International Journal on Document Analysis and Recognition (IJDAR)
https://doi.org/10.1007/s10032-024-00466-x

ORIGINAL PAPER

Deformity removal from handwritten text documents using variable cycle GAN

Shivangi Nigam[1] · **Adarsh Prasad Behera**[1,2] · **Shekhar Verma**[1] · **P. Nagabhushan**[1]

Received: 25 March 2022 / Revised: 3 April 2024 / Accepted: 5 April 2024

Abstract
Text recognition systems typically work well for printed documents but struggle with handwritten documents due to different writing styles, background complexities, added noise of image acquisition methods, and deformed text images such as strike-offs and underlines. These deformities change the structural information, making it difficult to restore the deformed images while maintaining the structural information and preserving the semantic dependencies of the local pixels. Current adversarial networks are unable to preserve the structural and semantic dependencies as they focus on individual pixel-to-pixel variation and encourage non-meaningful aspects of the images. To address this, we propose a Variable Cycle Generative Adversarial Network ($VCGAN$) that considers the perceptual quality of the images. By using a variable Content Loss (Top-k Variable Loss (TV_k)), $VCGAN$ preserves the inter-dependence of spatially close pixels while removing the strike-off strokes. The similarity of the images is computed with TV_k considering the intensity variations that do not interfere with the semantic structures of the image. Our results show that $VCGAN$ can remove most deformities with an elevated $F1$ score of 97.40% and outperforms current state-of-the-art algorithms with a character error rate of 7.64% and word accuracy of 81.53% when tested on the handwritten text recognition system

Keywords Handwritten text · Strike-off · Semantics · Generative adversarial network · Image-to-image translation

1 Introduction

Handwritten Text Recognition (HTR) is an active area of research due to its wide range of applications, for example, digitization for digital libraries [25], text restoration [5, 6, 36], etc. It aims to transform the text in graphic forms such as images of handwritten notes, scene text, memos, whiteboards, medical records, and historical documents into its symbolic representation. The constraints of handwritten text dictate the complexity of the problem. Diverse free-flow writing styles of individuals, innumerable characters and their combinations, divergent backgrounds such as ruled pages, the image behind text (scene text), and various image acquisition practices are some challenges to recognition systems [7, 19, 29, 30].

The current state-of-the-art handwritten text recognition has been performed on images acquired by ideal and supervised means, for example, on the IAM dataset [21]. HTR systems have achieved high accuracy on such datasets; however, they need the notion of the intricacies of handwritten text documents [23]. Almost all Handwritten Text Recognition systems assume document texts are flawlessly written and captured. One such intricacy is the strike-off components in a handwritten text. An example of such occurrence is unrestricted handwritten text in students' examination notebooks, which may have these strike-off errors. With such a sample, HTR may produce irrelevant outputs. Due to the lack of such unrestricted data, it was challenging to develop HTR systems that could perform well on these irregularities in the data. Almost every handwritten document is expected

✉ Shivangi Nigam
rsi2018506@iiita.ac.in

Adarsh Prasad Behera
pwc2015004@iiita.ac.in

Shekhar Verma
sverma@iiita.ac.in

P. Nagabhushan
pnagabhushan@iiita.ac.in

[1] Department of Information Technology, Indian Institute of Information Technology Allahabad, Jhalwa, Prayagraj, Uttar Pradesh 211015, India

[2] Edge Networks Group, IMDEA Networks Institute, Av. Mar Mediterráneo, 22, 28918 Leganes, Madrid, Spain

Published online: 07 May 2024

to have such intricacies that a current OCR system cannot process. The pre-processing of these strike-off words can be favorable for applications like (a) OCR and digital transcription tasks: Strike-off words should be identified and removed for better transcription [2, 3] and (b) Document forensics: an aid to psychological clues for the forensic experts [8] and (c) Cognitive Analysis: helpful for analyzing behavioral and psychological patterns [10, 11].

There are different kinds of writing errors. The most common is the strike-off error. The strike-off is a mark-down indication to discard the content of the concerning text. It may be on a single character, word, multiple characters, multiple words, or multiple lines. The style of strike-off is a characteristic of individual writing fashion, which is profoundly indiscriminate. Some typical examples of challenges are shown in Fig. 1. Figure 1a shows the strike-off images with various types of strokes. Various image acquisition and environment-related flaws (ruled pages, background image, blurriness, skewness) are significant reasons for degrading document image quality. Most of the strike-off removal work considers the datasets as images in Fig. 1a. The strike-off elements may have various lengths and shapes. A more significant portion of text is mainly struck off with straight lines, such as single-line strike-offs, multiline strike-offs, or cross-strike-offs. In contrast, the words may have stroke types like a wave, zig-zag, cross, lined, scratch, etc. Also, different persons have different styles of strike-off.

Recent developments use generative modeling [15] for Document Image enhancement [34] and Handwritten Text Generation tasks [12, 14]. Generative models are unsupervised probabilistic models that attempt to learn patterns in a dataset and use these observations to generate new data. Generative adversarial networks (GAN) [15] use these generative models supervised by classifying generated samples as real and fake. GANs can generate novel and meaningful text while preserving the semantic and syntactic properties required by natural language processing and other document-related applications. Handwritten text Image generation using GANs has helped reduce the gap of datasets for training deep learning models [12, 23, 24]. Lately, Image-to-Image translation (I2I) has helped to translate from a source domain representation to a target domain representation. The research works in [17, 27, 28] have used I2I for translating from a strike-off image to a clean version of the corresponding image. The images produced by these [17, 27, 28] algorithms achieve acceptable perceived visual quality. However, precise matching with the ground truth has yet to be achieved. Simple element loss functions supervise the restoration task addressed by these works based on Mean Squared Error (MSE) or Structural Similarity Indices ($SSIM$) [27, 28]. These loss functions encourage perceptually meaningless aspects by accounting for the overall structure of the input data rather than considering the multi-modal distribution of data [13]. Hence, the restoration task must be improved to produce quality perceptually pleasing images.

In this work, we pose the strike-off strokes as changes in structural information and semantic content reflected in intensity variations in a clean handwritten text. A well-known way to consider intensity variations is the $L1$ norm. However, it cannot consider the importance of perceptual quality as it includes perceptually non-meaningful aspects. The perceived visual quality of a cleaned image is directly related to the removal of strike-off strokes and the preservation of interdependency of spatially close pixels. Consequently, we need a measure to evaluate the intensity variations while maintaining the structural information of the local pixels. We propose Top-k Variable loss TV_k as a new norm to measure the similarity of the image. TV_k focuses on the highest intensity variations to account for the significant structural differences between the strike-off image and its clean counterpart. The significant contributions of this work are:

(1) Strike-off removal using Unpaired Image-to-Image translation with a weakly supervised adversarial model: $VCGAN$.
(2) Content loss: Top-k Variable Loss (TV_k) for measuring the similarity of strike-off image and its clean counterpart.
(3) A CNN-LSTM-CTC model to perform recognition tasks on cleaned images generated by $VCGAN$.

The rest of the paper is structured as follows. Section 2 presents some of the recent developments and related works. In Sect. 3, some preliminaries or essential background knowledge are presented. The objectives and problem definition are

a) Strike-off words with clean background

b) Strike-off and alike words with complex background

Fig. 1 Strike-off words

stated in Sect. 4. Section 5 contains the proposed methodology. The datasets and implementation details are explained in Sect. 6. Section 7 discusses the experimental and comparison results, and finally, Sect. 8 concludes the paper.

2 Related works

The research in strike-off identification has primarily used hand-crafted manual methods such as SVM and HMM [3]. Then, image-in-painting methods have been utilized to restore the text. It had been a comparatively less explored area due to insufficient handwritten strike-off datasets. Recently, to resolve data scarcity, data augmentation techniques [26, 37] have been adequate to augment input data and produce new data in the input data space. Salient data augmentation techniques such as cropping, adding noise, resizing, flipping, rotating, and changing the color of an image were formerly used. However, there is a drawback: no new data are introduced, and more data are needed to improve the model's generalizability [26]. Lately, many data augmentation techniques have been explored to generate strike-off datasets. Recently, a Resnet-BiLSTM-CTC-based method was proposed for generating strike text in [23, 33].

The studies on strike-off text processing have been specific to scripts or styles of strike-off. Very early work [2] used K-Nearest Neighbor(K-NN) to identify and reject noise elements such as scribbles, crossed-outs, and isolated strokes. In [22], authors present a Markov random field(MRF) based MAP framework to determine joint energy distribution between labels and observation fields. In [4], a probabilistic contextual relationship model using a patch-based MRF was proposed to restore the printed documents' degradations, such as cuts, merges, blobs, and erosion. Another work in [20] proposed HMM-based wave and line stroke recognition, although the detection of strike-off is not considered in this work. Another work in [9] used a decision tree-based binary classifier to remove crossed-out handwritten text components. A US patent [35] claimed to recognize crossed-out English characters by a feature-based classifier. The work [1] has utilized morphological and graph-based features computed from deformed text images to identify and remove strike-offs. These works consider most strokes possible in a handwritten text, but the proposed solution requires a priori knowledge of the strike off being handled. In contrast to these manually crafted measures, deep learning approaches have been explored to address the problem of strike-off identification and recognition. The works in [17, 27, 28] have used I2I for translating from a strike-off image to a clean version of the corresponding image. They have created a synthetic dataset for this purpose under supervised conditions. These datasets do not acknowledge that free-form handwritten text has many more obstructions than posed in their work. Therefore, the methodologies do not consider the natural denotation of the problem.

3 Preliminaries/background

3.1 GANs for image-to-image translation

In I2I translation, the models seek to learn the mapping from input domain $\mathcal{X}$ to a target domain $\mathcal{Y}$, using paired or unpaired training samples $\mathcal{S} = \{(x_i, y_j)_{i,j}^{N,M}\}$ where $x \epsilon \mathcal{X}$, $y \epsilon \mathcal{Y}$ and N, M are the number of samples in domains $\mathcal{X}$ and $\mathcal{Y}$ respectively. In a GAN there is a $Generator\ Gen(\mathcal{X})$, which seeks to learn the mapping $\hat{h} : \mathcal{X} \rightarrow \mathcal{Y}$ during an adversarial training. The generator $Gen(\mathcal{X})$ competes with a discriminator $Discriminator$: $Disc(\mathcal{Y})$ with the intention of fooling the discriminator by producing indistinguishable fake samples. On the contrary, the discriminator learns to differentiate between real and fake samples. A min-max objective function achieves this:

$$\min_{Gen} \max_{Disc} \ f_{\vartheta,\varphi}(\breve{\theta}_g, \breve{\theta}_d) \tag{1}$$

where ϑ is a sample from real data $\vartheta \epsilon \mathcal{X}$ and $\varphi \epsilon \mathcal{N}[0, I_d]$ is a noise vector used by the generator to produce fake samples. The min-max objective learns weights $\breve{\theta}_g$ ($Gen(\mathcal{X})$ weights) such that they minimize the rate at which $Disc(\mathcal{Y})$ learns to differentiate real and fake. Simultaneously, it also learns weights $\breve{\theta}_d$ ($Disc(\mathcal{Y}$ weights) that maximize this rate (computed by f). Out of the various choices of the loss function f proposed in the literature, the cross entropy adversarial loss function by Goodfellow [15]:

$$\begin{aligned}&\mathcal{L}_{adv}^{\hat{h}}(Gen, Disc, \mathcal{X}, \mathcal{Y}) \\ &= \mathbb{E}_{\vartheta}[log(Disc(\vartheta))] + \mathbb{E}_{\varphi}[log(1 - (Disc(Gen(\varphi))))]\end{aligned} \tag{2}$$

I2I translation task with paired or unpaired samples aims to preserve the source content features and translate them into the target domain's style. The lack of paired ground truth images for most I2I applications, such as male-to-female translation, has inspired unpaired I2I translation. A translation task requires two generators $Gen(\mathcal{Y})$, $Gen(\mathcal{X})$ and discriminators $Disc(\mathcal{Y})$ $Disc(\mathcal{X})$ corresponding to mappings $\hat{h} : \mathcal{X} \rightarrow \mathcal{Y}$ and $\hat{f} : \mathcal{Y} \rightarrow \mathcal{X}$ to help the translation task cycle between the domains $\mathcal{X}, \mathcal{Y}$ to and fro. The Adversarial Loss is defined as:

$$\begin{aligned}\mathcal{L}_{adv_{\hat{h},\hat{f}}} =& \mathcal{L}_{adv}^{\hat{h}}(Gen_{\mathcal{Y}}, Disc_{protect}\mathcal{Y}, \mathcal{X}, \mathcal{Y}) \\ &+ \mathcal{L}_{adv}^{\hat{f}}(Gen_{\mathcal{X}}, Disc_{\mathcal{X}}, \mathcal{X}, \mathcal{Y})\end{aligned} \tag{3}$$

For a translation task, the adversarial loss computed between generated samples and the original samples is insufficient to ensure that the input x_i will be mapped to the specific target y_i. The cycle consistent loss [38] $\mathcal{L}_{cycle}$ regularizes the loss function by penalizing the inconsistencies of different domain translations. Without Cycle Consistent Loss, the generator produced images in the target domain that could not be translated between the two domains. Cycle consistency is imposed to ensure that the image-to-image translation is transitive, that is, $x \rightarrow \hat{h}(x) \rightarrow \hat{\hat{f}}(\hat{h}(x)) \sim x$. $\mathcal{L}_{cycle}$ is the L1 norm between the generated sample $fake(x)$ or $fake(y)$ and the corresponding original samples x or y.

$$\begin{aligned} \mathcal{L}_{cycle_x} &= \mathbb{E}_{x \sim p(\mathcal{X})} \| fake(x) - x \|_1 \\ \mathcal{L}_{cycle_y} &= \mathbb{E}_{y \sim p(\mathcal{Y})} \| fake(y) - y \|_1 \end{aligned} \tag{4}$$

The $\mathcal{L}_{cycle}$ gives the total of all the absolute errors between each pixel. Each pixel value is compared with other corresponding pixel values to produce a total representation of the translated pixel loss.

The total loss of I2I translation (CycleGAN [38])is the sum of adversarial losses for both mappings $\hat{h}$, $\hat{f}$ along with the cycle losses for the respective domains.

$$Total\ loss = \mathcal{L}_{adv_{\hat{h},\hat{f}}} + \mathcal{L}_{Cycle_{x,y}} \tag{5}$$

4 Problem definition

A distorted image can be considered a sum of an undistorted reference and an error component. The loss of perceptual quality is related to the visibility of the error component. MSE objectively quantifies the strength of the error and has a precise physical interpretation. However, distorted images with different visible or invisible errors may have the same MSE. Images are highly structured, and spatially proximate pixels carry information about the structure of the content and, therefore, exhibit strong dependencies. MSE is independent of the underlying signal structure and does not match perceived visual quality. Different measures of assessment of perceptual image quality weigh different aspects of errors to give different quantitative measures. The Content Loss of current adversarial networks cannot preserve the structure and semantic dependencies. Metrics like $L1$, MSE, and $SSIM$ consider each individual pixel-to-pixel variation. These metrics encourage the perceptually non-meaningful aspects rather than considering the multi-modal distribution of data of the images and thus limit the performance of the restoration.

5 VCGAN and handwritten text recovery

The objective of restoring strike-off images is to recover clean images while maintaining the structural information of the local pixels. This problem can be optimized to achieve perceptually pleasing target images, although the perceived visual quality of these target images should match the ground truth distribution. There is one common denominator in all metrics, such as MSE, $SSIM$, etc. These assessments measure the amount of structural information preserved in the distorted version of the reference image. Lost structural information leads to a lower quality score. These indices are applied locally as distortions may vary spatially, and statistical features of the image may be spatially non-stationary. For instance, the $SSIM$ Index defines the structural information in an image as those attributes representing the structure of objects in the scene, independent of the local average luminance and contrast. Objective structural similarity indices can capture the characteristics of subjective measures and yield a better assessment compared to MSE. However, these indices try to discount distortions that do not affect the local structures. Thus, we need indices based on intensity variations and structural information that can be used to measure the similarity between images that overcome the limitations of the intensity-based MSE index. This work focuses on preserving the interdependence of spatially close pixels while removing the strike-off strokes. To adhere to this goal, we design an objective function such that the target images are on the raw image manifold while maintaining the similarity to the ground truth distribution.

In this work, we propose Variable CycleGAN ($VCGAN$) that seeks to learn the mappings $\hat{h} : \mathcal{S} \rightarrow \mathcal{C}$ and $\hat{f} : \mathcal{C} \rightarrow \mathcal{S}$ from the input domain of the strike-off images $\mathcal{S}$ to a target domain of clean images $\mathcal{C}$ and vice versa. The generators $Gen(\mathcal{S})$ and $Gen(\mathcal{C})$ aim to translate from one domain to the other so that the images produced are indistinguishable for the discriminators $Disc(\mathcal{S})$ and $Disc(\mathcal{C})$. The idea is to utilize the unpaired Image-to-Image translation ability of adversarial models [38] to restore the text image.[1] The generated clean image (c) and the actual strike-off image (s) have already been defined before (Sect. 5, second paragraph) and are referenced in Section 5.1 . Consequently, we do not have the exact ground-truth pair of the strike-off text images we are training $VCGAN$. Persuading with this information, rather than having the target image precisely match the ground truth, we encourage the similarity of underlying semantic structural distributions. We achieve this by a new similarity norm Top-k Variable loss $\mathcal{TV}_k$.

[1] The model is weakly supervised, as it uses unpaired training data samples $\{(s_i, c_j)_{i,j}^{N,M}\}$ where $s \in \mathcal{S}$ and $c \in \mathcal{C}$

5.1 VCGAN objective function

The objective function of $VCGAN$ comprises two loss functions: Adversarial loss, similar to CycleGAN [38] and Content loss defined as Top-k Variable loss $\mathcal{TV}_k$. $\mathcal{TV}_k$ empowers the $VCGAN$ with the flexibility (with variable k) of the expanse of semantic dependencies to be preserved to improve perceived image quality in various applications.[2] The cross-entropy adversarial loss for $VCGAN$ can be defined by the following:

$$\begin{aligned}\mathcal{L}_{adv_{\hat{h},\hat{f}}} =& \mathcal{L}_{adv}^{\hat{h}}(Gen_{\mathcal{C}}, Disc_{\mathcal{C}}, \mathcal{S}, \mathcal{C}) \\ &+ \mathcal{L}_{adv}^{\hat{f}}(Gen_{\mathcal{S}}, Disc_{\mathcal{S}}, \mathcal{S}, \mathcal{C})\end{aligned} \tag{6}$$

The Content Loss of a CycleGAN is an image similarity measure that computes a distance metric that evaluates the differences between the corresponding pixels of two images. The Sum of Absolute Errors (SAE) or $L1$ norm computes the absolute difference between the corresponding pixels and then takes the mean over all pixels. Average loss is the most widely used metric to fairly approximate abrupt anomalies. The pixel-wise loss between two images (a, b) of dimension $(m * n * 3)$ with $L1$ is given by:

$$\|L\|_1 = \frac{1}{p}\sum_{i=0}^{p} l(a_i, b_i) \tag{7}$$

where $p = (m * n * 3)$ and $l(a_i, b_i)$ is the L1 norm between every pixel of a and b.

Although an important issue in designing an objective function for an $I2I$ translation is to deal with the high level of the interdependent semantic content of an image, the perceived visual quality of an image is related to the visibility of any deformity (strike-off) in the image and the semantic structural content of the image. Such content is more salient to maintain the perceptual quality of an image. Thus, the judgment of perceptual similarity is influenced by removing the deformities and preserving the salient semantic structures. To achieve this objective, we draw inspiration from the work [13, 31] to design an objective function sensitive to an image's structural information. We propose Top-k Variable loss $\mathcal{TV}_k$ to compute the similarity of images by the intensity variation between strike-off image s and its clean counterpart c. This loss detects the structural information change, accounting only for those variations that should not interfere with the interdependence of spatially close pixels. The $\mathcal{TV}_k$ Variable Loss is an aggregate loss, the mean over the k top individual losses over data. $\mathcal{TV}_k$ norm is a natural generalization of Average and Maximum Loss. It is shown in [13] due to the additional flexibility provided by the different choices of k, $\mathcal{TV}_k$ Variable Loss can adapt to different data distributions. Thus, it can preserve the interdependence of spatially close pixels. Moreover, it is a convex function of overall individual losses, and the computation is only marginally more than individual losses like MSE and $SSIM$. These properties make it suitable for strike-off removal.

We propose $\mathcal{TV}_k$ to preserve the semantics of images at the pixel level rather than the sample level by measuring the variations in intensity per pixel. $\mathcal{TV}_k$ addresses the shortcomings of the $L1$ norm by measuring the similarity of only top $k\%$ intensity variations. By doing this, we ignore the variations that interfere with the images' semantic structures. The per-pixel intensity variation is given by a set $\mathcal{V}$, which is computed by $l(s_i, c_i)$ where $0 \leq I \leq p$. As shown in Eq. 9 $\mathcal{TV}_k$ is the average of k, $(0 \leq k \leq 100)$ maximum per pixel intensity variations ($\mathcal{T} : \mathcal{T} \subset \mathcal{V}$) between images s and c (eg. $k =$ 50 means $\mathcal{T}$ is top 50% values in $\mathcal{V}$) and $card(\mathcal{T}) = (\frac{k}{100} \cdot p)$ is the cardinality of $\mathcal{T}$.

$$\mathcal{V} = \{v \epsilon\, l(s_i, c_i) : 0 \leq i \leq p\} \tag{8}$$

$$\|\mathcal{TV}_k\| = \frac{1}{card(\mathcal{T})}\Big[\arg\max_{\mathcal{T}\subset\mathcal{V}} \sum_{t\epsilon\mathcal{T}} t\Big] \tag{9}$$

$\mathcal{TV}_k$ can be scaled to match the standard $L1$ norm when $k = 100$. The coefficient k is a meta-parameter that provides flexibility to adapt to diverse data distributions.

The objective function of $VCGAN$ is:

$$\begin{aligned}\mathcal{L}(Gen, Disc, \mathcal{S}, \mathcal{C}) =& \mathcal{L}_{adv_{\hat{h}}} + \mathcal{L}_{adv_{\hat{f}}} \\ &+\lambda(\mathcal{TV}_{k_s} + \mathcal{TV}_{k_c})\end{aligned} \tag{10}$$

Where λ is a coefficient to control the relative importance of the two losses in the objective function. The aim is to solve the objective:

$$\min_{Gen}^{\theta_g}\max_{Disc}^{\theta_d}[\mathcal{L}(Gen, Disc, \mathcal{S}, \mathcal{C})] \tag{11}$$

The generators $Gen(\mathcal{C})$ and $Gen(\mathcal{S})$ learns the mapping $\hat{h}$ and $\hat{f}$ during a min-max game with the discriminators $Disc(\mathcal{C})$ $Disc(\mathcal{S})$ respectively and learn the parameters θ_g, θ_d. The model tries to restore strike-off image $\hat{x}$ to cleaned image $\hat{y}$ with minimum restoration error while encouraging the target $\hat{y}$ to be perceptually similar to the ground truth distribution of x.

[2] $VCGAN$ computes Adversarial Loss between the generated clean image (c) and the actual strike-off image (s) as discussed above in Sect. 3

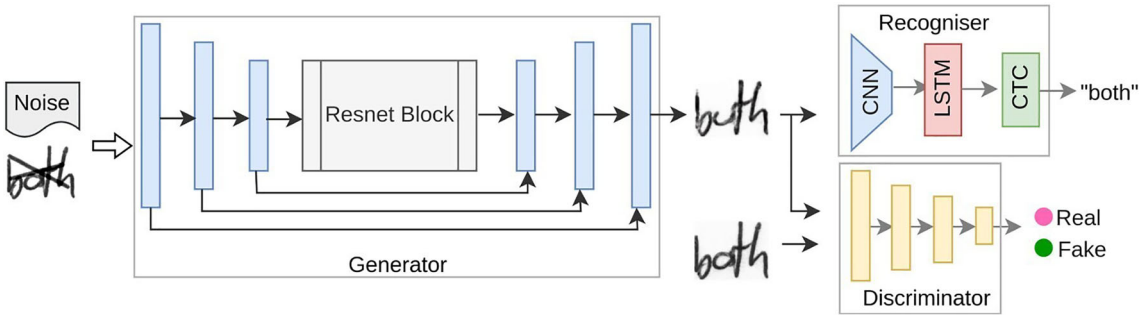

Fig. 2 A network architecture of VCGAN with Generator, Discriminator and Recognizer

6 Implementation details

6.1 Adversarial network

The generative networks of our model (Fig. 2) are adapted from [18]. The generator network contains stride-2 convolutions for down-sampling and several residual blocks, which are used to capture relevant information and flow that information from the initial layers to the last ones and at last convolutions with stride $\frac{1}{2}$ for up-sampling. We use the instance normalization technique in [18]. The discriminator network uses 70×70 Patch GAN to identify real and fake images. Unlike a normal GAN discriminator, a patch GAN outputs a $N \times N$ output array O in which O_{ij} indicates whether the respective patch belongs to real or fake.

6.2 Recognition network

To evaluate the performance of $VCGAN$, we use a subsequent recognition network. For recognition tasks, a CNN-LSTM-CTC network is implemented. The CNN layers perform feature extraction, which LSTM layers use to identify the temporal patterns in the feature set. The network comprises 5 CNN layers, 2 LSTM layers, and, lastly, connectionist temporal classification (CTC) is used to predict the final output.

6.3 Dataset

In this work, we have used various text data sources to enhance the model's ability. The strike-off words are generated by superimposing actual strokes over the clean word images using the strike-off generation algorithm proposed in the work [17]. The datasets used in this work are described as follows:

- *Augmented IAM dataset* This is the primary dataset on which the model is trained on [16]. We have synthetically augmented the IAM dataset [21] by producing strike-off images for all types of stroke, such as single line, double line, cross, wave, zig-zag, and scratch.
- *Real-world handwritten text* We have manually collected handwritten text images from student notebooks to increase the diversity of training samples [32]. These samples are actual strike-off words and are not biased toward supervised conditions, unlike the IAM dataset. The data are collected as documents, which are segmented into words to train the proposed model.

Table 1 Performance metrics for strike-off removal for $k = 10, 50$

Stroke type	Pr	Re	$F1$s
(a) $k = 10$			
Cross	97.856	96.963	97.410
Wave	91.589	90.591	91.090
Scratch	89.576	88.580	89.078
Single line	96.401	96.817	96.609
Double line	96.747	95.849	96.298
Diagonal	97.833	96.921	97.377
Zig-zag	90.541	89.136	89.839
(b) $k = 50$			
Cross	97.711	95.952	96.832
Wave	91.565	90.562	91.064
Scratch	89.565	88.562	89.064
Single line	96.972	96.805	96.889
Double line	96.938	95.927	96.433
Diagonal	96.328	96.069	96.199
Zig-zag	89.526	88.008	88.767

6.4 Training details

We have applied some techniques to optimize the training of the $VCGAN$ model. First, the generator aims to minimize the probability of images being predicted as fake. In other words, the generator seeks to maximize the probability of images being predicted as real. Thus non-saturating loss function of the generator is to maximize $log(Disc(Gen(s)))$

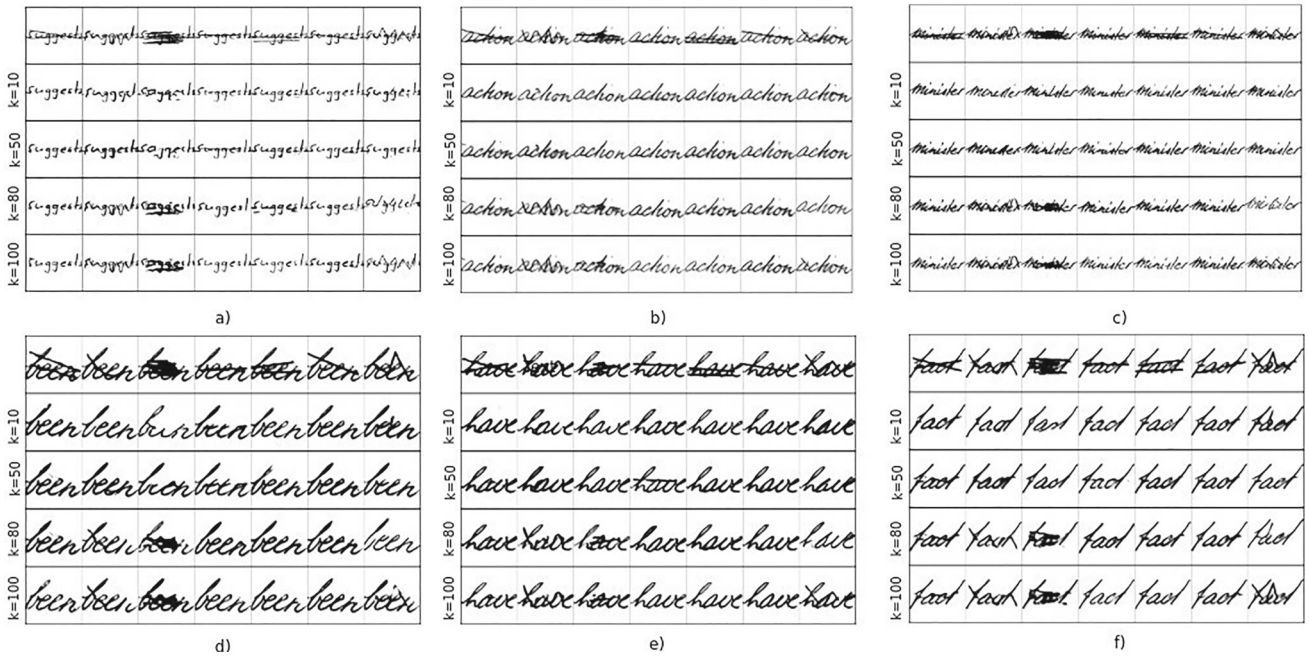

Fig. 3 Strike-off removal with $VCGAN$ for various values of k

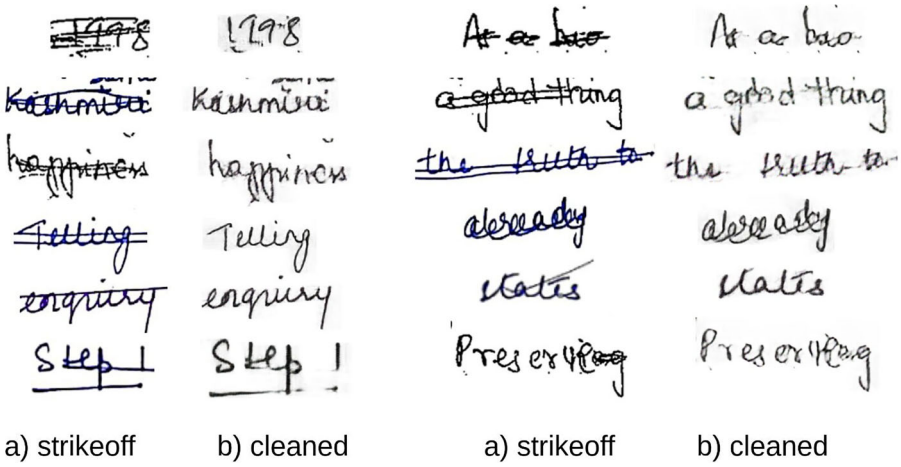

Fig. 4 Strike-off samples taken form student notebooks and corresponding generated cleaned text

$$\mathcal{L}_{Gen}(\mathcal{S}, \mathcal{C}) = E_{\varphi \sim P_{\varphi}}[log(Disc(Gen(\varphi)))] \quad (12)$$

Secondly, we have used two approaches to reduce the model oscillation while training: a)Instead of learning with a fixed learning rate, we use a learning rate scheduler to reduce the learning curve. This is a way to reduce the primacy effect of early training examples. Without it, you may need to run a few extra epochs to achieve the desired convergence, as the model untrains those early superstitions. b) Provide the discriminator with images that have been produced recently, instead of just those from the last iteration. We maintain a buffer to hold some images of recent epochs. So instead of providing images produced in the last iteration, we provide a mixture of images from recent iterations.

7 Results

This work considers seven types of deformities (strike-offs) in the handwritten text: cross, wave, scratch, single line, double line, diagonal, and zig-zag. The model is tested on the proportional combination of all types of deformities. The proposed model is also tested on underlined text images. It can differentiate between a strike-off and an underline. We have observed high performance of deformities like a cross, single line, double line, and diagonal, while scratch, wave, and zig-zag have seen average performances. Figure 3 shows some generated samples of our model. We have also tested our approach on a handwritten dataset collected from student notebooks. It can be observed in Fig. 4 that our approach

achieves impressive results on student notebooks as well. The approach is tested for two objectives:

- Performance of strike-off detection and removal
- Performance of Handwritten Text Recognition

7.1 Performance of strike-off removal

An authentic image quality assessment compares the target images with the ground truth images. Due to the absence of ground truth data in all the test data, we access the image quality using various reference metrics. Image similarity metrics such as MSE (Mean Square Error), $PSNR$ (Peak Signal to Noise Ratio), $SSIM$ (Structured Similarity Index Method) evaluate images for their structural similarity. Pixel-based metrics like Precision, Recall, and $F1$ score measure the quality of restoration (strike-off removed images) by directly comparing pixels of corresponding images. Image restoration metrics such as deformity detection rate (DDR), restoration accuracy (RA), and F measure (FM) are used to analyze the performance of the strike-off image removal. Overall, these assessment techniques measure the deviation of quality of generated clean images for the ideal/ground truth clean images with different perspectives.

In a generated clean image, we define: *True Positives (TP)*: the foreground area of the image which is correctly identified as strike-off and is removed in the generated clean image *False Positives (FP)*: the foreground area of the image, which is incorrectly labeled as strike-off and hence is removed in the generated clean image. *False Negatives (FN)*: the missing area of the image that could not be labeled as strike-off and hence is not removed from the generated clean image. *Intersection measure (IM)*: Intersecting pixels of generated image and true image *True pixels (Tr)*: Foreground pixels in true image *Pred pixels (Pr)*: Foreground pixels in the generated image

Our proposed model, $VCGAN$, demonstrates impressive results in removing the strikes. We present the experimentation results in the form of the following measures, which are shown in Fig. 3 and Table 1 representing results for coefficient $k = (10, 50)$.[3]

- Pixel-based measures
  - Precision (Pr)= $TP/TP + FP$
  - Recall (Re): $TP/TP + FP$
  - $F1$ score ($F1$s): $2 * Pr * Re/(Pr + Re)$

The $F1$ score represents a balanced assessment of precision (correctly identified strike-off regions) and recall (correctly restored text). The average $F1\ score$ achieved is 93.95% with a standard deviation of (±3.766). Average Precision and average Recall achieved are 94.363%(±3.635), 93.55%(±3.766) respectively.[4] The highest precision (96.85%) and recall (96.93%) values are obtained for cross-type with k size of 10. Figure 5 illustrates that the results for k size 10, 50 are close to the best precision and recall. This suggests that $10-50$ is an optimum range of k. A high $F1$ score of 97.41% for cross-type strike-offs indicates the exceptional performance of $VCGAN$ in detection and restoration. Among all strike-offs, scratch stroke has been the most challenging, with the lowest $F1$ score of 89.55%, highlighting an area of further research to improve $VCGAN$'s performance (Tables 2 and 3).

- Image similarity measures
  - Mean squared error (MSE) between generated $g(x, y)$ and true $t(x, y)$ image as defined in equation. $MSE = \frac{1}{MN}\sum_{n=0}^{N}\sum_{m=0}^{M}[t(n, m) - g(n, m)]^2$
  - Peak Signal to Noise Ratio ($PSNR$) $PSNR = 10 \cdot \log_{10}\left(peak_{value}^2\right)(MSE)$
  - The structural similarity index measure ($SSIM$) computes luminance $\mathcal{L}$, contrast $\mathcal{C}$ and structure $\mathcal{S}$ of the reference images x, y. $SSIM(x, y) = [\mathcal{L}(x, y)^{\alpha} \cdot \mathcal{C}(x, y)^{\beta} \cdot \mathcal{S}(x, y)^{\gamma}]$

Image similarity metrics MSE and $PSNR$ are absolute errors that can have the same values for different deformations in an image. Therefore, these cannot discriminate between the structural content of the images. $SSIM$ is better at these scenarios as it captures perception and saliency-based variations by incorporating structural information. The results demonstrate that high $SSIM$ scores (0.78) are obtained for $k = 10$ and $k = 50$. This indicates that the results of $SSIM$ with these k sizes align with the $F1$ scores obtained above. With image restoration tasks like text images, these $SSIM$ scores are acceptable as the loss of information in this case is inevitable, but the core structure of images is preserved. The result Tables 4, 5, 6 and 7 of the image similarity metrics for $k = 10, 50, 80, 100$ are in Appendix section.

- Image Restoration measures [1]
  - Deformity detection Rate: $DDR = IM/Tr$
  - Reconstruction Accuracy: $RA = IM/Pr$
  - $FM = (2 * DDR * DDR)/(DDR + RA)$

We analyze DDR, RA, and FM to analyze the model beyond pixel-based and similarity-based methods. DDR

[3] Additional results for $k = (80, 100)$ are reported in the Appendix.

[4] The standard deviation values are confidence scores for the variation across various strike-off types considered in this work

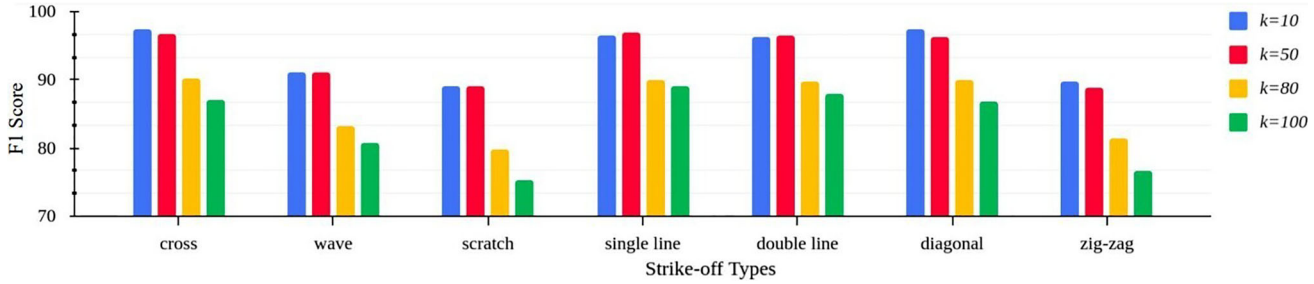

Fig. 5 $F1$ scores of $VCGAN$ with different k values across various types of strike-off

depicts the model's performance in detecting strike-off regions, while the restoration accuracy measures how well the text is recovered from the strike-off image. The F measure conveys the overall accuracy of detection and restoration. Tables 4b and 5b show good recognition accuracy and F measure, while the results for $k = 80, 100$ in Tables 6b and 7b have mixed performance of these metrics. Despite the mixed results for larger k sizes, $VCGAN$ achieves a peak DDR of 97.85%, indicating impressive detection capabilities. The model obtained top RA of 96.96%, demonstrating excellent restoration of text portions. The corresponding $FMeasure$ is 97.407%, signifying overall performance. The overall $F1score$ is 93.50%(±3.01), with a detection rate and restoration accuracy of 93.863%(±2.824), 93.151% (±3.185) which shows consistent overall performance across all strike-offs.

Our experiments demonstrate that the proposed model performs consistently over various strike-off types for specific k sizes. Figure 3 showcases results of various k values across different strike-offs, whereas Fig. 4 presents examples of actual notebooks, which include multiple-word strike-off, multiple-line strike-off, partial strike-off, and underlined samples. Figure 6 compares various reference works [1, 17, 28] with our proposed model. The work in [28] proposes TexRGAN, which utilizes a CycleGAN-based approach and overlooks three strike-off categories, scratch, wave, and zig-zag, which are prevalent in most handwritten documents. Another work in [1] has not considered scratch strike-offs, although the work tends to produce comparable results on other strike-offs considered in this work. The work in [17] has a similar approach as [28], but their model does not account for the semantics of the content in their objective function. Consequently, their translation process is not seamless. A comprehensive comparison of $F1$ scores of $VCGAN$ with different k values over various strike-offs is shown in Fig. 5. The elaborate results of various performance measures are presented in Tables 4 5, 6 and 7.

7.2 Performance of handwritten text recognition (HTR)

To evaluate the recognition capability of images generated by the proposed model, we used a CNN-LSTM-CTC-based handwritten text recognition module. This module is trained on the IAM words dataset. The $VCGAN$ generated images are passed to the HTR module, which produces transcribed text. This text is evaluated with Character Error Rate (CER) and Word Accuracy(WA). These error rates provide granularity to measuring the difference between transcribed and actual text labels. We have manually labeled the generated images into their respective transcribed text labels to perform recognition tasks.

- Character Error Rate (CER): This represents the percentage of characters incorrectly predicted during OCR. We require lower CERs' for better recognition performance.

$$CER = \frac{(S + D + I)}{Characters\ in\ ground\ truth}$$

where, S is number of substitutions, D is number of Deletions and I is number of Insertions.

- Word Accuracy (WA):

$$WA = \frac{Correctly\ transcribed\ words}{Total\ words} \tag{13}$$

WA represents the ratio of correctly transcribed words. Correctly transcribed words are labeled CORRECT. The higher the accuracy, the better the recognition performance.

Table 2 shows the CER and WA values for various strike-off types for different k sizes. The results show that the diagonal stroke exhibited the highest accuracy of 81.53% for $k = 10$, followed by a double line with 75.78% for $k = 50$. The model struggles with scratch stroke with lowest WA of 31.47% and highest CER of 32.65%. The best CER of 4.76% is observed for diagonal stroke with $k = 10$.

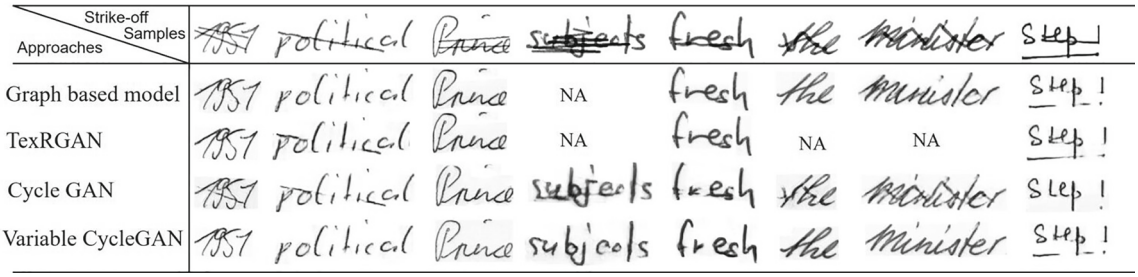

Fig. 6 Comparison of various state-of-the-art methods on various types of strike-off

This highlights that higher values of k have observed higher CER and lower accuracies. Once again, the values of metrics CER and WA justify the results of other performance measures and prove that optimal k size is crucial for balancing model complexity and performance for different strike-offs.

7.3 Discussion

The key findings of our work can be summarized as follows:

Effectiveness of TV_k *Loss* $VCGAN$ leverages Top-k Variable loss TV_k norm that focuses on capturing the deformities while maintaining the perceived visual quality. TV_k incorporates coefficient k that is a meta-parameter that provides flexibility to adapt to different data distributions. In this work, we experimented with $k = 10, 50, 80, 100$. We have observed that range of k between 10 and 50 gives the best results in removing the strike (Fig. 5). This signifies that these k values are optimal for balancing high intensity variations. Larger values of k tend to include meaningless intensity variations when measuring the similarity of images. Conversely, lower values of k (below 10) affect the clarity of underlying text by removing minor intensity variations.

TV_k allows the model to be flexible while effectively balancing the strike-off removal with maintaining the image quality. The experiments indicate that the best results are obtained by setting k in range (10, 50). Interestingly, $k = 100$ makes the model similar to CycleGAN, thus demonstrating the adaptability of $VCGAN$ for a variety of distributions.

$VCGAN$'s *Performance* $VCGAN$ exhibits language-agnostic and script-/grammar-independent properties. It can be extended to different writing systems (such as Bengali, Devanagari, etc.). We experimented the model with seven strike-offs settings: cross, wave, scratch, diagonal, single line, double line, and zig-zag. The results show that $VCGAN$ removes most of the strike-offs, while, stroke types wave, scratch, and zig-zag were proved to be challenging than others. The model obtained best performance on the cross, diagonal, single line, and double line stroke types with a $F1$ score of 97.40% which indicates an exceptional balance of efficient strike-off removal (precision) and accurate text reconstruction (recall). This demonstrates that $VCGAN$ effectively removes strike-off regions while preserving the actual text. This is justified with impressive detection capabilities (DDR: 97.85%); less character recognition errors (CER: 7.64%) and high word recognition accuracies (WA: 81.53%) (Table 2). Figure 3 shows that variation in k significantly changes the perceptual quality of the image. The results of the unconstrained data collection can be seen in Fig. 4. The intricate strike-off patterns scratch, wave, and zig-zag were the most challenging ones (Fig. 7). However, our model demonstrates significant removal ability even in these challenging cases.

Fig. 7 Some examples of challenging samples of strike-off text with corresponding generated and ground truth text

8 Conclusion

This work introduced $VCGAN$ model with a Top-k Variable loss, a novel approach to remove strike-off from handwritten images. TV_k effectively captures essential distributions of images while ignoring the intensity variations that tend to interfere with the semantic structures of the image. Our experiments suggested that smaller values of k (within the range of 10-50) leads to optimal performance achieving impressive scores for different performance metrics like $F1$ score (97.40%), DDR (97.85%), CER (7.64%) and WA (81.53%) on the generated images. This showed $VCGAN$'s impressive ability in strike-off removal, even for the most complex patterns such as wave, zig-zag, and scratch. Additionally, the comparison results signified that our approach has outperformed the current state-of-the-art methodologies.

In conclusion, $VCGAN$ is empowered with the flexibility to expand semantic dependencies with TV_k preserving the perceived image quality. The model has achieved an

exceptional balance between strike-off removal and text restoration.

Appendix

Table 2 Handwritten text recognition (Character error rate (CER) and Word Accuracy(CER)) of generated images for strike-off removal on various k values

Strike off type	Measure	$k = 10$	$k = 50$	$k = 80$	$k = 100$
	$CER\%$	1.75			
Clean	$WA\%$	89.52			
	$CER\%$	**7.64**	8.45	30.87	34.01
Cross	$WA\%$	**72.45**	71.38	20.56	19.09
	$CER\%$	**20.31**	23.54	22.74	26.95
Wave	$WA\%$	**54.35**	51.31	17.16	15.67
	$CER\%$	**4.76**	6.12	18.91	19.34
Diagonal	$WA\%$	**81.53**	80.15	18.61	17.38
	$CER\%$	7.96	**7.81**	15.47	17.29
Double Line	$WA\%$	70.12	**75.78**	28.95	25.23
	$CER\%$	10.78	**10.57**	16.85	18.74
Single line	$WA\%$	70.96	**71.19**	27.39	19.82
	$CER\%$	**32.65**	39.10	55.67	59.48
Scratch	$WA\%$	**31.47**	27.91	14.14	15.00
	$CER\%$	**15.41**	19.89	58.70	60.88
Zig-zag	$WA\%$	52.36	**53.23**	18.03	15.11

Bold indicate the best performance of the proposed algorithm for a particular case

Table 3 Comparison of performance measures of various state-of-the-art methods

Approaches	$F1$ score %	$CER\%$	$WA\%$
Graph based model[1]	89.44	–	–
TexRGAN[28]	96.76	12.74	65.28
CycleGAN[17]	83.40	17.29	25.23
VCGAN(ours)	**97.40**	**7.64**	**81.53**

Bold indicate the best performance of the proposed algorithm for a particular case

Table 4 Performance Metrics for strike-off removal for $k = 10$

(a) Image Similarity Metrics

Stroke type	MSE	$PSNR$	$SSIM$
Cross	0.067	11.957	0.706
Wave	0.070	11.712	0.780
Scratch	0.064	12.149	0.799
Single line	0.063	12.178	0.704
Double line	0.063	12.178	0.707
Diagonal	0.063	12.178	0.710
Zig-zag	0.065	12.064	0.703

Table 4 continued

(b) Deformity detection Metrics

DDR	RA	FM
97.856	96.963	97.407
91.589	90.591	91.087
89.576	88.580	89.075
95.401	96.817	96.104
96.747	95.849	96.296
97.833	96.921	97.377
90.541	89.136	89.833

Table 5 Performance Metrics for strike-off removal for $k = 50$

(a) Image Similarity Metrics

Stroke type	MSE	$PSNR$	$SSIM$
Cross	0.069	12.835	0.684
Wave	0.076	12.991	0.676
Scratch	0.074	11.991	0.670
Single line	0.069	12.172	0.702
Double line	0.068	11.902	0.711
Diagonal	0.072	12.961	0.600
Zig-zag	0.079	12.799	0.687

(b) Deformity Detection Metrics

DDR	RA	FM
97.711	95.952	96.824
91.565	90.562	91.061
89.565	88.562	89.064
95.372	96.805	96.389
96.438	95.827	96.132
96.328	96.069	96.199
90.526	89.008	89.761

Table 6 Performance Metrics for strike-off removal for $k = 80$

(a) Image Similarity Metrics

Stroke type	MSE	$PSNR$	$SSIM$
Cross	0.081	14.913	0.582
Wave	0.082	14.865	0.575
Scratch	0.080	14.664	0.584
Single line	0.072	13.580	0.584
Double line	0.078	13.834	0.594
Diagonal	0.075	13.080	0.599
Zig-zag	0.080	14.889	0.634

(b) Deformity Detection Metrics

DDR	RA	FM
97.856	96.963	97.410
91.589	90.591	91.090
89.576	88.580	89.078
95.401	96.817	96.109

Table 6 continued

(b) Deformity Detection Metrics		
DDR	RA	FM
96.747	95.849	96.298
97.833	96.921	97.377
90.541	89.136	89.839

(c) Pixel based Metrics		
Pr	Re	$F1$s
91.555	88.791	90.173
83.766	82.834	83.300
80.653	79.157	79.905
91.621	88.492	90.057
90.399	89.137	89.768
90.741	89.069	89.905
81.999	80.959	81.479

Table 7 Performance Metrics for strike-off removal for $k = 100$

(a) Image Similarity Metrics			
Stroke type	MSE	$PSNR$	$SSIM$
Cross	0.086	12.023	0.574
Wave	0.086	11.963	0.565
Scratch	0.084	11.682	0.532
Single line	0.073	12.215	0.517
Double line	0.079	12.216	0.521
Diagonal	0.087	11.907	0.556
Zig zag	0.085	12.037	0.537

(b) Deformity detection Metrics		
DDR	RA	FM
97.711	95.952	96.832
91.565	90.562	91.064
89.565	88.562	89.064
95.972	96.805	96.389
96.438	95.827	96.133
96.328	96.069	96.199
90.526	89.008	89.767

(c) Pixel based Metrics		
Pr	Re	F!s
88.456	85.758	87.107
81.732	79.821	80.777
76.642	74.145	75.394
89.610	88.449	89.030
88.313	87.619	87.966
88.702	85.043	86.873
77.579	75.847	76.713

Author Contributions All authors have equally contributed in the manuscript.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

1. Adak, C., Chaudhuri, B.B.: An approach of strike-through text identification from handwritten documents. In: 2014 14th International Conference on Frontiers in Handwriting Recognition, pp 643–648. IEEE (2014)
2. Arlandis, J., Pérez-Cortes, J. C., Cano, J.: Rejection strategies and confidence measures for a k-nn classifier in an OCR task. In: Object Recognition Supported by User Interaction for Service Robots, pp 576–579. IEEE (2002)
3. Banerjee. J., Namboodiri, A.M., Jawahar, C.: Contextual restoration of severely degraded document images. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp 517–524. IEEE (2009a)
4. Banerjee, J., Namboodiri, A.M., Jawahar, C.: Contextual restoration of severely degraded document images. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp 517–524. IEEE (2009b)
5. Bannigidad, P., Gudada, C.: Restoration of degraded historical kannada handwritten document images using image enhancement techniques. In: International Conference on Soft Computing and Pattern Recognition, pp 498–508. Springer (2016)
6. Bannigidad, P., Gudada, C.: Restoration of degraded kannada handwritten paper inscriptions (hastaprati) using image enhancement techniques. In: 2017 International Conference on Computer Communication and Informatics (ICCCI), pp 1–6. IEEE (2017)
7. Bathla, A,K., Gupta, S.K., Jindal, M.K.: Challenges in recognition of devanagari scripts due to segmentation of handwritten text. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp 2711–2715. IEEE (2016)
8. Brink, A., Schomaker, L., Bulacu, M.: Towards explainable writer verification and identification using vantage writers. In: Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), pp 824–828. IEEE (2007)
9. Brink, A., van der Klauw, H., Schomaker, L.: Automatic removal of crossed-out handwritten text and the effect on writer verification and identification. In: Document Recognition and Retrieval XV, International Society for Optics and Photonics, p 68150A (2008)
10. Caligiuri, M.P., Mohammed, L.A.: The Neuroscience of Handwriting: Applications for Forensic Document Examination. CRC Press, Boca Raton (2012)
11. Chaudhuri, B.B., Adak, C.: An approach for detecting and cleaning of struck-out handwritten text. Pattern Recognit. **61**, 282–294 (2017)
12. Eltay, M., Zidouri, A., Ahmad, I., et al.: Generative adversarial network based adaptive data augmentation for handwritten arabic text recognition. PeerJ Comput. Sci. **8**, e861 (2022)
13. Fan, Y., Lyu, S., Ying, Y., et al.: Learning with average top-k loss. In: Advances in neural information processing systems 30 (2017)
14. Fogel, S., Averbuch-Elor, H., Cohen, S., et al.: Scrabblegan: Semi-supervised varying length handwritten text generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 4324–4333 (2020)
15. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: Advances in neural information processing systems 27 (2014)
16. Heil, R., Vats, E., Hast, A.: Iam strikethrough database. (2021). https://doi.org/10.5281/zenodo.4767095

17. Heil, R., Vats, E., Hast, A.: Paired image to image translation for strikethrough removal from handwritten words. arXiv preprint arXiv:2201.09633 (2022)
18. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, pp 694–711. Springer (2016)
19. Khobragade, R.N., Koli, N.A., Lanjewar, V.T.: Challenges in recognition of online and off-line compound handwritten characters: a review. In: Smart Trends in Computing and Communications, pp 375–383 (2020)
20. Liao, M., Shi, B., Bai, X., et al.: Textboxes: A fast text detector with a single deep neural network. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
21. Marti, U.V., Bunke, H.: The iam-database: an english sentence database for offline handwriting recognition. Int. J. Doc. Anal. Recognit. **5**(1), 39–46 (2002)
22. Nicolas, S., Paquet, T., Heutte, L.: Markov random field models to extract the layout of complex handwritten documents. In: Tenth International Workshop on Frontiers in Handwriting Recognition, Suvisoft (2006)
23. Nisa, H., Thom, J.A., Ciesielski, V., et al.: A deep learning approach to handwritten text recognition in the presence of struck-out text. In: 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ), pp 1–6. IEEE (2019)
24. Nisa, H., Ciesielski, V., Thom, J., et al.: Annotation of struck-out text in handwritten documents. In: Proceedings of the 25th Australasian Document Computing Symposium, pp 1–7 (2021)
25. Pande, S.D., Jadhav, P.P., Joshi, R., et al.: Digitization of handwritten devanagari text using CNN transfer learning-a better customer service support. Neurosci. Inform. **2**(3), 100016 (2022)
26. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621 (2017)
27. Poddar, A., Chakraborty, A., Mukhopadhyay, J., et al.: Detection and localisation of struck-out-strokes in handwritten manuscripts. In: International Conference on Document Analysis and Recognition, pp 98–112. Springer (2021a)
28. Poddar, A., Chakraborty, A., Mukhopadhyay, J., et al.: Texrgan: a deep adversarial framework for text restoration from deformed handwritten documents. In: Proceedings of the Twelfth Indian Conference on Computer Vision, Graphics and Image Processing, pp 1–9 (2021b)
29. Rajiv, K.S., Amardeep, S.D.: Challenges in segmentation of text in handwritten gurmukhi script. In: International Conference on Business Administration and Information Processing, pp 388–392. Springer (2010)
30. Rusu, A.I., Govindaraju, V.: On the challenges that handwritten text images pose to computers and new practical applications. In: Document Recognition and Retrieval XII, International Society for Optics and Photonics, pp 84–91 (2005)
31. Shalev-Shwartz, S., Wexler, Y.: Minimizing the maximal loss: how and why. In: International Conference on Machine Learning, pp 793–801. PMLR (2016)
32. Shivangi, N., Adarsh, B., Shekhar, V., et al.: Real-strikeoff dataset. https://github.com/shivii/Real-Strike-off-dataset.git (2024)
33. Shonenkov, A., Karachev, D., Novopoltsev, M., et al.: Handwritten text generation and strikethrough characters augmentation. arXiv preprint arXiv:2112.07395 (2021)
34. Souibgui, M.A., Kessentini, Y.: De-gan: a conditional generative adversarial network for document enhancement. IEEE Trans. Pattern Anal. Mach. Intell. **5**, 55 (2020). https://doi.org/10.1109/TPAMI.2020.3022406
35. Tuganbaev, D., Deriaguine, D.: Method of stricken-out character recognition in handwritten text. US Patent 8,472,719 (2013)
36. Wadhwani, M., Kundu, D., Chakraborty, D., et al.: Text extraction and restoration of old handwritten documents. In: Digital Techniques for Heritage Presentation and Preservation, pp 109–132. Springer (2021)
37. Wigington, C., Stewart, S., Davis, B., et al.: Data augmentation for recognition of handwritten words and lines using a CNN-LSTM network. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), pp 639–645. IEEE (2017)
38. Zhu, J.Y., Park, T., Isola, P., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp 2223–2232 (2017)