Received: March 24, 2023. Revised: April 23, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Optimizing Image Rectangular Boundaries with Precision: A Genetic Algorithm-Based Approach with Deep Stitching

Muntasser A. Wahsh[1]* **Zainab M. Hussain**[2]

[1]*Department of Computer Science, Information Institute for Postgraduate Studies (IIPS),*
Iraqi Commission for Computers and Informatics (ICCI), Baghdad, Iraq
[2]*Department of Computer Engineering, Al-Mansour University College, Baghdad, Iraq*
* Corresponding author’s Email: phd202010549@iips.icci

Abstract: This paper presents a novel method for optimizing rectangular boundaries in stitched images using a combination of deep learning and genetic algorithms. The proposed method utilizes an enhanced structure of convolutional neural network (CNN) in order to extract features for residual regression and a genetic algorithm to tune CNN parameters. The performance of the proposed method was evaluated using various metrics, including mean squared error, peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Fréchet inception distance (FID). The outcomes from the conducted experiment demonstrated that the suggested method achieved high accuracy in rectangular boundary optimization, even for images with complex shapes and orientations. The method was also tested with different parameters, and results indicated that increasing the number of generations and population size led to improved performance. Also, the combination of deep learning algorithm specifically (CNN) and optimization algorithm specifically (genetic algorithm (GA)) led to an increase in the accuracy of the rectified images where the average PSNR reached 23.98, SSIM is 0.8066, and FID is 18.72. The proposed method has potential applications in various areas of study, such as image cropping, object detection, and image segmentation.

Keywords: Image stitching, Image rectifying, Deep learning, Genetic algorithm.

1. Introduction

As computer vision techniques continue to advance quickly, image stitching has gained significant attention as a research area. Additionally, due to its crucial role in creating realistic representations of scenes in some applications, image stitching has become one of the most commonly applied topics in computer vision and graphics. Over the past few years, the use of image stitching algorithms has become increasingly prevalent in several fields, including image processing, computer vision and multimedia, it depends on images from Unmanned Aerial Vehicles and also is used in their applications [1]. These algorithms have become intertwined with people's everyday lives, allowing for the creation of panoramic images with mobile phone applications, the production of wide field-of-view (FOV) videos for surveillance, and the development of assistive technologies for automobiles [1]. Photo stitching involves merging multiple images that have overlapping fields of view to create a segmented panorama or a high-resolution composite image. This process is typically carried out using computer software, and it is crucial for the images to have near-perfect overlap and identical exposure levels to achieve a seamless end result [2]. Image stitching has become a widely used technique in modern applications such as document mosaicking, image stabilization in camcorders through frame-rate image alignment, high-resolution photomosaics in digital maps and satellite imagery, medical imaging, multiple-image super-resolution imaging, video stitching, and object insertion. Most studies on image stitching have focused on resolving issues like identifying key points, extracting and matching features, estimating homography, addressing image misalignment and outlier problems, and performing image registration. However, stitched images may exhibit irregular boundaries due to the warping of images to correct any distortions that may occur during the stitching process. While irregular boundaries were once accepted in certain image stitching applications like remote sensing or medical images, it is now recognized as a limitation of the technique [4]. Therefore, users of image stitching had to accept the presence of irregular boundaries in the resulting images. Authors in [5] developed an unmanned aerial vehicle (UAV) picture stitching technique based on the best seam algorithm and half-projective warp, which successfully preserves the image's original details while producing the best stitching results. The ghosting and blurring issues on the stitched photos can be fixed using the current seam stitching techniques, but the deformation and angle distortion brought on by image registration will still be present. They provided an ideal seam and half-projective warp sewing approach to get over this problem. The colour, structural, and line difference information are first defined in a new difference matrix in the overlapped area of the aligned picture. Then, in order to acquire the seam, they suggested a seam search algorithm based on the global minimum energy and restrict the search range of the seam by the minimum energy. Lastly, the form of the stitched picture is corrected in order to preserve more areas' original shapes in conjunction with the seam position and half-projective warp. Their technique offers a superior stitching effect, as demonstrated by the testing results of multiple sets of UAV photos. When using content-aware fill rectifying, one of the main drawbacks is that it can produce artifacts or distortions in the image, which can be noticeable in certain cases. This can be especially problematic when dealing with complex textures or patterns, which may be difficult for the algorithm to replicate accurately.

Local and global warping methods can also introduce distortions in the image, especially when the warping is not applied correctly. In some cases, the result may look unnatural or even unrealistic, making it unsuitable for certain applications.

Using deep learning with random hyperparameters for rectifying irregular boundaries in stitched images can be a powerful technique, but it also has some drawbacks. The accuracy of the algorithm is highly dependent on the quality and quantity of the training data, as well as the chosen hyperparameters. This means that there is a risk of overfitting the model to the training data, which can lead to poor performance on new, unseen data. Additionally, training deep learning models can be computationally expensive and time-consuming, which can be a barrier to adoption for some users.

The demand for joining images with regular boundaries has become crucial in fields such as virtual reality [4]. To address this issue, a hybrid method that uses well-built convolutional neural network (CNN) with genetic algorithm will be proposed. By combining these two algorithms, it aims to achieve train with optimal values of parameter. The initial objective was to find a suitable model that would meet the requirements for obtaining rectangular images. To achieve this, a method [5, 6] was proposed to warp the stitched images to a rectangular shape using mesh deformation. However, previous methods were limited to deal with images containing straight structures, such as buildings, boxes, pillars, and so on. Images with nonlinear structures, like portraits, exhibited clear distortions as shown in Fig. 1. Additionally, the learning baseline utilized by [5] relied on randomly chosen initial parameters for model training, resulting in a loss of optimality.

To address this issue, this study implemented a genetic algorithm to obtain optimal initial parameters for achieving the best possible training outcomes of CNN. Specifically, the best learning rate and decay rate were tuned. One of the strengths of genetic algorithms is their ability to search a large and diverse search space efficiently. By using genetic operators such as mutation and crossover, genetic algorithms can explore different combinations of hyperparameters and architectures, which can lead to better performance compared to other optimization methods. Another advantage of using genetic algorithms is their ability to handle non-linear and non-convex optimization problems, which are common in CNN optimization. Also, the structure of CNN is also tested; many structures were tested to get the best structure; which led to reduced global loss, mesh loss, performance loss, perception loss, and mask loss. This study is divided into several sections, the first section, each of which addresses a specific aspect of the research topic. The first section provides an introduction to the topic. The second section

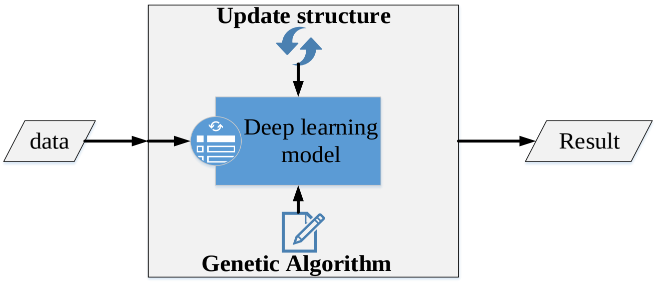

Figure. 1 Main approach

Received: March 24, 2023. Revised: April 23, 2023.

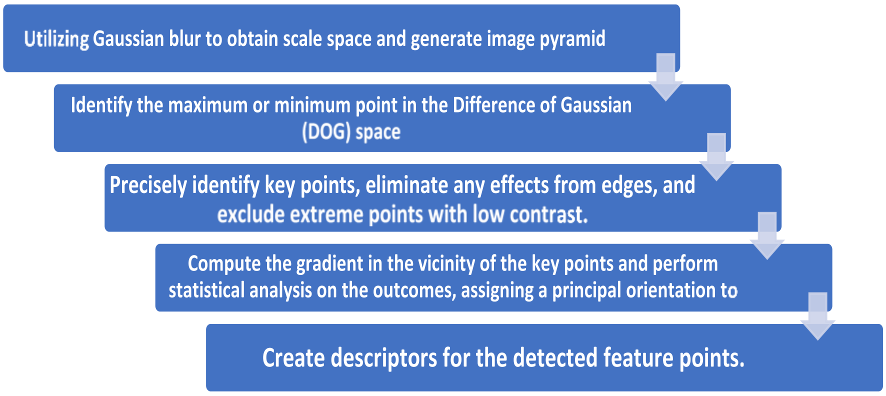
Utilizing Gaussian blur to obtain scale space and generate image pyramid
Identify the maximum or minimum point in the Difference of Gaussian (DOG) space
Precisely identify key points, eliminate any effects from edges, and exclude extreme points with low contrast.
Compute the gradient in the vicinity of the key points and perform statistical analysis on the outcomes, assigning a principal orientation to
Create descriptors for the detected feature points.

Figure. 2 SIFT features detection steps

presents the background of the study to give a clear look about the issue of irregular boundaries as a result of image stitching and the most popular methods to rectify it. The third section discusses the methodology followed in this study where how to choose best model structure with explanation, and the optimization algorithm (genetic algorithm and its usage as an optimizer is discussed. Results are introduced in the fourth section with details about training and testing, while section five discusses the experimental metrics (PSNR, SSIM, and FID) were used to evaluate the results. Finally, the conclusion is reviewed.

CNN with GA have been used in many fields, but in image stitching field, we thought that it is new. GA with CNN in this field with whole process that optimize parameters of training and parameters of models in feature extractions phase and feature regression models. In addition to all, best selection of model structure of both feature extraction and regression has been made.

The use of genetic algorithms in image stitching can improve the efficiency and effectiveness of the process, as it can quickly search a large search space of potential solutions. However, the performance of the algorithm can depend on several factors, such as the size and complexity of the input images, the number of rectangular regions, and the fitness function used to evaluate the solutions.

In general, genetic algorithms can provide a useful tool for optimizing the rectangular tiling of images in image stitching, but the results may need to be evaluated and refined further using other techniques to achieve the desired quality and accuracy.

2. Study background

To find out the reasons of irregular boundaries as a result of image stitching, the step-by-step process of image stitching will be discussed, from start to finish, and to identify the causes of irregular boundaries. Following this, various methods for obtaining rectangular images will be reviewed to provide a comprehensive understanding of the methods employed, as well as to identify the strengths and weaknesses of each method.

2.1 Image stitching

The majority of image stitching techniques fall into one of two categories: intensity-based methods or feature-based methods [4]. Intensity-based methods involve finding suitable similarity metrics to measure the similarity between the two input images, and then optimizing the overlaps of the images directly in order to maximize their similarity. Feature-based methods, on the other hand, are generally considered the more popular method for image stitching. In the following sections, the processes involved in image stitching will be discussed.

2.1.1. Feature detection algorithm

David Lowe proposed the scale invariant feature transformation (SIFT) algorithm in 1999, which was later improved in 2004 [9, 10]. This algorithm is used to detect local features in an image, also referred to as "key points". The SIFT algorithm can be broken down into several steps, which are discussed below: a) Feature point detection. b) Feature point localization. c) Orientation assignment. d) Feature descriptor generation.

Feature detection is the first and most important step in image stitching due to the importance of finding the similarity between two overlapped images and the rest of the process is highly dependent on it. Fig. 2 shows the steps of SIFT features detection According to [11]. In 1962, concept of scale space was introduced and developed. Gaussian blur was used, which is a linear and distinctive function, to create the scale space. By convolving the original image with the Gaussian function, a two-dimensional image's Gaussian scale space is obtained:

$$L(x\,,y\,,\sigma) = l(x\,,y)\ \otimes G(x,y,\sigma) \qquad (1)$$

$$(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\ exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) \qquad (2)$$

$L(x\,,y\,,\sigma)$ represents the scale space, grey value, and Gaussian function of the image respectively. Parameters x, y are the space coordinates; σ is the scale coordinate, which controls the picture sharpness. To detect the extreme point of scale space, it is essential to create a difference of Gaussian (DOG) to efficiently detect the extreme points in space. It is described as the ratio of two neighbouring scale images

$$D(x\,,y\,,\sigma) = L(x\,,y,k\sigma\,) - L(x,y,\sigma) \qquad (3)$$

$$= [G(x\,,y\,,k\sigma) - G(x\,,y\,,\sigma) \otimes I(x\,,y)] \qquad (4)$$

Several other feature detection algorithms have been created subsequently, including speed up robust feature (SURF), oriented BRIEF (ORB), and features from accelerated segment test (FAST), which are used to categorize images based on their distinct properties, such as brightness, image size, and rotation. Once these features have been identified, a descriptor is generated for each feature by calculating its relationship with the surrounding pixels, which describes the feature information [12].

2.1.2. Feature matching

To find a corresponding feature in two images, the descriptors are compared and the closest match is identified using feature matching. The RANSAC algorithm is utilized to select the matching model with the highest number of inliers and to remove any outliers that do not fit this model. A homography matrix is then generated to represent the relationships between the matching pairs of features in the images. Using this homography matrix [7]. The rotation matrix and focal length are computed for each image. To ensure that the key points are evenly distributed across both images, the nearest neighbours are identified. However, in some cases, the second-closest match may be almost as close as the first match due to noise or other factors. In such cases, the ratio of the nearest-to-second-nearest distances is used [8]. Figs. 3, 4, and 5 show the steps of finding key points and then match them.

2.1.3. Homography based warping and blending

Homography-based warping is an image processing technique used to convert an image from one coordinate system to another [11, 10]. This transformation is based on the homography idea, which is a matrix that represents the connection between two images of the same scene taken from various perspectives. The following steps are included in the homography-based warping process:

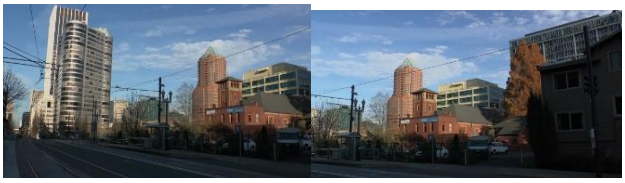

Figure. 3 Two overlapped images

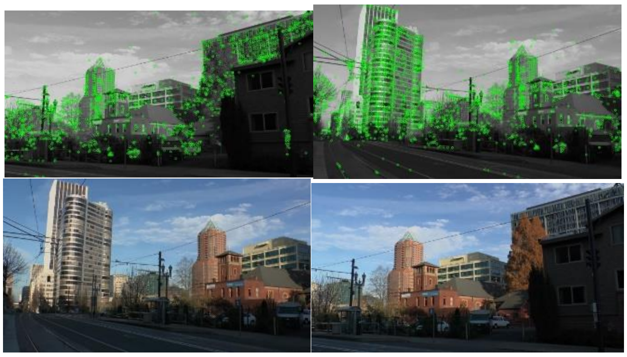

Figure. 4 Extreme points of the overlapped images

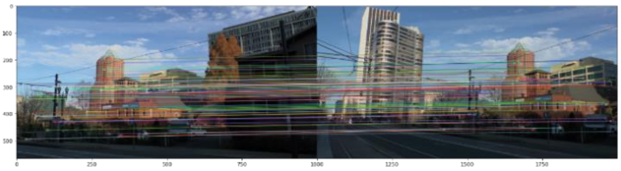

Figure. 5 Features matching

1. Recognize points that appear in both photos in the same position, such as the corners.
2. Compute the homography matrix to represent the transformation that converts points in one picture to points in the other.
3. To warp the picture, use the homography matrix. This requires repositioning each pixel in the original image to its new location in the converted image.

Homography-based warping is widely utilized in image stitching. It is especially beneficial for matching photos acquired from various perspectives since it allows for the correction of perspective distortion as well as the alignment of features in the two images [13, 14]. The result will be a wider view stitched image that can be treated as one image. However, visible seams or ghosts in the pre-aligned image may arise as a result of the exposure difference or a moving object in the overlapped region. As a result, the stitching quality suffers significantly as shown in Fig. 6.

The warping is done by regenerating the

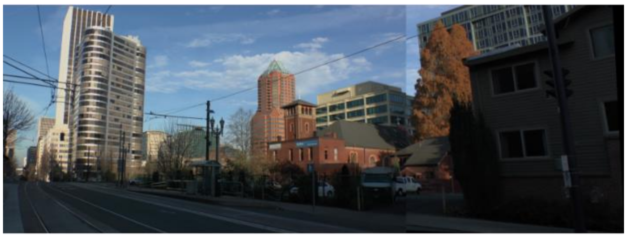

Figure. 6 Two stitched images with different exposure

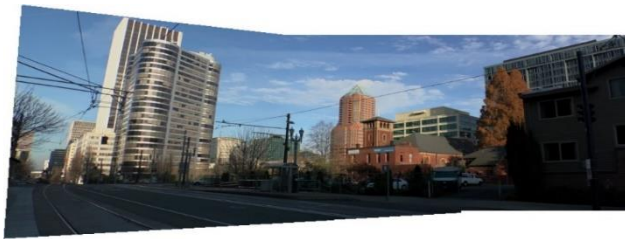

Figure. 7 stitched, warped, and blended image

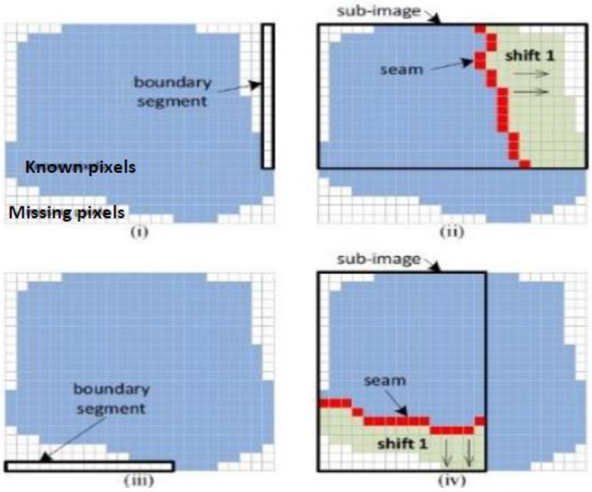

Figure. 8 Local seam carving approach

projective image by transforming each pixel at the appropriate height, projected to a ground plane with a specific quantity [13]. While blending techniques are utilized to unify the colors of the image, the final result can be seen in Fig. 7. Irregular boundaries are clearly shown in Fig. 6.

$$s\begin{bmatrix}\bar{x}\\ \bar{y}\\ 1\end{bmatrix} = H\begin{bmatrix}x\\ y\\ 1\end{bmatrix} = \begin{bmatrix}h_{11} & h_{12} & h_{13}\\ h_{21} & h_{22} & h_{23}\\ h_{31} & h_{32} & h_{33}\end{bmatrix}\begin{bmatrix}x\\ y\\ 1\end{bmatrix} \quad (5)$$

2.2 Image rectifying

Upon examining Fig. 8, it is evident that the resulting image exhibits irregular boundaries due to the warping process that is necessary to address any stretching or distortion that may have occurred during image stitching. As previously mentioned in this research, certain applications that employ image stitching techniques require users to tolerate irregular boundaries [14]. To address this issue, a number of solutions have been suggested to obtain rectangular boundaries. These solutions range from basic techniques such as cropping [15] the image to more complex deep learning algorithms that aim to eliminate irregular boundaries [16], known as image rectifying. While cropping the image can produce a rectangular image, important content may be lost in the process, which may be relevant to the end user. Thus, this technique is not recommended. In the upcoming sections, some of the proposed methods for achieving rectangular stitched images will be discussed.

2.2.1. Content-aware fill rectifying

This method for image rectifying relies heavily on the content of the irregular boundary image. According to [17], This method for image rectifying relies heavily on the content of the irregular boundary image. The process begins by identifying the longest segment of the boundary that needs to be filled. The next step is to compute the seam to find the highest contour of the sub-image that corresponds to the longest boundary. Next, all pixels are shifted in a single direction by one pixel. Finally, the next seam is computed, and the previous steps are repeated. Fig.8 illustrates the steps involved in content-aware image rectifying, which operates by using the concept of seam carving for local warping.

One of the drawbacks of this method is that it is only effective for images whose seams can be duplicated, such as those with homogeneous patterns in the seam area [18]. Moreover, the seam carving method may not be suitable for images with sharp edges or fine details, such as those with text or faces, as removing the seams can cause blurring or distortion in these areas, affecting the overall quality of the image and distortions will occur in other images, such as those shown in Figs. 9, 10, and 11.

2.2.2. Local and global warping rectifying

The alternative method proposed for handling distortion in image rectifying, as described in the statement, is based on local and global warping. This method was developed by [6] using local warping. This method was based on previous studies on image completion, image retargeting, and warping. Image completion involves generating panoramic images by filling in missing regions using parts copied from other known regions to synthesize the missing content of the image [21-23]. The local warping step in this method involves identifying and removing

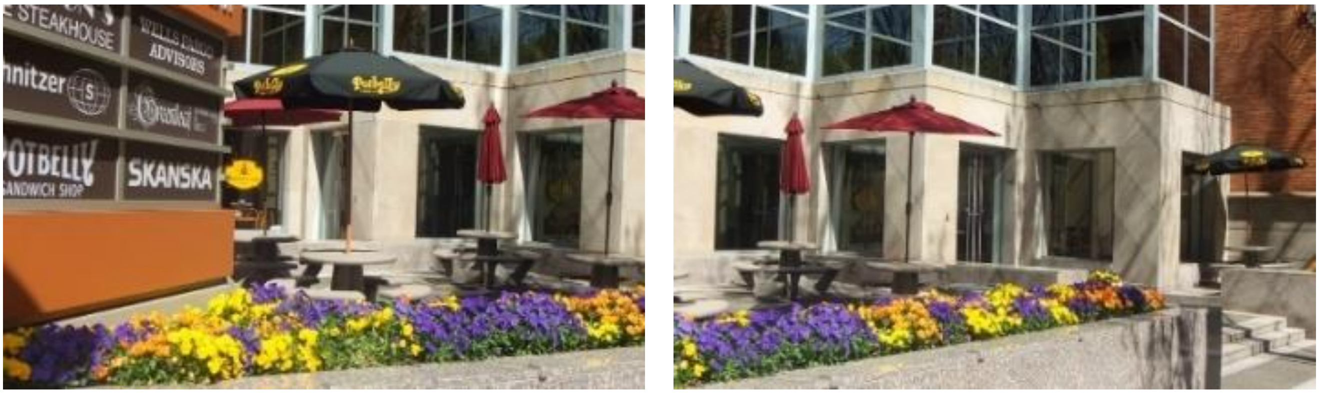

Figure. 9 Local input images

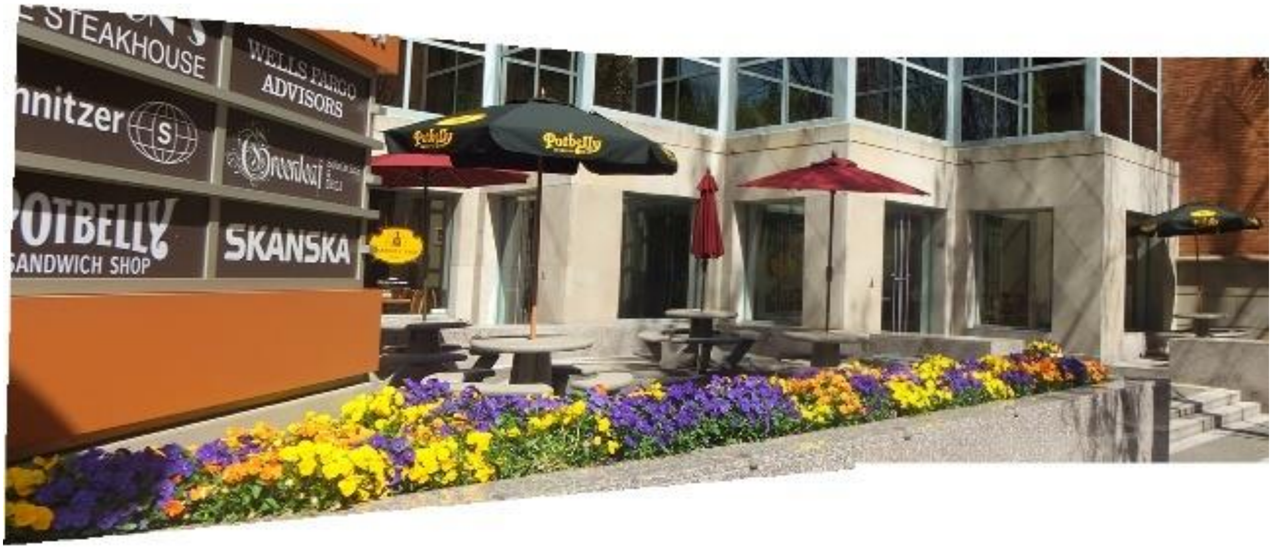

Figure. 10 Input images after stitching

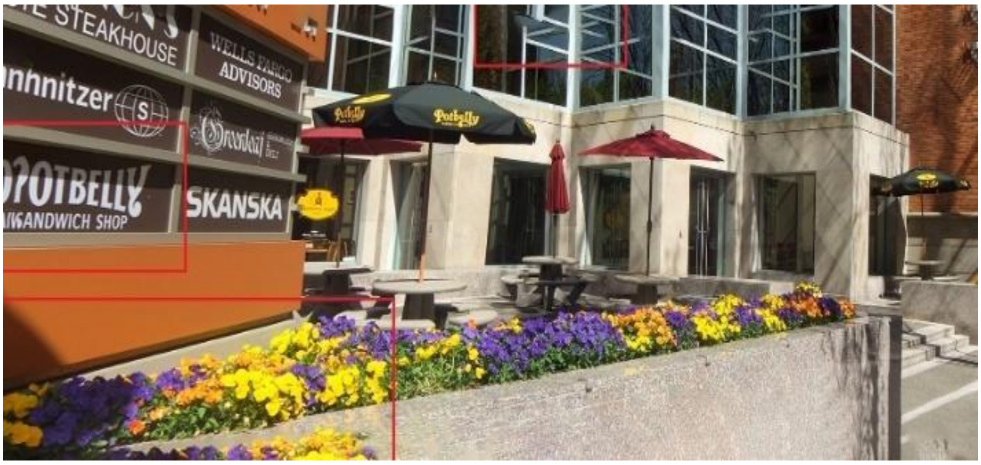

Figure. 11 image rectifying based on local warping via seam carving

seams from the image using the seam carving technique. Seam carving is a content-aware resizing technique that involves removing or duplicating seams from the image based on their importance to the content. By using seam carving for local warping, the algorithm can ensure that the image is rectified without introducing additional distortions.

The global warping step in this method uses a mesh-based approach to further refine the image rectification. Mesh-based warping involves mapping the image to a grid of vertices and edges that can be manipulated to deform the image. In this method, the mesh is drawn on the distorted image and then warped backward to obtain a mesh that fits the input image. This process can help to correct any remaining distortions in the image that were not addressed by

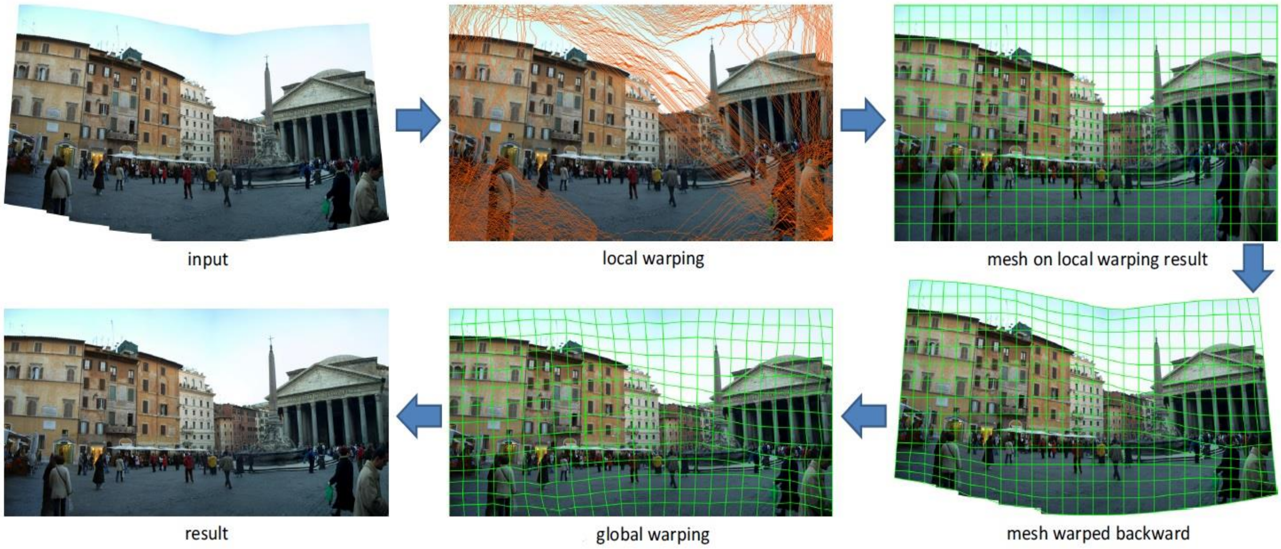

Figure. 12 The pipeline of local and global warping algorithm (The source: He. Et al.)

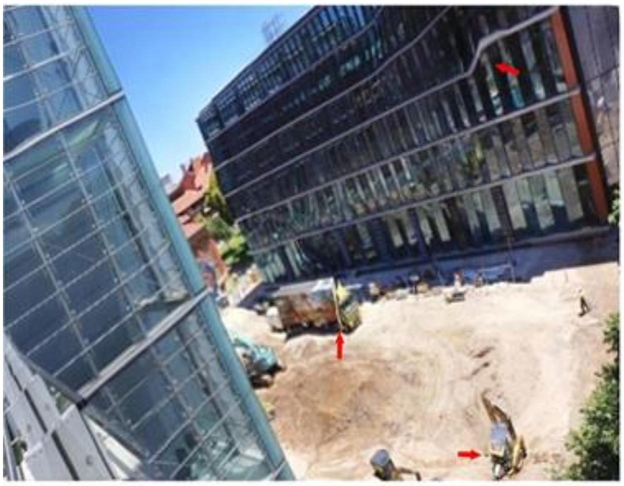

Figure. 13 Distortion after applying local and global warping algorithm

the local warping step.

The final step in this method is mesh optimization, which involves adjusting the vertices and edges of the mesh to obtain a rectangular boundary image. Mesh optimization can be performed using a variety of techniques, such as energy minimization or geometric constraints. The result is a rectangular image that is free from distortions and can be used for further processing or analysis. Fig. 12 illustrates the pipeline of the algorithm.

The local and global warping method, as proposed by [5], is known to be computationally expensive due to its two-step process. Additionally, the energy function proposed by the same study is limited in its ability to handle linear structures in panoramas with ERP format. As illustrated in Fig. 10, the energy function can only handle linear structures, which may result in straight lines appearing bent in panoramas with ERP format. To address this limitation, [20] introduced a line-preserving energy term to the geodesic-preserving energy term. The line-preserving energy term improves the accuracy of the final panorama by preserving the straightness of lines in the input images during the warping process. However, despite this improvement, the geodesic lines cannot be detected from the stitched image, and the improvement can only be directly observed in the panorama. Furthermore, the local and global warping method may not be suitable for real-time processing due to its computational complexity. Therefore, feature-based approaches may be a more suitable alternative for applications that require real-time processing.

2.2.3. Image rectifying based on deep learning

With the advancement of computer hardware and the capability to collect, analyse, and extract valuable insights from large amounts of data, many applications have turned to AI algorithms to obtain accurate and dependable results. This trend is not excluded in the fields of computer vision and image processing. As a result, [5] proposed a framework that utilizes a one-step algorithm to obtain a rectangular image boundary, as opposed to the two-step process of local and global warping. This method involves using a neural network to predict the initial mesh, given the stitched image. A predefined mesh with a rigid shape is then used as the target mesh. Furthermore, the computation of the backward interpolation can be expedited by utilizing the rigid mesh shape through simple matrix computation [21].

3. Methodology

The methodology used in this study is outlined

Received: March 24, 2023. Revised: April 23, 2023.

beginning with an explanation of the best selection of deep learning model. The embarking point is from [5]. The methodology consists of three main components: feature extraction, mesh motion regression, and residual progressive regression. Feature extraction involves identifying key features in the image that are important for the deformation correction process. Mesh motion regression involves predicting the movement of the mesh points that define the image. Residual progressive regression involves using residual networks to refine the deformation correction. The objective function is a critical component of the methodology as it guides the entire process. The objective function is designed to minimize the error between the predicted and ground-truth deformation maps. This helps to ensure that the resulting image is as close as possible to the original, without any significant deformation. Finally, the study proposes an optimization technique that uses genetic algorithms to refine the deformation correction.

Genetic algorithms are a type of heuristic search algorithm that mimics the process of natural selection to find the optimal solution to a problem. The proposed optimization technique aims to improve upon previous studies' outcomes by refining the deformation correction process further.

3.1 Model structure

The main goal of the model is to output the initial mesh motion. To achieve this, the input includes both the stitched image and its mask, similar to previous studies [28, 29], The image is processed to extract features by a series of eight convolutional layers with filters set to 64, 64, 64, 64, 128, 128, 128, 128 from layer 1 to layer 8 respectively, and max-pooling is applied after every two convolutional layers. An adaptive pooling layer is used after feature extraction to adjust the resolution of extracted features. A convolutional structure is then used to predict the horizontal and vertical motion of each vertex based on the regular mesh, resulting in an output volume size of (U + 1) × (V + 1) × 2. The residual progressive regression strategy is employed to improve the accuracy of mesh motion estimation. The warped image is not used directly as input to a new network to avoid doubling computational complexity. Instead, intermediate feature maps are warped, which increases speed with a minor increase in computation. Two regressors are created using the same structure to predict primary mesh motions and residual mesh motions, respectively, even though they are designed for different tasks due to the differences in input characteristics. The proposed methodology consists of a feature extractor, mesh motion regressor, and residual progressive regression, and it is explained in detail in this section. Figure below illustrate the overall structure to rectify stitched images.

3.2 Objective function

In order to evaluate the effectiveness of a network, a mathematical equation called an objective function is created based on the methodology outlined in [14] as follow:

$$\lambda_{total} = \lambda_b + \lambda_m + \lambda_c \tag{6}$$

Where λ_b, λ_m, and λ_c *are* boundary constraint, mesh constraint, and content constraint respectively.

3.2.1. Content constraint

Previous methods such as [8, 28] were unable to handle non-linear structures due to their focus on preserving straight or geodesic lines in order to maintain image content. To address this limitation, the proposed method employs content-preserving learning from two perspectives: appearance loss and perception loss. For appearance loss, a formula is used to ensure that the rectified image output is similar to the rectangular image label.

$$\ell_c^a = ||\mathcal{R} - w\left(I, m_p\right)|| + ||\mathcal{R} - w\left(I, m_f\right)|| \tag{7}$$

$w(x, y)$ is the operation of warp. On the other hand, to ensure that the results are perceptually accepted, the distance between the rectangular images and rectangular labels as shown in Eq. (8):

$$\ell_c^p = ||\varphi(\mathcal{R}) - \varphi(w(I, m_c))|| + ||\varphi(\mathcal{R}) - \varphi(w\left(I, m_f\right))|| \tag{8}$$

Where φ(x) is the feature extraction operation from the VGG19 layer ('conv4-2) [30]. By this method, multiple perceptual properties can be perceived and not only linear structures. In brief, the content loss is produced by emphasizing both the similarity in appearance and semantic perception, which can be expressed as shown in Eq. (9):

$$\ell_c = w_a \ell_c^a + w_p \ell_c^p \tag{9}$$

3.2.2. Mesh constraint

In order to prevent distortion of the contents in rectangular images, it is crucial to ensure that the predicted mesh is not excessively deformed. To tackle this issue, two types of constraints have been introduced - an intra-grid constraint and an inter-grid constraint - to maintain the shape of the deformed mesh. The intra-grid constraint involves placing restrictions on the magnitude and direction of grid edges within a grid. For instance, horizontal edges are encouraged to be projected towards the right and have a norm greater than a threshold value (α_v^W) that corresponds to the resolution of the stitched image (H×W). This constraint is applied to ensure that the mesh maintains its rectangular shape and that no part of the content is excessively stretched or compressed.

Additionally, an inter-grid constraint has also been incorporated to ensure that neighbouring grids transform consistently. This constraint is designed to maintain consistency between adjacent grids so that they blend together seamlessly, without any noticeable discontinuities or abrupt changes in the content. By enforcing both intra-grid and inter-grid constraints, the algorithm can produce accurate and visually appealing results.

3.2.3. Boundary constraint

Concerning the boundary term, instead of limiting the predicted mesh, a binary mask is restricted. To do this, a binary mask is obtained from a combined image, warped, and then constrained to be similar to a matrix E that contains all ones. This ensures that the boundaries of the binary mask remain consistent with the original image.

To elaborate further, in order to handle the boundary term, a binary mask is used instead of the predicted mesh. This binary mask is obtained by combining multiple images and then applying warping to it. The warped binary mask is then constrained to be similar to a matrix E, which contains all ones. By doing this, the binary mask is kept consistent with the original image, and the boundaries of the mask remain unchanged.

$$\ell_c^a = \left|\left|E - \mathcal{w}\left(M, m_p\right)\right|\right|_1 + \left|\left|E - \mathcal{w}\left(M, m_f\right)\right|\right|_1 \tag{10}$$

The model has been structured after all with next specifications, the model of feature extraction:

- Adding batch normalization between layers after each convolution and pooling layers.
- Relu activation function for each convolution layer

The model of regression:

- Build regression layer as network
- Add batch normalization to after each convolution layer
- Small genetic algorithms were used to tune then kernel size of convolution layer and pooling layers in both feature extraction and regression net
- Kernal size of convolution (3) (bigger size is useless according to GA and smaller size does not make model converge)
- Kernal size of pooling (2) (bigger size is useless according to GA and smaller size does not make model converge)

Each model seems like structure in Fig. 14

3.3 Optimization

After building a CNN model, machine learning researchers manually adjust various externally set parameters, which include the trained CNN network and hyperparameters. Hyper parameters refer to the variables that are adjusted to optimize the performance of the model. Genetic algorithm, a metaheuristic algorithm inspired by the natural selection process, is commonly used for generating a group of solutions for optimizing and searching problems, such as the Traveling Salesman Problem (TSP). In this study, the genetic algorithm is used to determine the best learning rate, decay rate, and mesh and mask loss that produce the minimum appearance and perception loss. Choosing the appropriate hyper parameters is crucial to ensure that the model is optimized and produces optimal performance measurements. To further elaborate, hyper parameters are a set of variables that affect the model's learning process, and they are typically adjusted manually by machine learning researchers. These hyper parameters include learning rate, decay rate, and loss functions, among others. The goal of hyper parameter tuning is to find the best combination of these parameters to optimize the model's performance. In this study, the genetic algorithm is used to determine the optimal values for these hyper parameters. The genetic algorithm is a metaheuristic algorithm that mimics the natural selection process to generate a group of solutions for optimization problems. By using the genetic algorithm, the researchers can efficiently search the hyper parameter space to find the best combination of hyper parameters that minimize the appearance and perception loss. Fig. 15 depicts the model with the optimization algorithm, which shows how the genetic algorithm is used to optimize the model's performance.

The Table 2 below displays some of the results that were generated using genetic algorithm. Random iterations were selected to show learning rate and

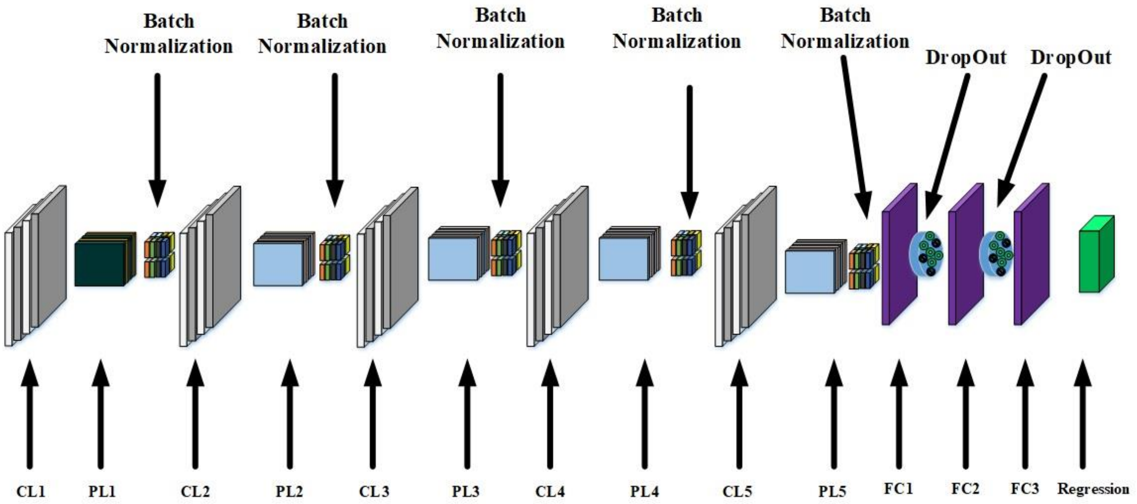

Figure. 14 Models of feature extraction and regression

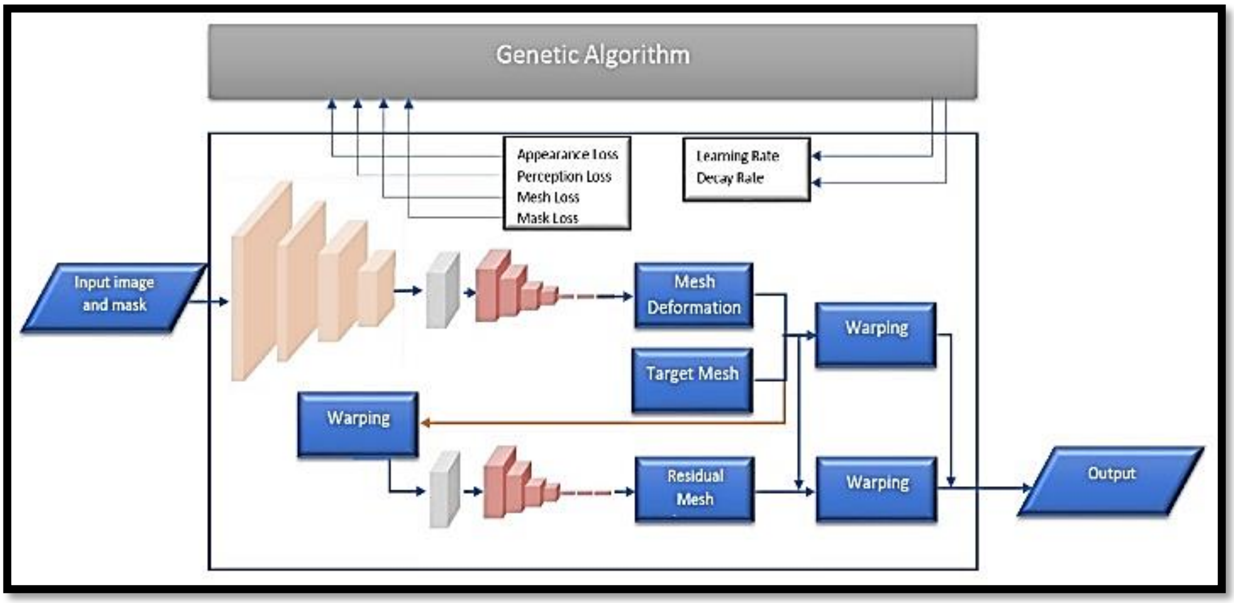

Figure. 15 Using genetic algorithm as an optimizer

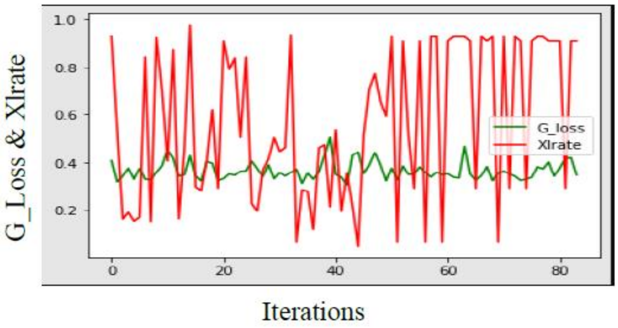

Figure. 16 Graph of learning rate and decay rate

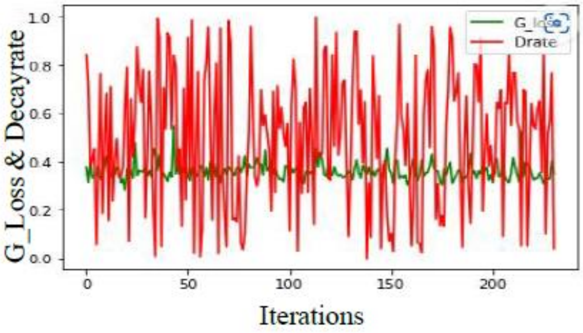

Figure. 17 Graph of learning rate and decay rate

decay rate based on the given loss functions. Figs. 16 and 17 shows the graph of the table analysing.

4. Experiments

It is important to outlines the details of the model training, including the neural network parameters, hardware, dataset, and number of iterations used. Initially, to ensure a fair comparison of the proposed technique with existing methods, it is important to provide detailed information about the dataset used in

Table 1. Experiments settings

Activation function	No. of training images	No. of testing images	Batch size		Mesh Dimensions U×V
			Training	Testing	
Relu	5839	519	4	1	16×12

this study. Since there are two types of datasets, structured and non-structured, the used dataset is unstructured data, it has the true images, images with irregular boundaries and irregular boundaries as mask. It is used by a study and it is mentioned in the study. The dataset consists of 5839 images that were collected for the purpose of rectifying irregular boundaries in images. The images were obtained from various sources, including public databases and in-house collections. Each image is in JPG format and has a resolution of 384 x 512 pixels. The images are of varying quality, with some exhibiting noise, blurring, or other artifacts. The images are annotated with bounding boxes that indicate the irregular boundaries that need to be rectified. The dataset is diverse, with images representing a range of objects, scenes, and lighting conditions. Before using the dataset, preprocessing was applied steps such as cropping, resizing, and color normalization to ensure consistency and improve the quality of the images. RELU was used as an activation function except last layer without activation function. Batch size set to 4 to avoid the issue of exceeding the limit of the memory while U and V set to 16, 12 respectively, and number of training iterations is 85K. Table below shows the experiments settings.

Regarding the genetic algorithm parameters, the population size set to 500, while number of iterations was 500, mutation probability is 0.01, elite ratio equal to 0.001, and crossover probability is 0.5 The laptop in which was used to finish the training was use 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz with 16 GB RAM and NVIDIA GeForce RTX 3060 laptop GPU and the operating system is windows 10 home edition 64 bit. Based on adopted dataset in this study, the proposed algorithm (CNN) with the optimization algorithm (Genetic algorithm) were tested in different datasets to ensure that this approach is valid with different datasets that consist image stitching by using different method such as (SPW[22] , LCP[23], and UDIS[24]). To track the enhancement of PSNR (peak-to-signal-noise-ratio) and SSIM (structural similarity index measure) of resulted images, the training conducted in stages and the results were saved every 5000 iterations. Training, results and the comparison with related studies will be discussed in the next section.

5. Experimental metrics

The experimental measurements used to evaluate the results of the paper include peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and fréchet inception distance (FID), which are widely used in image processing. PSNR is calculated as follows:

$$PSNR = 10 \times log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{11}$$

where MAX is the maximum possible pixel value and MSE is the mean squared error between the original and reconstructed images. Higher PSNR indicates better image quality. While SSIM is calculated as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{12}$$

Where μ_x and μ_y are the mean pixel values of x and y, σ_x and σ_y are their standard deviations, $\sigma_{x,y}$ is their covariance, and C1 and C2 are small constants to avoid division by zero. Higher SSIM indicates better perceptual quality. Regarding FID, it can be calculated as follows:

$$FID(x,y) = \left\|\mu_x - \mu_y\right\|^2 + T_r\left(\sigma_x + \sigma_y - 2\left(\sigma_x\sigma_y\right)^{\frac{1}{2}}\right) \tag{13}$$

where x and y are the feature representations of the generated and real images, respectively, μ_x and μ_y are their mean values, σ_x and σ_y are their covariance matrices, and T_r is the trace operator. Lower FID indicates better image quality. Overall, these metrics provide a comprehensive evaluation of the quality of the generated images and help to validate the effectiveness of the proposed method in the paper.

6. Results

The study involved using a combination of convolutional neural network (CNN) and genetic algorithm to rectify images, and the results were found to be promising. The method was tested on a dataset of diverse images with varying shapes, content, and orientations and was able to accurately rectify them. The model was constructed by utilizing a pre-trained VGG19 network, (a type of

Table 2. Learning rate and decay rate that gives minimum losses

Iterations	Learning Rate	Decay Rate	Global Lose	Appearance Lose	Perception Lose	Mask Loss	Mesh Loss
10	0.00001	0.81	0.3379	0.2696	0.0615	0.0051	0.0017
20	0.00012	0.83	0.3292	0.2581	0.0655	0.0041	0.0015
30	0.00043	0.87	0.2227	0.1456	0.073	0.0029	0.0012
40	**0.0005**	**0.82**	**0.1608**	**0.0884**	**0.0682**	**0.0029**	**0.0013**
50	0.0006	0.89	0.2365	0.1673	0.0648	0.0028	0.0017
100	0.00062	0.91	0.4254	0.3012	0.1194	0.003	0.0018
200	0.00075	0.936	0.6618	0.5181	0.1362	0.0045	0.003
300	0.0008	0.953	0.7067	0.611	0.091	0.0027	0.002
400	0.001	0.971	0.7011	0.5986	0.0954	0.0041	0.003
500	0.002	0.991	0.8495	0.6485	0.1929	0.005	0.0031

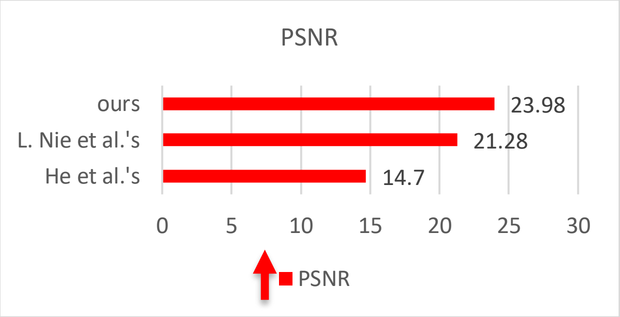

Figure. 18 PSNR Comparison

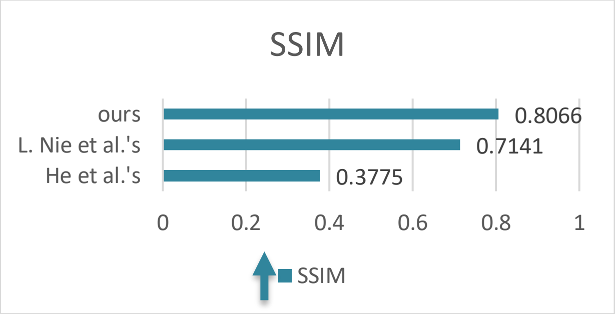

Figure. 19 SSIM Comparison

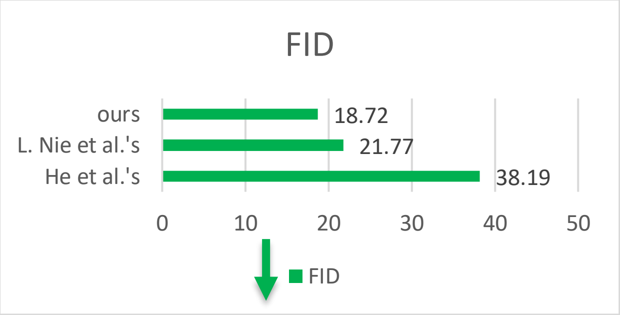

Figure. 20 FID Comparison

convolutional neural network (CNN)), that has been pre-trained on a large dataset of images, and it can be fine-tuned for specific image recognition tasks such as object detection or image segmentation for feature extraction where it takes an input tensor as a parameter, which represents an image to be processed. The input tensor is then passed through several layers of convolution and pooling, resulting in a set of feature maps that capture the high-level image features. While the genetic algorithm was employed to train the model using optimal values for CNN parameters that minimized the loss. The performance of the method was evaluated using several metrics, including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Fréchet inception distance (FID). The results indicate that the proposed method achieved a PSNR of 23.98 dB and an SSIM of 0.81. These results demonstrate the effectiveness of the proposed method in accurately rectifying images, even when they have complex shapes and orientations. A comparison of the achieved results with the related studies [7,8] results are shown in Figs. 18, 19, and 20. Furthermore, the method's performance was tested by conducting experiments with various parameters, including the number of generations and population size in the genetic algorithm. The findings revealed that augmenting the number of generations and population size resulted in better algorithmic performance.

7. Conclusion

To sum up, this research presented a method for rectifying projected stitched images to achieve regular boundaries by developing a model with optimal learning rate and decay rate values that minimize appearance loss, perception loss, mesh loss, and mask loss. Moreover, the outcomes indicated that this method is effective for rectifying images with straight and non-straight content, which may otherwise be distorted during the rectification process due to mesh. In addition to the findings described above, this study provides a significant contribution to the field of image processing by presenting a novel approach for rectifying irregular boundaries in images. By combining deep learning and genetic algorithm techniques, a method has been developed that achieves state-of-the-art performance in terms of accuracy and efficiency. The comprehensive evaluation, including experiments on a large and diverse dataset, demonstrates the effectiveness of proposed approach in real-world scenarios. It is believed that the proposed method has the potential to be used in a wide range of applications, including medical imaging, remote sensing, and industrial quality control. The results of this study provide a foundation for future research in this area and demonstrate the potential for further advancements in image processing.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

Muntasser A. Wahsh provided the idea, technique, software, formal analysis, materials, data collection, and writing-original version preparation. Zainab M. Hussain provided supervision, revision, and editing.

References

[1] X. Fan, J. Lei, Y. Fang, Q. Huang, N. Ling, and C. Hou, “Stereoscopic Image Stitching via Disparity-Constrained Warping and Blending”, *IEEE Trans. Multimed*, Vol. 22, No. 3, pp. 655–665, 2020.

[2] W. Lyu, Z. Zhou, L. Chen, and Y. Zhou, “A survey on image and video stitching”, *Virtual Real. Intell. Hardw*, Vol. 1, No. 1, pp. 55–83, 2019.

[3] X. Li, L. Wu, G. Wang, Y. Chen, K. Rong, and J. Lin, “Image Stitching and Quality Evaluation Algorithm for Large Size Parts”, *ACM Int. Conf. Proceeding Ser*, pp. 130–134, 2017.

[4] M. F. Sabzevar, M. Gheisari, and J. Lo, “Development and Assessment of a Sensor-Based Orientation and Positioning Approach for Decreasing Variation in Camera Viewpoints and Image Transformations at Construction Sites”, *Appl. Sci*, Vol. 10, No. 7, p. 2305, 2020.

[5] J. Li, K. Yu, Y. Zhao, Y. Zhang, and L. Xu, “Cross-reference stitching quality assessment for 360◦ omnidirectional images”, In: *MM 2019 - Proc. of 27th ACM Int. Conf. Multimed*, pp. 2360–2368, 2019.

[6] M. Meng and S. Liu, “High-quality Panorama Stitching based on Asymmetric Bidirectional Optical Flow”, In: *Proc. of 2020 5th Int. Conf. Comput. Intell. Appl. ICCIA 2020*, No. Iccia, pp. 118–122, 2020.

[7] L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, “Deep Rectangling for Image Stitching: A Learning Baseline”, In: *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5740–5748, 2022.

[8] K. He, H. Chang, and J. Sun, “Rectangling panoramic images via warping”, *ACM Trans. Graph*, Vol. 32, No. 4, 2013.

[9] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, *Int. J. Comput. Vis*, Vol. 60, No. 2, pp. 91–110, 2004.

[10] Y. G. X. Wenbo and W. Lian, “Distributed application based onSIFT feature image retrieval [J]”, *J. Guizhou Teach. Coll*, No. 32, pp. 13–17, 2016.

[11] Z. Zhang, “Image stitching algorithm based on combined feature detection”, In: *Proc. of 2020 IEEE Int. Conf. Adv. Electr. Eng. Comput. Appl. AEECA 2020*, pp. 966–971, 2020.

[12] H. Lee, S. Lee, and O. Choi, “Improved method on image stitching based on optical flow algorithm”, *Int. J. Eng. Bus. Manag*, Vol. 12, No. 17058, pp. 1–17, 2020.

[13] Y. Zhang, Z. Wan, X. Jiang, and X. Mei, “Automatic Stitching for Hyperspectral Images Using Robust Feature Matching and Elastic Warp”, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*, Vol. 13, pp. 3145–3154.

[14] C. Ravi and R. M. Gowda, “Development of Image Stitching Using Feature Detection and Feature Matching Techniques”, In: *Proc. of 2020 IEEE Int. Conf. Innov. Technol. INOCON 2020*, pp. 1–7, 2020.

[15] Y. Chen, H. Zheng, Y. Ma, and Z. Yan, “Image stitching based on angle-consistent warping”, *Pattern Recognit*, Vol. 117, p. 107993, 2021.

[16] Q. Zhao, Y. Ma, C. Zhu, C. Yao, B. Feng, and F. Dai, “Image stitching via deep homography estimation”, *Neurocomputing*, Vol. 450, pp. 219–229, 2021.

[17] H. Shi, L. Guo, S. Tan, G. Li, and J. Sun, “Improved parallax image stitching algorithm based on feature block”, *Symmetry (Basel)*, Vol. 11, No. 3, pp. 1–19, 2019.

[18] L. Deng, Y. Piao, and S. Liu, “Research on SIFT image matching based on mlesac algorithm”, In: *ACM Int. Conf. Proceeding Ser*, Vol. 2, No. 1, pp. 37–42, 2018.

[19] L. Song, J. Wu, M. Yang, Q. Zhang, Y. Li, and J. Yuan, “Stacked Homography Transformations for Multi-View Pedestrian

Detection", In: *Proc. of IEEE Int. Conf. Comput. Vis*, pp. 6029–6037, 2021.

[20] W. Zhang, Y. Wang, and Y. Liu, "Generating High-Quality Panorama by View Synthesis Based on Optical Flow Estimation", *Sensors*, Vol. 22, No. 2. 2022.

[21] J. Chen, G. Bai, S. Liang, and Z. Li, "Automatic image cropping: A computational complexity study", In: *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 507–515, 2016.

[22] C. H. Chao, P. L. Hsu, H. Y. Lee, and Y. C. F. Wang, "Self-supervised deep learning for fisheye image rectification", In: *Proc. of ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2248–2252, 2020.

[23] X. Zheng, X. Qiao, Y. Cao, and R. W. H. Lau, "Content-aware generative modeling of graphic design layouts", *ACM Trans. Graph*, Vol. 38, No. 4, 2019.

[24] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "CARAFE++: Unified Content-Aware ReAssembly of FEatures", *IEEE Trans. Pattern Anal. Mach. Intell*, Vol. 44, No. 9, pp. 4674–4687, 2022.

[25] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting", *IEEE Trans. Image Process*, Vol. 13, No. 9, pp. 1200–1212, 2004.

[26] T. Igarashi, T. Moscovich, and J. F. Hughes, "As-rigid-as-possible shape manipulation", *ACM Trans. Graph*, Vol. 24, No. 3, pp. 1134–1141, 2005.

[27] R. Carroll, A. Agarwala, and M. Agrawala, "Image warps for artistic perspective manipulation", *ACM SIGGRAPH 2010 Pap. SIGGRAPH 2010*, Vol. 1, No. 212, pp. 1–9, 2010.

[28] D. Li, K. He, J. Sun, and K. Zhou, "A geodesic-preserving method for image warping", In: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, Vol. 07-12-June, pp. 213–221, 2015.

[29] L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, "Depth-Aware Multi-Grid Deep Homography Estimation with Contextual Correlation", *arXiv Prepr. arXiv2107.02524*, 2021.

[30] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky, "Resolution-robust Large Mask Inpainting with Fourier Convolutions", In: *Proc. of 2022 IEEE/CVF Winter Conf. Appl. Comput. Vision, WACV 2022*, pp. 3172–3182, 2022.

[31] P. Teterwak, A. Sarna, D. Krishnan, A. Maschinot, D. Belanger, C. Liu, and W. Freeman, "Boundless: Generative adversarial networks for image extension", In: *Proc. of IEEE Int. Conf. Comput. Vis*, Vol. 2019-Octob, pp. 10520–10529, 2019.

[32] L. Xie and A. Yuille, "Genetic CNN", In: *Proc. of the IEEE International Conference on Computer Vision*, pp. 1379–1388, 2017.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv Prepr. arXiv1409.1556*, 2014.

[34] T. Liao and N. Li, "Single-perspective warps in natural image stitching", *IEEE Trans. image Process*, Vol. 29, pp. 724–735, 2019.

[35] Q. Jia, Z. Li, X. Fan, H. Zhao, S. Teng, X. Ye, and L. Latecki, "Leveraging line-point consistence to preserve structures for wide parallax image stitching", In: *Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12186–12195, 2021.

[36] L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, "Unsupervised deep image stitching: Reconstructing stitched features to images", *IEEE Trans. Image Process*, Vol. 30, pp. 6184–6197, 2021.