Ya	Nya	Ma	Ga	Ba	Tha	Nga
Real images																				
Enhanced Bagan																				
SkelBAGAN v1																				
SkelBAGAN v2																				
SkelBAGAN v3																				

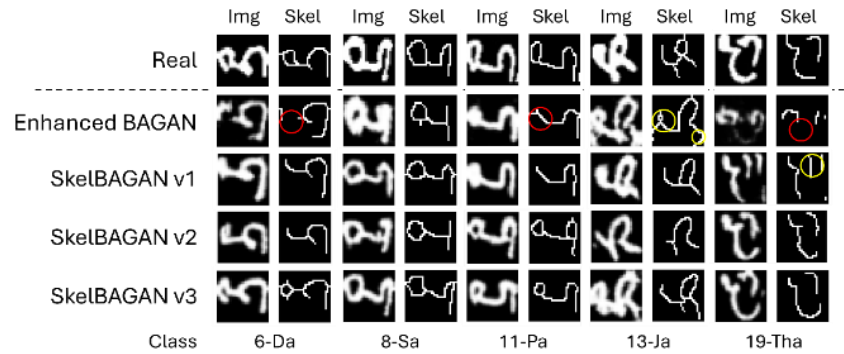


Figure 10. Comparison between sample real images, Enhanced BAKAN, and proposed SkelBAGAN

Figure 11 shows the results of the user study, which yielded 1,380 completed rankings (3 samples $\times$ 20 classes per respondent) with 23 respondents. The frequencies of each method are calculated using ranks 1–5, with rank 1 being the most realistic and rank 5 being the least realistic. The results of the study show that SkelBAGAN v2 received the highest number of Rank 1 votes with 318, surpassing the Enhanced-BAGAN, which received 307 Rank 1 votes, and the other methods evaluated. In addition, SkelBAGAN v3 yields the best results in the top three ranks compared to all other methods, with a total of 879 votes consisting of 271 votes in Rank 1, 301 votes in Rank 2, and 307 votes in Rank 3. These results indicate that SkelBAGAN v2 and v3 are the most effective approaches in terms of qualitative evaluation, and that the proposed methods are highly compatible with human perceptual preferences.

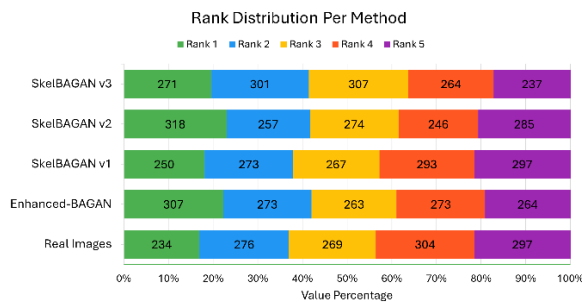


Figure 11. Rank distribution on user study results

3.4 Discussions

The results from the study demonstrate the effectiveness of SkelBAGAN as a GAN-based augmentation method in addressing data imbalance in Javanese manuscript characters, improving FID and SSIM scores by 176.12 and 0.552, compared to previous work using Enhanced BAKAN [18], which achieved improvements of 189.50 and 0.511. This improvement shows that the proposed method is feasible for generating characters in Javanese manuscripts. For Javanese character classification or recognition, the SkelBAGAN can be used to augment the data with an imbalanced distribution. Qualitative examinations in human perception show that the proposed method is more realistic than the previous

method with 879 votes to 843 votes in the top three ranks. These observations confirmed that the GAN-produced samples closely resemble the original authentic manuscript characters in real-world scenarios, which are critical for recognition tasks such as digitization and restoration of manuscripts.

However, while the method shows potential improvements, some limitations remain. The usage of the skeletonization process and the additional autoencoder layer introduces extra complexity, which is not good and could limit the scalability of the approach in more complex tasks, such as large-scale or real-time applications. Moreover, some unexpected results in terms of variability between the results of different character classes require further investigation to determine the optimal balance between real and synthetic data.

In terms of broader societal impact, this research contributes to Sustainable Development Goal 4 (Quality Education) by supporting the creation of digital educational resources that include underrepresented scripts, enabling broader access to regional cultural knowledge. It also aligns with SDG 11 (Sustainable Cities and Communities) by promoting cultural preservation through digital restoration and accessibility of ancient manuscripts. By enabling more accurate and scalable digitization methods, this work helps safeguard intangible heritage for future generations.

Future work may explore further refinements in adversarial training and consider additional forms of regularization to enhance the visual quality of the generated images. Additionally, future works may explore the development of simpler architectures optimized for generating high-quality samples with reduced computational complexity.

4. Conclusions

This paper proposes a novel method to improve the generation of Javanese manuscript characters using several skeleton-based enhancements of the previous techniques called Skeleton Balancing GAN (SkelBAGAN). The proposed approach incorporated a skeleton extraction layer, a pre-trained skeleton

autoencoder, and skeleton-based loss refinements, including skeleton L1 loss and skeleton adversarial loss to overcome the challenges of structural preservation and class imbalance in the dataset.

The experimental results demonstrated that the SkelBAGAN outperforms the baseline model in both FID (Fréchet Inception Distance) and SSIM (Structural Similarity Index) metrics, indicating better distributional similarity and structural fidelity. The combination of the enhancement components led to an improvement in the quality of the generated images by increasing the structural similarity and maintaining the class distribution. The analysis of data characterization and evaluation on both quantitative and qualitative aspects validates SkelBAGAN's effectiveness in generating samples for underrepresented classes. This study contributes a promising direction for handling data imbalance in historical manuscript datasets using GANs.

Future research could explore further optimizations in adversarial training and skeleton extraction to refine the model's performance. Additionally, expanding the approach to other types of imbalanced datasets could further validate their versatility and applicability across different domains.

Acknowledgements

This research was supported by the Ministry of Higher Education, Science, and Technology of Indonesia through the Master's Thesis Research Program, implemented by Institut Teknologi Sepuluh Nopember, under agreement number 1130/PKS/ITS/2025.

References

- [1] S. Anistya Ori and H. W. Susianti, "Preserving Javanese Language and Culture in The Digital Age: Challenge and Future Prospects," *LACULTOUR: Journal of Language and Cultural Tourism*, vol. 2, no. 2, pp. 79–88, Nov. 2023, doi: 10.52352/lacultour.v2i2.1212.
- [2] A. B. Perdana, "Diverse Scripts, Same Mistakes: Identifying Common Pitfalls in the Typographic Designs of Indonesian Traditional Scripts," *Nirmana*, vol. 24, no. 2, pp. 167–178, Jul. 2024, doi: 10.9744/nirmana.24.2.167-178.
- [3] M. A. Rasyidi, T. Bariyah, Y. I. Riskajaya, and A. D. Septyani, "Classification of handwritten Javanese script using random forest algorithm," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3, pp. 1308–1315, Jun. 2021, doi: 10.11591/eei.v10i3.3036.
- [4] D. Iskandar, S. Hidayat, U. Jamaludin, and S. Mukti Leksono, "Javanese script digitalization and its utilization as learning media: an etnopedagogical approach," *International Journal of Mathematics and Sciences Education*, vol. 1, no. 1, pp. 21–30, Jun. 2023, doi: 10.59965/ijmsed.v1i1.24.
- [5] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of Imbalanced Data: Review of Methods and Applications," *IOP Conf Ser Mater Sci Eng*, vol. 1099, no. 1, p. 012077, Mar. 2021, doi: 10.1088/1757-899X/1099/1/012077.
- [6] S. Cicak and U. Avcı, "Handling Imbalanced Data in Predictive Maintenance: A Resampling-Based Approach," in *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, IEEE, Jun. 2023, pp. 1–6, doi: 10.1109/HORA58378.2023.10156799.
- [7] J. Jason Jeong, B. Patel, and I. Banerjee, "GAN augmentation for multiclass image classification using hemorrhage detection as a case-study," *Journal of Medical Imaging*, vol. 9, no. 03, Jun. 2022, doi: 10.1117/1.JMI.9.3.035504.
- [8] M. Tomaszewski and J. Osuchowski, "Effectiveness of Data Resampling in Mitigating Class Imbalance for Object Detection," in *Proc. ITTAP '23: 3rd International Workshop on Information Technologies: Theoretical and Applied Problems*, Nov. 2023, pp. 197–07.
- [9] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image Data Augmentation for Deep Learning: A Survey," *arXiv preprint*, Apr. 2022, doi: 10.48550/arXiv.2204.08610.
- [10] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network," *Remote Sens (Basel)*, vol. 12, no. 9, p. 1432, May 2020, doi: 10.3390/rs12091432.
- [11] P. Sharma, M. Kumar, H. K. Sharma, and S. M. Biju, "Generative adversarial networks (GANs): Introduction, Taxonomy, Variants, Limitations, and Applications," *Multimed Tools Appl*, vol. 83, no. 41, pp. 88811–88858, Mar. 2024, doi: 10.1007/s11042-024-18767-y.
- [12] R. Nagaraju and M. Stamp, "Auxiliary-Classifer GAN for Malware Analysis," *arXiv Preprint*, Jul. 2021.
- [13] A. Bourou, V. Mezger, and A. Genovesio, "GANs Conditioning Methods: A Survey," *arXiv preprint*, Aug. 2024, doi: 10.48550/arXiv.2408.15640.
- [14] T. Chakraborty, U. Reddy K S, S. M. Naik, M. Panja, and B. Manvitha, "Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art," *Mach Learn Sci Technol*, vol. 5, no. 1, p. 011001, Mar. 2024, doi: 10.1088/2632-2153/ad1f77.
- [15] G. Zhu, K. Zhou, L. Lu, Y. Fu, Z. Liu, and X. Yang, "Partial Discharge Data Augmentation Based on Improved Wasserstein Generative Adversarial Network With Gradient Penalty," *IEEE Trans Industr Inform*, vol. 19, no. 5, pp. 6565–6575, May 2023, doi: 10.1109/TII.2022.3197839.
- [16] B. Kamsu-Foguem, S. L. Msouobu Gueuwou, and C. A. K. A. Kounta, "Generative Adversarial Networks based on optimal transport: a survey," *Artif Intell Rev*, vol. 56, no. 7, pp. 6723–6773, Jul. 2023, doi: 10.1007/s10462-022-10342-x.
- [17] V. Sampath, I. Maurtua, J. J. A. Martín, and A. Gutierrez, "A Survey on Generative Adversarial Networks for Imbalance Problems in Computer Vision Tasks," *J Big Data*, Aug. 2020, doi: 10.21203/rs.3.rs-45616/v1.
- [18] G. Huang and A. H. Jafari, "Enhanced balancing GAN: minority-class image generation," *Neural Comput Appl*, vol. 35, no. 7, pp. 5145–5154, Mar. 2023, doi: 10.1007/s00521-021-06163-8.
- [19] A. Zhang, L. Su, Y. Zhang, Y. Fu, L. Wu, and S. Liang, "EEG data augmentation for emotion recognition with a multiple generator conditional Wasserstein GAN," *Complex & Intelligent Systems*, vol. 8, no. 4, p. 3059–3071, Aug. 2022, doi: 10.1007/s40747-021-00336-7.
- [20] A. R. Widiarti and R. Pulungan, "A method for solving scriptio continua in Javanese manuscript transliteration," *Heliyon*, vol. 6, no. 4, p. e03827, Apr. 2020, doi: 10.1016/j.heliyon.2020.e03827.
- [21] P. Basa Pati, G. V Chandana, C. Rithish Reddy, G. Balaji Subash, and J. SriHarsha, "Handwriting Quality Assessment using Structural Features and Support Vector Machines," in *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, IEEE, Apr. 2022, pp. 1–5, doi: 10.1109/I2CT54291.2022.9824125.
- [22] J. Gan, Y. Chen, B. Hu, J. Leng, W. Wang, and X. Gao, "Characters as graphs: Interpretable handwritten Chinese character recognition via Pyramid Graph Transformer," *Pattern Recognit*, vol. 137, p. 109317, May 2023, doi: 10.1016/j.patcog.2023.109317.
- [23] D. H. Ko, A. Ul Hassan, S. Majeed, and J. Choi, "SkelGAN: A Font Image Skeletonization Method," *Journal of Information Processing Systems*, vol. 17, no. 1, pp. 1–13, Feb. 2021, doi: 10.3745/JIPS.02.0152.
- [24] D. H. Ko, A. U. Hassan, J. Suk, and J. Choi, "SKFont: skeleton-driven Korean font generator with conditional deep

- adversarial networks,” *International Journal on Document Analysis and Recognition (IJДАР)*, vol. 24, no. 4, pp. 325–337, Dec. 2021, doi: 10.1007/s10032-021-00374-4.
- [25] S. Rana and M. Gatti, “Comparative Evaluation of Modified Wasserstein GAN-GP and State-of-the-Art GAN Models for Synthesizing Agricultural Weed Images in RGB and Infrared Domain,” *MethodsX*, vol. 14, p. 103309, Jun. 2025, doi: 10.1016/j.mex.2025.103309.
- [26] P. Milewski, “A Review of Image Thinning Algorithms,” in *Computer Information Systems and Industrial Management: 23rd International Conference, CISIM 2024*, 2024, pp. 3–16, doi: 10.1007/978-3-031-71115-2_1.
- [27] N. Mamatov, M. Jalelova, V. Fayziev, A. Samijonov, and B. Samijonov, “Hybrid quadratic diagonal algorithm for thinning contour lines,” *EPJ Web Conf*, vol. 321, p. 03005, Mar. 2025, doi: 10.1051/epjconf/202532103005.
- [28] H. A. Abdo, A. Abdu, R. Manza, and S. Bawiskar, “An approach to analysis of arabic text documents into text lines, words, and characters,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 2, p. 754, May 2022, doi: 10.11591/ijeecs.v26.i2.pp754-763.
- [29] L. O. Guertler, A. Ashfahani, and A. T. Luu, “How to train your draGAN: A task oriented solution to imbalanced classification,” *arXiv preprint*, Nov. 2022, doi: 10.48550/arXiv.2211.10065.
- [30] Q. Wan, W. Guo, and Y. Wang, “SGBGAN: minority class image generation for class-imbalanced datasets,” *Mach Vis Appl*, vol. 35, no. 2, p. 22, Mar. 2024, doi: 10.1007/s00138-023-01506-y.
- [31] S. Yuan, R. Liu, M. Chen, B. Chen, Z. Qiu, and X. He, “SE-GAN: Skeleton Enhanced Gan-Based Model for Brush Handwriting Font Generation,” in *IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, Jul. 2022, pp. 1–6, doi: 10.1109/ICME52920.2022.9859964.
- [32] Y. Xu, W. Luo, A. Hu, Z. Xie, X. Xie, and L. Tao, “TE-SAGAN: An Improved Generative Adversarial Network for Remote Sensing Super-Resolution Images,” *Remote Sens (Basel)*, vol. 14, no. 10, p. 2425, May 2022, doi: 10.3390/rs14102425.
- [33] G. Huang and A. H. Jafari, “Enhanced Balancing GAN: Minority-class Image Generation,” *ArXiv Preprint*, Oct. 2020, doi: 10.1007/s00521-021-06163-8.
- [34] I. Bakurov, M. Buzzelli, R. Schettini, M. Castelli, and L. Vanneschi, “Structural similarity index (SSIM) revisited: A data-driven approach,” *Expert Syst Appl*, vol. 189, p. 116087, Mar. 2022, doi: 10.1016/j.eswa.2021.116087.
- [35] M. 'Arif Faizin and D. Herumurti, “An Initial Study in Historical Javanese Script Recognition Data Synthetic using Neural Style Transfer,” in *2024 11th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, IEEE, Sep. 2024, pp. 787–792, doi: 10.1109/EECSI63442.2024.10776278.
- [36] M. 'Arif Faizin, N. Suciati, and C. Fatichah, “Evaluation of Synthetic Data Effectiveness using Generative Adversarial Networks (GAN) in Improving Javanese Script Recognition on Ancient Manuscript,” *JUTI: Jurnal Ilmiah Teknologi Informasi*, pp. 95–109, Feb. 2025, doi: 10.12962/j24068535.v23i1.a1256.
- [37] M. 'Arif Faizin, “Handwritten Javanese Characters Sewaka Detection (HJCS_DET),” Apr. 2025, *Zenodo*. doi: 10.5281/zenodo.15226702.
- [38] O. Fedoruk, K. Klimaszewski, A. Ogonowski, and R. Możdżonek, “Performance of GAN-based augmentation for deep learning COVID-19 image classification,” in *AIP Conference Proceedings*, Apr. 2023, doi: 10.1063/5.0203379.
- [39] R. Zhu, Y. Guo, and J.-H. Xue, “Adjusting the imbalance ratio by the dimensionality of imbalanced data,” *Pattern Recognit Lett*, vol. 133, pp. 217–223, May 2020, doi: 10.1016/j.patrec.2020.03.004.
- [40] M. Arabboev, S. Begmatov, M. Rikhsivoev, K. Nosirov, and S. Saydiakbarov, “comprehensive review of image super-resolution metrics: classical and AI-based approaches,” *Acta IMEKO*, vol. 13, no. 1, pp. 1–8, Mar. 2024, doi: 10.21014/actaimeko.v13i1.1679.