Chapter 6

Inpainting For Image Enhancement

In computer vision, image inpainting is crucial in various research areas, such as restoring damaged or old images, removing unwanted objects, and filling in missing regions using the surrounding pixels [148]. It involves analyzing the surrounding context and leveraging the image inherent structures, patterns and texture to generate plausible completion. Initially, various filters are used with many iterations applied to identify the missing portion on the images. Later, machine learning and advanced techniques algorithms were used to generate the realistic and visually appealing reconstructed process. This chapter focuses on filling the segmented region with the neighboring pixel to enhance the standard of the images. So, this chapter proposed the "*Bilateral-based Convolutional Inpainting Model"* for filling up the eliminated region with the adjacent pixels. The proposed method is compared with the existing CNN inpainting method to check the performance analysis. Additionally, the proposed method is implemented to the other digital images to remove a specific object from the images to evaluate its performance further. It is also analyzed on different loss functions and different datasets. Based on the analysis, the proposed model outperforms the existing method, filling the smart colposcopy images with higher quality.

6.1. Convolutional Neural Network Models for the Filling up the Eliminated reflection region on Smart Colposcopy Images

This section explores existing techniques for inpainting tasks, specifically for filling random missing image pixels. The model like P.Conv, GMCNN, and D.Conv are applied to smart colposcopy images to replace the removed glare with neighboring pixels. The initial method is the partial convolutional neural network (P.Conv), which utilizes convolutional neural networks with partial convolutions to handle missing data. By considering only valid pixels in the convolutional filters, the network can effectively pinpoint the missing regions in an image. This approach finds the available neighboring pixels to generate plausible completions for the missing areas. The second method is the generative multicolumn neural network, a traditional convolutional neural network variant. This technique incorporates multiple columns within the network architecture, enhancing inpainting capabilities. The network is trained to learn the underlying patterns and structures in the image data and

generate accurate predictions for the missing pixels based on the neighboring information. The third technique is the dilated convolutional neural network, which employs dilated convolutions to expand the receptive field of the network. It helps to capture the larger context from the neighboring pixels, enabling more accurate inpainting results. By adjusting the dilation rates of the convolutions, the network can effectively capture local and global information, ensuring the coherent completion of the missing regions. These CNN models have been applied explicitly to smart colposcopy images to address the challenge of filling removed glare pixel. The next session discusses the CNN inpainting method used to fill the smart colposcopy images in detail.

6.1.1. Partial Convolutional Inpainting Model (P.Conv) for Inpainting Missing Pixels

Guilin introduced a digital image inpainting using a partial CNN model [91][149]. This method aims to fill in the missing pixels in cervical images by combining stacked partial convolutional operations and an updating binary mask. The U-Net architecture trains the model by replacing the conventional encoder-decoder with the partial convolutional neural network. This approach involved several key steps, including the binary mask updating, partial convolutional operation, and overall network architecture. It differs from regular convolutions by considering only the valid pixels for computation, excluding the missing or masked regions. It permits the network to concentrate only on the available information when generating the inpainted pixels, preventing the corruption of the missing areas with inconsistent values.

Step 1 Partial convolution operation

The partial convolution operation, as defined in equation (6.1)

$$x' = \begin{cases} W^T(I_{out} \odot M)\frac{Sum(1)}{Sum(M)}, & if sum(M) > 0 \\ 0, & Ot\square erwise \end{cases} \quad (6.1)$$

In the equation, W and B represent the weight and bias of the convolutional kernel, respectively. X denotes the pixel value, which refers to the feature value of the current sliding window. The M represents a binary mask, crucial in distinguishing the masked (0) and unmasked (1) regions in the cervical images. This mask aims to identify the areas where reflections or missing information exist. To compute the partial convolution, the valid pixels

in the image are considered, excluding the masked regions. It is achieved by multiplying the input pixel values (*X*) with the corresponding mask values (*M*), and then scaling the result by the sum of the unmasked pixel values to the sum of the mask values. This scaling factor ensures that the convolutional operation is adjusted based on the available valid input pixels, effectively preventing the corruption of the eliminated area when inpainting.

Step 2 Updating of the Binary Mask

After the partial convolution operation, the next step involves updating the binary mask area of the images. The binary mask is dynamically updated throughout the inpainting process. The mask helps to indicate the regions with missing pixels and guides the partial convolutions by specifying which pixels to consider during the convolution operation. By iteratively updating the binary mask, based on the current inpainted image, the network gradually refines its understanding of the missing regions and enhance the standard of the inpainting results. This update is based on a condition described in equation (6.2).

$$m' = \begin{cases} 1, & if sum(M) > 0 \\ 0, & ot\square erwise \end{cases} \qquad (6.2)$$

If the output of the convolution operation comprises of at least one valid input pixel, it is considered as the valid pixel for the subsequent partial convolution operation. This process is repeated until all the initially masked regions become valid pixels. This iterative process aims to refine the binary mask and improve the inpainting results gradually. By considering the valid pixels from the partial convolution outputs, the network better understands the missing regions and can generate more accurate inpainted pixel values. Once the binary mask has been updated, the U-Net architecture trains the cervical images and fills in the missing pixels.

Step 3 Network Architecture

The U-Net architecture is a widely used CNN architecture, particularly in biomedical imaging [129][150]. It has proven adequate for various image-related tasks, including image inpainting. The U-Net architecture is adapted for inpainting by replacing the regular U-Net convolutional layers with partial convolutional layers while retaining the encoder-decoder structure and incorporating skip links. This architecture typically consists of an encoder section that gradually reduces the spatial dimensions while capturing image features, followed by a decoder section that recovers the spatial information and generates the desired output. In the inpainting scenario, the convolutional layers of the U-Net are

replaced with partial convolutional layers, which utilize the stacked partial convolutional operations.

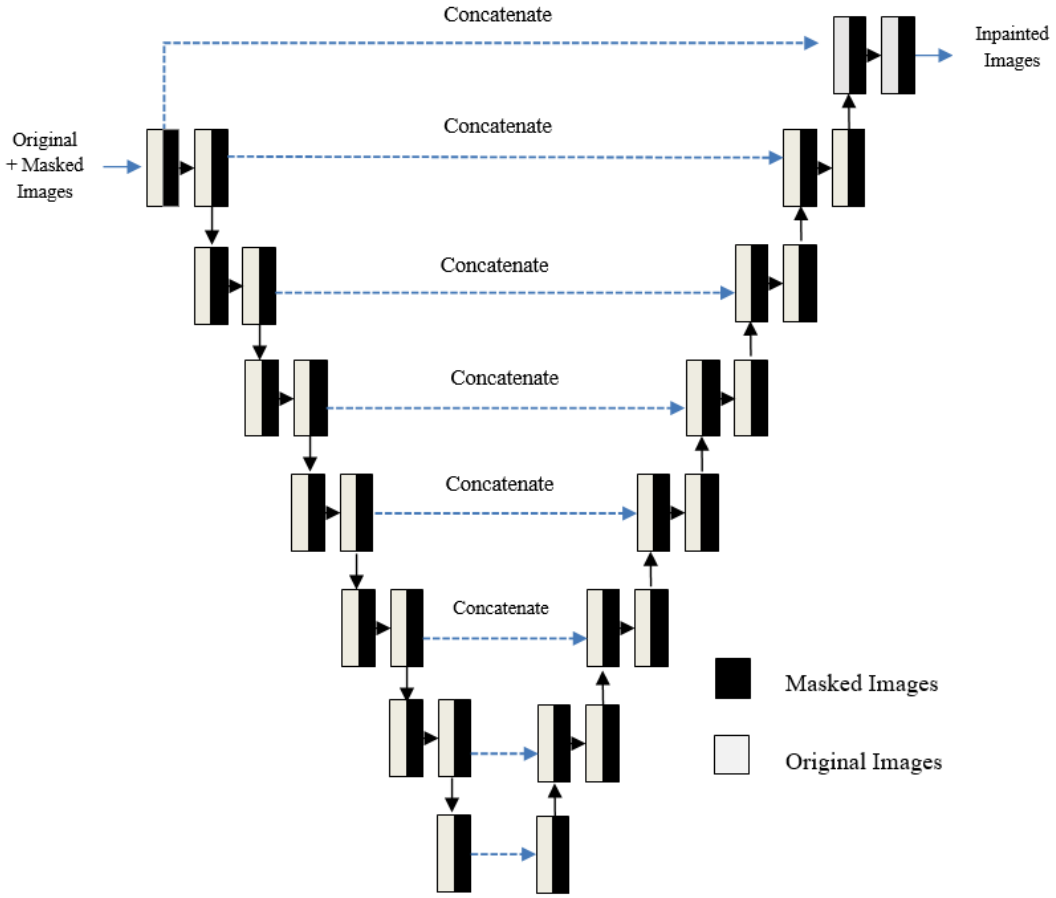


Figure 6.1 Partial Convolutional Deep Learning Model for Inpainting

Furthermore, skip links are employed in the U-Net architecture to facilitate information flow between corresponding feature maps and masks. These skip connection link the encoders feature information to the respective decoder stage, acting as a feature reference for the subsequent partial convolutional layers. It permits the network to retain important details from the encoder stage and utilize it in decoding for accurate inpainting. In the decoding step, nearest neighbor upsampling is often used to recover the spatial feature of the feature maps. The upsampling is followed by combining the respective skip link feature maps, enabling the network to combine low- and high-level data for enhanced inpainting results. The last layer of the P.Conv layer combines the cervical input image with a hole, indicating the region to be filled in the image. The layer is masked to ensure that only the missing or empty areas are inpainted using the partial convolutional operations. The architecture for completing the eliminated part on the images is shown in Figure 6.1.

Refilling the empty pixels using the partial convolutional layer can be visualized in the Figure 6.2. By incorporating the U-Net architecture with partial convolutions, skip links and masking techniques, the algorithm effectively leverages the strengths of both approaches to inpaint missing regions in cervical images, resulting in accurate and visually coherent results. However, some of the pixel gaps appear on the inpainted images, showing the loss of pixelation on the images.

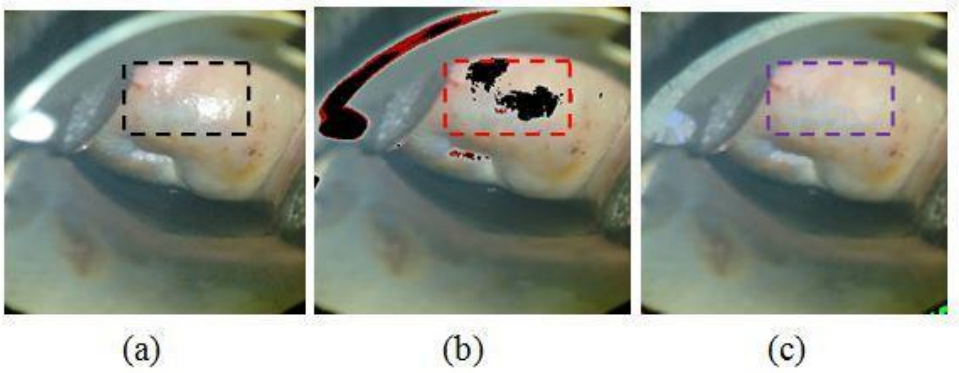


Figure 6.2. SR Inpainting using P.Conv. (a) Original images with SR are indicated utilizing the "black box". (b) Segmented SR utilizing the Finetuned U++ model, marked utilizing the "red box". (c) P.Conv inpainting fill-up the eliminated region indicated using the violet box.

6.1.2 Inpainting using Generative Multicolumn Convolutional Neural Network (GMCNN)

An approach called the GMCNN for inpainting has been introduced to fill invalid regions in digital images [151]. Their approach aimed to enhance the standard of these images by effectively refilling the missing or invalid regions. It incorporates a multi-branch CNN and a dilated CNN to achieve the desired results. The method involves several steps, which are explained below.

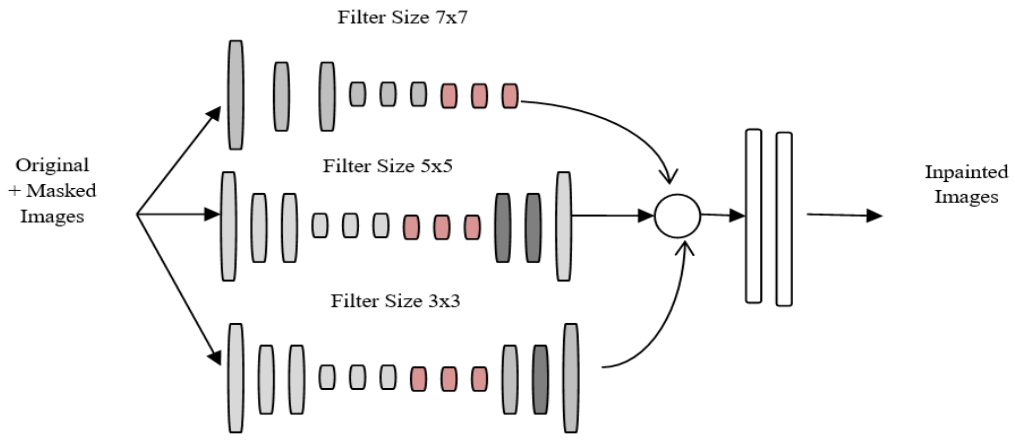


Figure 6.3 Generative Multicolumn Convolutional Model for Inpainting

Step 1 Input and Masked images

The input cervical images with the removed specular reflection region are considered *X*, and a binary mask *M* as input. The unknown pixels of the empty area of the cervical image *X* are filled with zero. After creating the input and masked images, the images are refilled using the generative network architecture.

Step 2 Network architecture

The GMCNN model for inpainting missing pixels in cervical images consists of three parallel encoder and decoder sections. Each section has a receptive field that utilizes three different kernel sizes: 3x3, 5x5, and 7x7, with dilation rates of 1, 2, and 3, respectively. These kernel sizes allow for extracting features at different resolutions of the cervical images. The three branches, each corresponding to a different kernel size, capture features at various scales to enhance the understanding of the missing regions. The features extracted from the three branches are concatenated and passed through two convolutional layers, denoted as d(.), to produce the output image *Y*. The completed image *Y* can be represented as $Y = d(F)$, where *F* represents the concatenated features from the branches. To increase the output image's resolution, upsampling is applied at layers n=1 and n=2, allowing the model to restore the original resolution of the cervical images. The generative networks from the three branches are merged at one point and further fed into the convolutional network to address the missing pixels. This GMCNN model combines multi-scale and global contextual information, leveraging the features extracted by the multi-branch and dilated convolutional networks. Figure 6.3 illustrates the overall process of the GMCNN for inpainting the eliminated pixels in the cervical images, demonstrating how the removed cervical region is refilled using this approach.

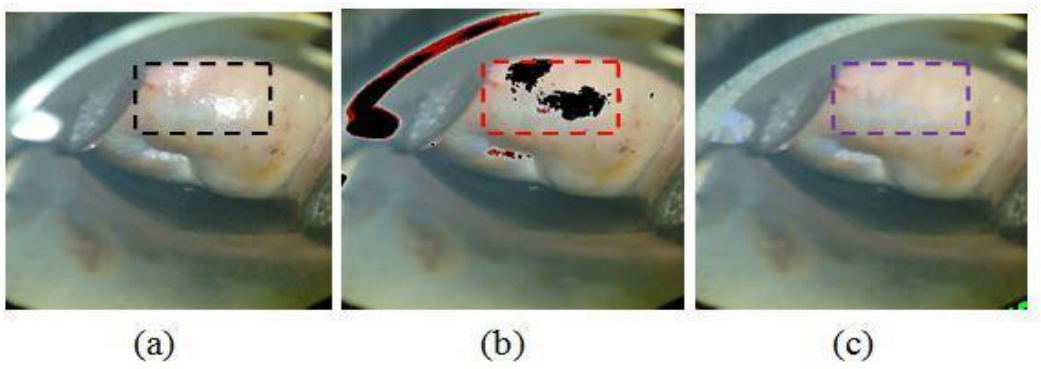


Figure 6.4 The SR inpainting using GMCNN. (a) SR is indicated using the black box (b). The red box means the detected reflection region using CNN (c). The inpainted images using GMCNN are defined utilizing the violet box.

The SR was removed to create random holes in the cervical images. Some of the cervical regions are small holes, and some regions are large holes. The refilling process is successful for the area with small holes, and it has no impact on the resolution of the cervical images, as shown in Figure 6.4(C). The large holes are also refilled, but the filled large holes appear as small patches on the cervical images. So, it will affect the quality of lesion detection, leading to the wrong analysis of the cervical images.

6.1.3 Inpainting using Dilated Convolutional Neural Network (D.Conv)

The dilated convolutional neural network was addressed for randomly shaped missing regions in cervical images [152]. This approach refills the removed reflection pixels by leveraging the neighboring pixels. First, the input images and their respective masked versions are extracted from the dataset. The masked images contain the randomly shaped missing regions where the reflection pixels have been removed. The model employs a GNN to perform the inpainting process. The network is learned using the extracted input and masked images. During the training phase, the generator learns to predict and generate the missing reflection pixels by considering the information from the surrounding pixels. The D.CNN architecture is employed in the generator to effectively capture contextual information and handle the randomly shaped missing regions. The dilated convolutions allow the network to increase the area of receptive without significantly maximizing the count of parameters. It enables the network to capture long-range dependencies and incorporate information from a broader area surrounding the missing region. The proposed approach aims to generate visually coherent and realistic results by training the GNN, filling in the eliminated reflection pixels based on the neighboring pixel information. The utilization of dilated convolutions helps the network effectively address randomly shaped missing regions in the cervical images, improving the quality and completeness of the inpainted results.

Step 1 Input and Masked Images

The specular reflections in these images were identified and converted into a binary mask, representing the original images with missing regions. The GNN was trained to address these missing regions using a pair of input images: the original cervical image and the corresponding binary-masked image. This pair of images is considered corrupted since the binary mask represents the missing pixels. By training the generator network on these

corrupted images, the model learns to paint and refill the missing pixels based on the information currently in the original image. The generator utilizes the contextual information from the surrounding pixels to generate coherent and realistic results, filling in the missing regions of the cervical images.

Step 2 Network Architecture

In this algorithm, the Generator Inpainting (GI) network is utilized to load the empty area of the cervical images. The GI network is a FCNN that inputs corrupted cervical images. An encoder-decoder architecture is adopted to optimize memory usage and computational time. During the inpainting process, the resolution of the resulting images may be affected due to the refilling operation on the empty regions. However, maintaining the resolution of the images is vital for accurate dysplasia analysis. To address this, a dilated convolutional neural network is employed. Dilated convolutions are a modified version of standard convolutions that enable faster training with fewer parameters. By utilizing dilated convolutions, the algorithm retains the resolution of the cervical images while effectively refilling the removed specular reflections. The training images are resized to a size of 255x255 pixels. The encoder region of the GI network employs the ReLU activation function, while the decoder region employs the leaky ReLU activation function with a parameter of $\alpha = 0.2$.

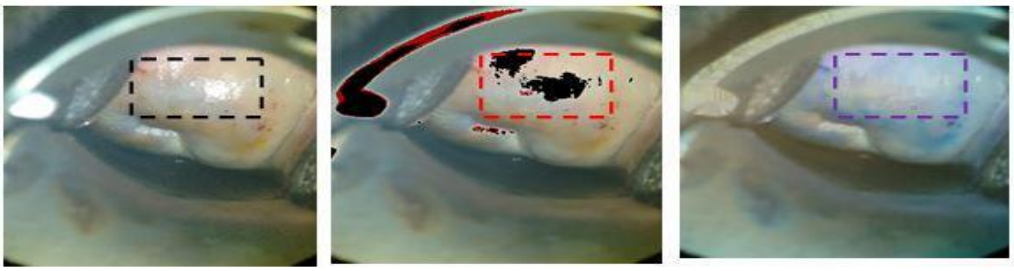


Figure 6.5 SR inpainting using a DCNN. (a) Original cervix images with the black box representing the specular reflection(b) Red box means the detected SR (c) The refilled pixel using a DCNN.

These activation function helps to introduce non-linearity and enable better representation learning in the network. Through the learning process, the GI network learns to paint the removed specular reflections on the cervical images using dilated convolutional neural networks. The resulting images preserve the resolution necessary for accurate dysplasia analysis. The result obtained using the D.Conv helps to refill the holes in the cervical images. The drawback is that the color of some cervix images is affected and

appears as some blue region on the images, as illustrated in Figure 6.5(C). The hue of the cervix images is more critical for neoplasm detection and grading. In the refilling process, the small holes of the cervical images are filled, but the large holes in the refilled regions appear as patches on the cervical images.

6.2. Overview of the Proposed Method

On analysis, the inpainting model significantly removes noise in smart colposcopy images. This chapter proposed the bilateral based convolutional filter for identifying empty regions on smart colposcopy images and filling the area with the neighboring pixel. It removes the glare region, which helps to enhance the standard of the images. The model is trained by fitting the proposed filter in the deep learning U-Net model, which introduced a new model called a bilateral-based convolutional inpainting model to fill the region with the neighboring pixel on smart colposcopy images. The model is contrasted with various deep learning convolutional models to fill the missing area on the images. This model is analyzed with other loss functions to identify the suitable loss function for the proposed model. On analysis, the proposed model outperformed in both qualitative and quantitative analysis. Figure 6.6 provides an overview of the suggested procedure.

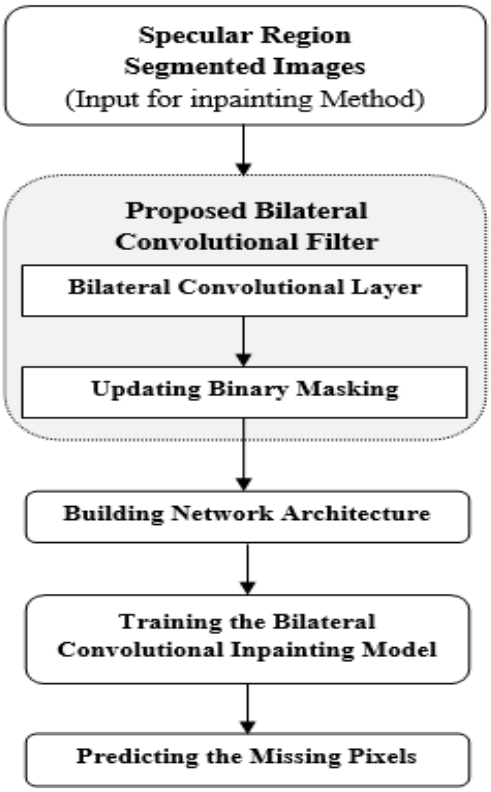


Figure.6.6 Outline of the Proposed Method for the Identification of SR

6.2.1 Inpainting using Filters

The image completion problem is typically formulated in equation 6.3, which includes the incomplete image (x), the complete image (y), and the noise (n). The equation considers the image index (i) and the ideals of other corrupted pixels (u) caused by various factors.

$$x(i) = \begin{cases} y(i) + n(i), & M(i) = 1 \\ u(i), & M(i) = 0 \end{cases} \quad (6.3)$$

A binary mask (M) is utilized, where 0 represents missing pixels, and 1 illustrates valid pixels in the digital image [148]. The conventional inpainting algorithms employ statistical and mathematical approximation methods such as biharmonic functions, Cahn-Hilliard equations, and anisotropic diffusion based on partial differential equations. These methods propagate boundary data to estimate missing pixel values. Additionally, filtering techniques like convolutional, bilateral, and median filters are often used iteratively to predict the valid pixel regions [153]. Several filters are commonly used in image inpainting to enhance the standard of the inpainted parts. On study, the bilateral and convolutional filters are highly used for the inpainting of digital images.

6.2.1.1 Convolutional Filter for Inpainting the Digital Images.

Convolutional filters, known as kernel filters, are widely used in image processing. It involves convolving a small matrix over the image to perform blurring, sharpening, or edge detection operations. The convolutional filters help to smooth out the inpainted regions and improve their visual coherence with the surrounding areas. The convolutional filter for inpainting the digital images is formulated in equation 6.4.

$$Output(x, y) = sum\big(kernel(i, j) * Input(x + i, y + j)\big) \quad (6.4)$$

The $Output(x, y)$ represents the output pixel value at coordinates (x, y) in the filtered image. The $kernel(i, j)$ refers to the kernel value at position (i, j) within the filter matrix. *The* $Input(x + i, y + j)$ denotes the original image's input pixel value at coordinates $(x + i, y + j)$. It computes the weighted sum of the products between the kernel values and the corresponding input pixels within a local neighborhood around each output pixel position. This process is iterated for every pixel in the image, giving the filtered output images [154][155]. The size and values of the kernel matrix determine the specific filtering

operation performed. The commonly used convolutional filters includes gaussian filters, edge detection filters, and box blurring filters. The selection of the kernel matrix depends on the desired effect and the characteristics of the inpainting task.

6.2.1.2 Bilateral Filter for Inpainting the Digital Images

The bilateral filter considers both spatial proximity and pixel intensity similarity. It preserves the edges while reducing noise in the image. This filter is commonly used for filling the eliminated area with the original image's structure and texture. The bilateral filter is formulated in the equation. 6.5.

$$Output(x,y) = \frac{1}{W(x,y)} Sum\big[input(i,j) * SpatialWeig\square t(x,y,i,j) * IntensityWeig\square t\big(Input(x,y), Input(i,j)\big)\big] \quad (6.5)$$

The $Output(x,y)$ represents the output pixel value at coordinates (x,y) in the filtered image. The $input(i,j)$ denotes the original image's input pixel value at coordinates (i,j). The (x,y,i,j) is the spatial weight, which measures the spatial proximity between the center pixel (x,y) and the neighboring pixel (i,j). It is calculated using a gaussian function using the Euclidean distance between the two-pixel positions [132][156]. The $IntensityWeight\big(Input(x,y), Input(i,j)\big)$ measures the similarity between the intensity values of the center pixel (x,y) and the neighboring pixel (i,j). The $W(x,y)$ represents the normalization factor, ensuring the weights sum is correctly normalized. The bilateral filter applies the spatial and intensity weights to each neighboring pixel and its corresponding intensity value, then performs a weighted average of the adjacent pixel values to compute the output pixel value. The weights are determined based on the spatial proximity and intensity similarity between the center and neighboring pixels [156].

The bilateral filter considers the distance between pixels in the image and the difference in their intensity values. It allows the filter to preserve edges and fine details while effectively smoothing out noise in the image. In the context of inpainting, the bilateral filter can help blend the inpainted regions with the surrounding areas, ensuring a visually coherent and realistic result. The bilateral filter is a popular technique used in image processing and, specifically, in the context of image inpainting. It effectively removes noise while preserving edges and details, making it suitable for enhancing the quality of inpainted regions. The bilateral filter considers spatial proximity and pixel intensity similarity when

applying the filter. Both plays are highly used filters for inpainting digital images based on the comparison analysis of the convolution and bilateral filters. The bilateral filter outperforms the convolutional filter in filling the missing portion of the images [157]. So, to pinpoint the eliminated region on the images, the bilateral-based method is proposed to improve the texture of the filled pixels.

6.2.2. Proposed Bilateral Based Convolutional Model (B.Conv)for Inpainting

This session proposed a novel approach combining traditional convolutional layers power with bilateral filtering's adaptability. As defined by equation (6.6), the standard convolutional layer has been widely used in various computer vision tasks to inpaint the missing portion of the images. Based on the comparison analysis of the bilateral and convolutional layers, the convolution filter fails to consider the spatial structure of the input images. So, the traditional model does not explicitly consider the spatial arrangement or the range differences in the input data, potentially leading to a loss of essential details and texture in the predicted images.

$$I_{out}(i) = I_{in}(i) \otimes W(i,j) \tag{6.6}$$

The i and j indicate the pixel value of the digital images $\otimes$ and denote the convolutional operator, I is the kernel filter, *and* I_{in} and I_{out} indicate the convolution layer's input and output images. The equation (6.6) is modified by explicitly taking the average weight of the pixel into account, as shown in equation (6.7). To consider the spatial information, the bilateral filter is incorporated to modulate the weight of the convolution layer, and equation (6.6) is modified and represented in equation (6.7). The bilateral filter kernel uses the element-wise multiplication between the original weight $W(i,j)$ and gaussian function of the differences in depth or range between the pixel z_i and z_j. The gaussian function is represented as σ, which controls the width of the function.

$$I_{out} = I_{in} \otimes \left(W(i,j) \odot \frac{\exp\left(-(z_i - z_j)^2\right)}{\sigma^2} \right) \tag{6.7}$$

In this equation, I_{in} represents the input image, and I_{out} represents the output image after the convolution operation. The symbol $\otimes$ denotes the convolution operator. The term $\left(W(i,j) \odot \frac{\exp\left(-(z_i - z_j)^2\right)}{\sigma^2} \right)$ represents the bilateral filter kernel, which modulates the weights

of the convolutional layer. The symbol ⊙ represents element-wise multiplication. The exponential term $\frac{\exp\left(-(z_i-z_j)^2\right)}{\sigma^2}$ is a gaussian function that takes into account differences in the depth or range of the input data, and σ is a parameter that governs the width of the gaussian function.

Using a bilateral filter kernel to modulate the convolutional layer weights, the model can adapt to the spatial structure of the input data while preserving edges and other features. It mainly helps in preserving the details and texture of the predicted images. In this technique, the weight of the CNN kernel using a gaussian filter is modified by considering the depth differences. By applying a gaussian filter that considers differences in depth, the kernel can adapt to the spatial arrangement of the input data. This approach is identical to the joint bilateral filter, a non-linear filter used in image processing to smooth images while preserving edges. It also smoothens the color differences of the image, especially in the complex structure, and keeps the edges. It modulates the kernel weights by regularization, which prevents the model from over fitting by adding constraints to the weights. It improves the model's ability to generalize the data and overall performance in predicting the missing pixels. The weight of the CNN kernel determines the processing of the input data, where a large weight gives more emphasis to certain features and a small weight gives less emphasis. The bilateral convolution layer is further modified to make the model additionally predict the image missing pixels. The proposed layer is fitted like the partial convolutional layer, as shown in equation 6.1. The partial convolutional layer includes the partial convolutional operation and mask update jointly functioned to predict the missing pixel. Similarly, the partial convolutional operation is replaced with the bilateral convolutional process along with the mask updates to predict the missing pixel on the image, as shown in the equation. 6.8.

$$x' = \begin{cases} W^T(I_{out} \odot M)\dfrac{Sum(1)}{Sum(M)}, & if sum(M) > 0 \\ 0, & Otherwise \end{cases} \qquad (6.8)$$

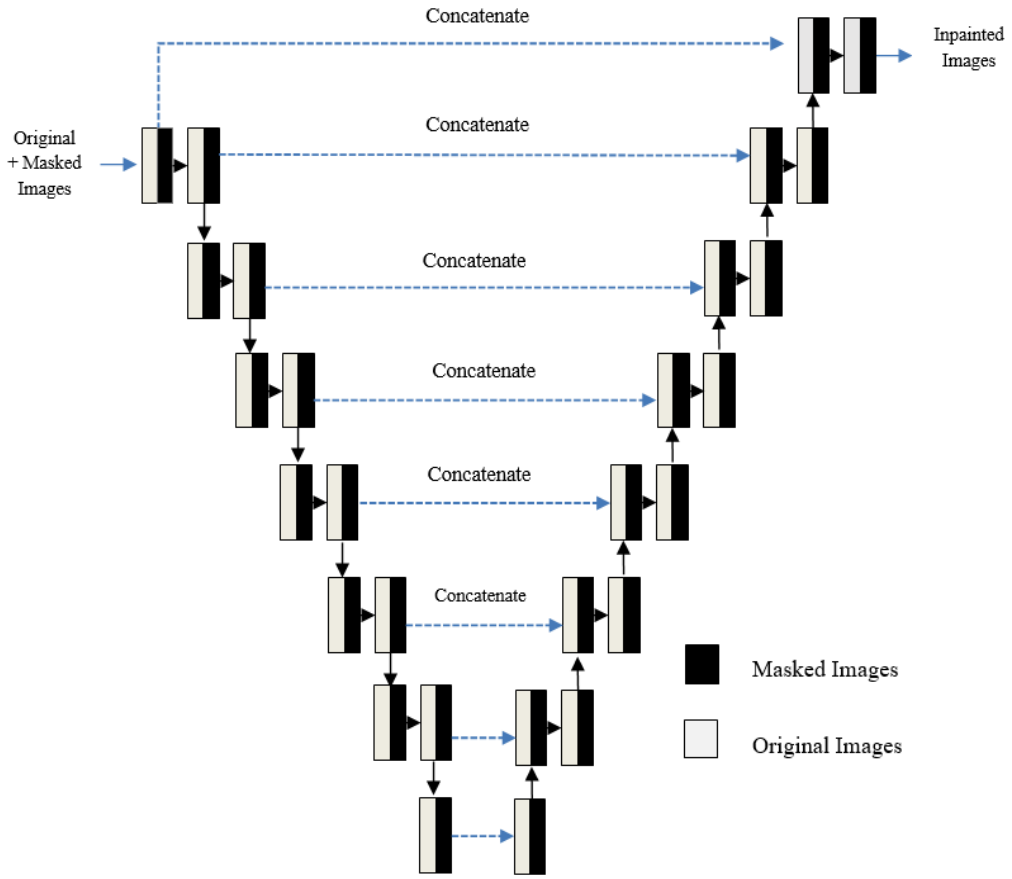
The symbol ⊙ represents element-wise multiplication, σ represents the standard deviation. By default, it is set as 1. The z_i and z_j are the averaging weights of the pixels i and j of the digital images. In order to forecast the absent pixels of the digital, the bilateral operation equation (6.7) is element-wise multiplied with the masked image *M*, as shown in equation (6.8). The M is the binary masked image which indicates the validity of each

feature value, i.e., 0 indicate the eliminated pixels, and 1 illustrates the rational pixels of the digital images. The binary mask checks the missing region and fills the non-filled region as shown in equation 6.9. The associated position will be considered legitimate for the following bilateral convolutional layer if the current convolution's result is dependent on at least one valid input value.

$$m' = \begin{cases} 1, & if sum(M) > 0 \\ 0, & otherwise \end{cases} \qquad (6.9)$$

6.2.3. Training the Proposed Model for Filling the Eliminated Region

The U-Net architecture is widely used in CNN architecture, especially in biomedical imaging. The U-Net has shown its effectiveness in various image-related tasks, including image inpainting [105] [158]. It is modified for inpainting by replacing regular convolutional layers with partial ones while maintaining its encoder-decoder structure and incorporating skip connections. The U-Net architecture typically consists of two main sections: the encoder and the decoder.



6.7 Proposed Bilateral Convolutional Inpainting Model for Inpainting the Smart Colposcopy Images

The encoder captures important aspects of the input image while progressively decreasing its spatial dimensions. On the other hand, the decoder aims to recover the spatial information and generate the desired output. Similar to the partial inpainting model, the traditional convolutional layers of the U-Net are replaced with bilateral convolutional layers. It is designed to handle masked regions by using only the valid parts of the convolution filter that fall within the known (unmasked) regions. The network focuses on inpainting the eliminated pieces of the image based on the information available.

Additionally, skip connections are introduced in the U-Net architecture to facilitate information flow between corresponding feature maps and masks. The U-Net retains essential information from the encoder stage and utilizes it in decoding, for accurate inpainting. The combination of bilateral convolutional layers and skip connections in the U-Net architecture allows the network to effectively inpaint missing regions in the image, providing high-quality results in various inpainting scenarios, especially in biomedical imaging tasks. In the decoding stage, nearest neighbor upsampling is often used to recover the spatial dimensions of the feature maps. This upsampling is followed by combining with the respective skip link feature maps, enabling the network to combine low- and high-level features for enhanced inpainting results. The last layer of the bilateral convolutional layer combines the cervical input image with a hole, indicating the region to be filled in the image. This layer is masked to ensure that only the missing or empty regions are inpainted using the bilateral convolutional operations. Refilling the empty pixels using the bilateral convolutional layer can be visualized in Figure 6.8. By incorporating the U-Net architecture with bilateral convolutions, skip links and masking techniques, the algorithm effectively leverages the strengths of both approaches to inpaint missing regions in cervical images, resulting in accurate and visually coherent results.

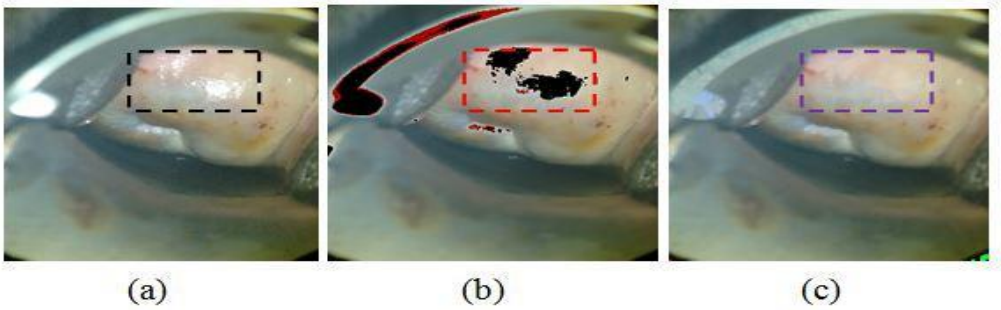


Figure 6.8 SR Inpainting using BConv. (a) SR is indicated using the black box (b). The red box represents the detected reflection region using Finetuned UNet++ (c). The inpainted images using B.Conv are represented using the violet box.

6.3. Result and Discussion

The experimental setting required for the proposed model is discussed in this section. The deep learning models were trained on a machine provision with a "*Tesla V100 PCIe GPU and CUDA version 11.4*", which provided the computational power necessary to train the models efficiently.

6.3.1 Training the Inpainting Models

This session briefly discusses the training data, optimizer model input batch size and learning rate to train the inpainting model.

- **Training Data**: The model is trained with smart colposcopy images in which the specular reflection is detected and removed from the images. The training dataset has 3582 images, the testing dataset has 448 images, and the validation dataset has 448 images.
- **Optimizer**: The Adam optimizer is utilized for learning the model [91]. The extension of stochastic gradient descent combines the advantage of adaptive learning rates and momentums. It automatically adapts the learning rate for each parameter during training. It calculates individual rates of learning according to historical gradients of each parameter, allowing it to converge faster and more effectively. It also incorporates the concept of momentum, which helps the optimizer accelerate the learning process by accumulating the influence of past gradients. This helps to navigate the flatter regions of the loss landscape and avoid getting stuck in local minima.
- **Model inputs**: The deep learning inpainting model takes the original images and their corresponding masked images (i.e., a region where the specular reflection region is removed from the surface).
- **Batch size and learning rate**: The model is learned with 32 as batch size, which means that the optimizer updates the model's parameters after processing 32 images at a time. The 0.001 is set as the learning rate, which controls the step size taken during parameter updates.
- **Loss function and activation function**: The loss function used is BCE, which compares the predicted output with the ground truth labels. The activation function used in the final layer of the model is sigmoid, which maps the model's output to a probability range between 0 and 1.

- **Training epochs**: The model is trained for 200 epochs, where the entire training dataset is passed through the model 200 times during the training process.

6.3.2. Comparison Analysis of the Inpainting Model for the Removal of SR

The comparison analysis of the model is carried out in both qualitative and quantitative analysis**.** In qualitative assessment, the SR removed region is inpainted employed the bilateral convolutional inpainting model and is visualized and inspected in this session. The specular reflection was deliberately removed from cervical images to create random holes, as shown in Figure 6.9(a). These holes vary in size, some being small while others are larger. The GMCNN refilling process was applied to address this issue, which successfully filled the small holes without affecting the color or resolution of the images, as shown in Figure 6.9(b). However, when it comes to larger holes, the GMCNN refilling process affects the texture of the digital images, which could lead to inaccurate lesion detection and wrong analysis of the cervical images. To mitigate this problem, the D.Conv algorithm was used to refill the holes in the cervical images. While it helped in filling the holes, it caused some of the cervix images to have blue regions, impacting the hue quality of the images, which is vital for neoplasm detection, as shown in Figure 6.9(c). Another method called P.Conv was employed for hole refilling, and it also showed promising results for small hole filling without compromising the image quality. However, some small pixels were left unfilled for larger holes, resulting in white dots on the images and affecting the image quality, as shown in Figure 6.9(d). The proposed BConv model was used for hole refilling and demonstrated positive outcomes. It managed to fill large holes effectively without compromising pixel or color quality, keeping the images similar to the original ones, as shown in Figure 6.9(e).

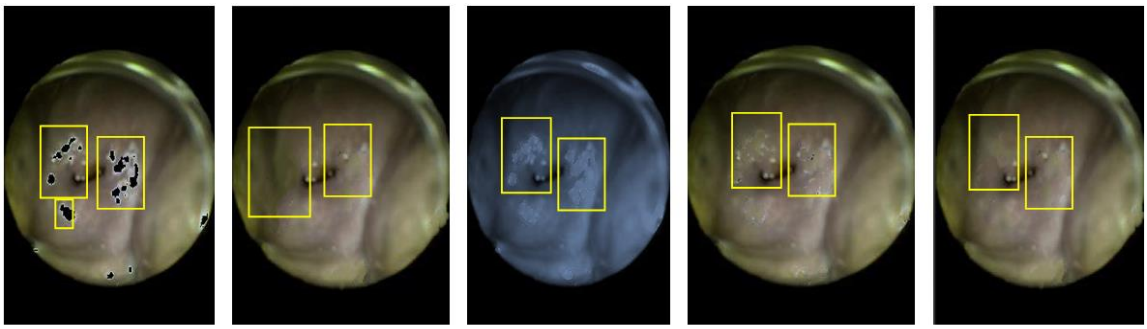
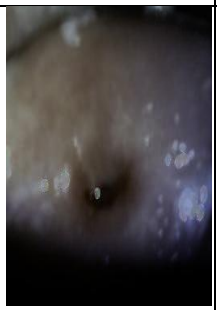
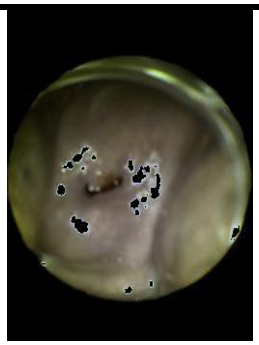
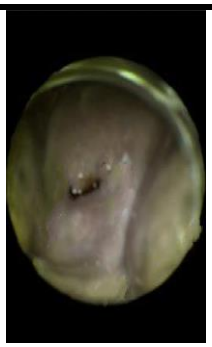
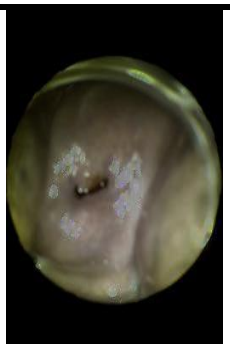
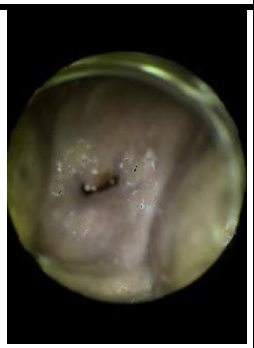
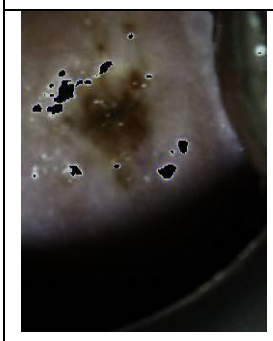


Figure 6.9 Comparison Evaluation (a) Original Images with the removed reflection region (b) Empty region refilled using GMCNN model (C) Empty region refilled using D.Conv model (d) Empty region refilled using P.Conv model (d) Empty region refilled using proposed B.Conv model

In summary, the proposed B.Conv model is the most promising for effectively filling large holes without adversely affecting image quality, particularly the color and texture, which are crucial for accurate diagnosis, especially in neoplasm detection. The comparison of the results is shown in Table 6.1.

Table6.1 Comparison Analysis for the Removal of SR on Smart Colposcopy Images

	SR Segmented Images	**GMCNN**	**D.Conv**	**P.Conv**	**Proposed B.Conv**
Type 1					
					
Type 2					
					

Type 3					
					

For quantitative analysis, various metrics like the mean of squared error, Peak Signal-to-Noise Ratio, Structural Similarity Index, L1-Loss, Coefficient of Variation, and Sum of Squared Differences are utilized.

A. **Mean of Squared difference (MSD)**

The target patch and the exemplar patch's MSD can be used to calculate the degree of differences between corresponding pixels at known positions (i.e., pixels that are already present) in both patches [159]. The mismatch mistake is likely to occur if there are noticeable variations between the current pixels in the target patch and the equivalent pixels in the exemplar patch. The MSD calculates how much the two patches differ from one another. The equation (6.10.) represents the MSD.

$$MSD(\varphi_p, \varphi_q) = \frac{\sum |M\varphi_p - M\varphi_q|^2}{\sum M} \qquad (6.10)$$

Where $\varphi_p$the target is the patch, and φ_q is the exemplar patch. M is the binary mask, which uses 1 to indicate the pixels that need to be filled and 0 to indicate the already existing pixels. The $M\varphi_p$ extracts the pixels that already exist in the target patch φ_p , $and\ M\varphi_q$ extracts corresponding pixels in the exemplar patch φ_q . To determine how different two patches are from one another, it computes the average of the squared disparities between corresponding pixels at known places in each patch. According to equation (6.10), if the MSD value is low, it indicates that the pixels in the two patches are

fairly comparable already. Conversely, if MSD is substantial, it indicates that the two patches' preexisting pixel compositions differ significantly.

B. Peak Signal-to-Noise Ratio (PSNR)

It is a widely used metric to evaluate the quality of image inpainting [160]. It determines how similar the inpainted and original image are to one another by quantifying the noise or distortion introduced during the inpainting process. Better inpainting performance is indicated by higher PSNR values. The PSNR is represented in the equation. 6.11.

$$PSNR = 10.log_{10}\left(\frac{MAX}{MSE}\right)^2 \qquad (6.11)$$

Where:

- MAX indicate the maximum possible pixel value of the image (e.g., 255 for 8-bit images).
- MSE is the Mean Squared Error, calculated as the average of the squared differences between corresponding pixel values in the original and inpainted images.

It provides a numerical measure of the inpainting quality, reflecting the information preserved during the process.

C. Structural Similarity Index

It is a widely used metric for evaluating the quality of image inpainting, particularly when assessing the similarity in perception between the original and inpainted images [160]. It measures two images structural and textural similarity, considering luminance, contrast, and structural information. Its index varies from -1 to 1, 1 indicating perfect image similarity. A closer resemblance between the inpainted and original images is indicated by higher SSIM values, which indicate superior inpainting ability. The formula to calculate SSIM is represented in the equation.6.12.

$$SSIM(x,y) = \frac{(2\mu X.\mu Y+C_1)(2\sigma XY+C_2)}{(\mu_X^2+\mu_X^2+C_1)(\sigma_X^2+\sigma_X^2+C_2)} \qquad (6.12)$$

Where:

- X and Y are the original and inpainted images, respectively.
- μ_X and μ_Y are the average pixel intensities of X and Y.
- σ_X and σ_Y are the standard deviations of X and Y pixel intensities.
- σ_{XY} is the cross-covariance between X and Y.
- c_1 and c_2 are constants to stabilize the division and avoid division by zero.

The SSIM can quantitatively assess how well your inpainting method preserves the visual content and texture of the original image.

D. L1 Loss

L1 Loss, or Mean Absolute Error (MAE), is a standard metric in image inpainting and other image processing tasks [161]. L1 Loss measures the average absolute differences between respective pixel values in the original and inpainted images. It quantifies the dissimilarity between the images at a pixel level. The L1 Loss is represented in the equation.6.13.

$$L1\ Loss = \frac{1}{N}\sum_{i=1}^{N}\left|X_i - X_y\right| \tag{6.13}$$

Where:

- X_i represents the pixel of the original image.
- Y_i represents the pixel of the inpainted image.
- N is the pixel count in the inpainted regions (defined by the binary mask).

It measures the average absolute discrepancy between corresponding pixels, indicating how effectively the inpainted image appears identical to the original image in the inpainted regions. Lower L1 Loss values suggest better inpainting performance, as it shows a closer resemblance between the inpainted and original images at a pixel level.

E. Coefficient of Variation (CoV)

It is a statistical measure used to assess a datasets relative variability or dispersion. It quantifies the ratio of the standard deviation to the mean of a dataset and is often expressed as a percentage. COV is commonly used in various fields to compare the variability of different datasets, mainly when their scales differ. The COV is represented in the equation.6.14.

$$COV = \left(\frac{StandardDeviation}{Mean}\right) * 100 \tag{6.14}$$

Where:

- ***Standard Deviation*** measures the spread or dispersion of the dataset values around the mean.
- ***Mean*** is the average of the dataset.

In the context of image inpainting, COV may not be a commonly used metric for evaluation [68]. It can be applied to the pixel intensities of inpainted regions to quantify their variability relative to the mean intensity, it may not provide direct insights into the quality of inpainting or the preservation of image content.

F. Sum of Squared Differences (SSD)

It is a metric used to assess the similarity between two images by calculating the sum of the squared differences between corresponding pixel values [159]. In image inpainting, it can evaluate how effectively the inpainted image identified similar to the original image in the regions where inpainting was performed. To calculate the SSD for inpainting is represented in the equation.6.15.

$$SSD = \sum_{i=1}^{N} (X_i - Y_i)^2 \quad (6.15)$$

Where:

- X_i represents the pixel of the original image.
- Y_i represents the pixel of the inpainted image.
- N is the pixel count in the inpainted regions (defined by the binary mask).

Lower SSD values indicate better inpainting performance, meaning the inpainted image is identified similar to the original image in the inpainted regions. SSD is a pixel-wise metric that assesses the differences between individual pixel values. It provides valuable information about the pixel-level similarity of the images.

Table 6.2 Comparison Analysis of the Deep Learning Inpainting Model for Inpainting the Smart Colposcopy Images

Metrics	P.Conv	GMCNN	D.Conv	Proposed B.Conv
SSD ↓	0.134	0.147	0.2387	**0.017**
MSD ↓	0.964	0.967	0.896	**0.978**
COV ↓	0.193	0.173	0.147	**0.112**
PSNR ↑	48.25	47.32	47.48	**48.95**
SSIM ↑	0.984	0.984	0.984	**0.999**
L1 Loss ↓	0.627	0.632	0.851	**0.536**

From the Table 6.2, it is observed that the proposed “B.Conv" model outperforms the other models (P. Conv, GMCNN, and D. Conv) across most of the evaluation metrics. It achieves the lowest values in SSD, COV, and L1 Loss, indicating better image similarity, overlaps, and lower pixel-level differences with the ground truth. Additionally, it has the highest PSNR and SSIM values, signifying better image quality and structural similarity than the other models. The performance of each model can vary based on the specific dataset and the nature of the images being processed. Based on this quantitative evaluation, the "*Proposed B.Conv"* model shows promising results for filling the removed reflection region and appears to be the most effective among the models considered in this comparison.

6.3.3. Implementation of the Proposed B.Conv Model on the Different Dataset

The proposed model is implemented to the other digital and medical images for further analysis to check the quality of the inpainting process performed on the digital images. The datasets used are the three medical images and three digital images. The glare region identified in the medical dataset, as discussed in Chapter 3, is considered for the analysis. The SR region, recognized and segmented on the medical images, fills the eliminated areas. The dataset description utilized for inpainting the digital images are discussed below:

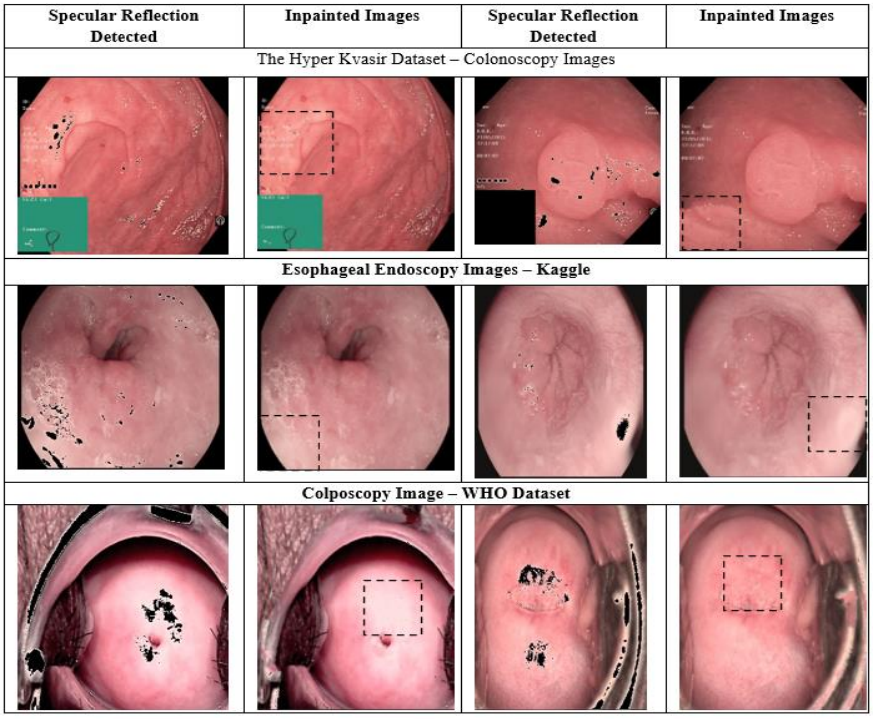


Figure 6.10 Specular Reflection Removed on Medical Images using the Proposed B.Conv Model

- **The Hyper Kvasir Dataset – Colonoscopy Images [139]**

The data is collected during real gastro- and colonoscopy examinations at a Hospital in Norway and partly labeled by experienced gastrointestinal endoscopists. For training the data, 4000 images are selected, and 1000 images are used for testing the model.

- **Esophageal Endoscopy Images – Kaggle [140]**

The dataset is collected from the Kaggle database. A total of 1000 photos are chosen for training, and 689 images are used for model testing.

- **Colposcopy Image – WHO Dataset[141]**

The images are acquired from the WHO dataset. The dataset consists of 1420 images, where 1000 images are used for training, and 420 images are used for testing the model.

The proposed method is specifically designed to address the issue of specular reflections commonly present in digital medical images captured through endoscopy, digital colposcopy, and colonoscopy procedures. These reflections can obstruct essential details in the images and hinder accurate diagnosis and analysis. The proposed method can effectively pinpoint and isolate these reflections from the rest of the image content by leveraging advanced image processing and segmentations models. The proposed B,Conv model is utilized to achieve specular reflection removal. Inpainting is filling in missing or removed regions in an image, and bilateral inpainting is a variant that considers both spatial and intensity information, allowing for smoother and more seamless inpainting results. The inpainting process successfully eliminates the identified specular reflections from the medical images, as shown in Figure 6.10.

Moreover, the inpainting method achieves this without compromising the medical images essential texture appearance and color information. The qualitative assessment of the inpainted images confirms the productivity of the proposed model in SR elimination. It demonstrates that the reflections are removed, leaving behind clear, unobstructed medical images that retain their original visual features. This improvement in image quality is expected to enhance the accuracy and reliability of subsequent medical image analysis, assisting healthcare professionals in making well-informed decisions and diagnoses. Similarly, the proposed bilateral inpainting model is utilized to the other images to check its quality performance on the different digital images. The Place2 dataset, OMUL-OpenLogo dataset, and cloud satellite images are considered for this analysis.

- **Places 2,** it has more than 1.8 million photos from 365 different types of scenes. Because of its intricate scenes, it is one of the hardest datasets to inpaint images. The training/testing splits (1.8 million/36,500) that correspond to the parameters most inpainting models utilize [162].
- **QMUL-Open Logo:** It contains 27,083 images from 352 logo classes. The fine-grained bounding box annotations of logos annotates each image. In this proposed model, 15,975 training images for training and 2,777 validation images for testing [163].
- **Satellite Images - Kaggle Dataset** contains 9,244 satellite images collected from Kaggle. The 5546 are employed for training, and 3698 are employed for testing the model [164].

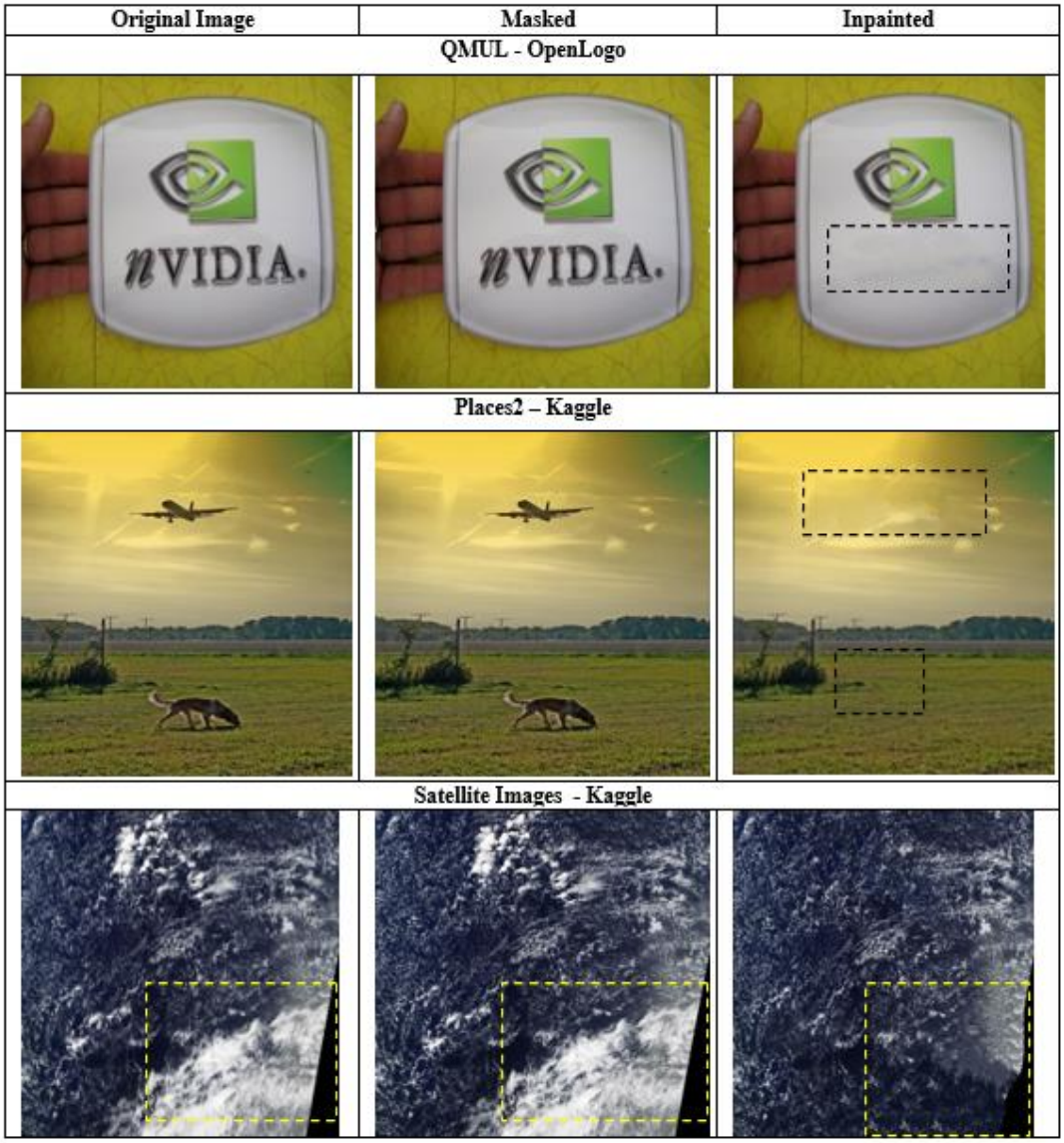


Figure 6.11 Implementation of Proposed Bilateral Convolutional Inpainting Model on the Other Digital Images

In the QMUL-OpenLogo dataset, the NVIDIA logo is selected and completely removed from digital images (Figure 6.11, first row). The logo is seamlessly refilled with a white background, preserving the original texture of the images. Similarly, on the Place 2 Dataset, scenario category images are chosen, and the dog and airplane are masked and removed using our proposed method (Figure 6.11, second row). The image color quality is fully restored, and no visible pixel changes are observed in the images after inpainting. Moreover, in the case of satellite images, cloud regions are identified and masked in the images (Figure 6.11, third row). The proposed method successfully removes the cloud regions and restores the empty regions without any noticeable impact on the image texture. The restoration is performed effectively even in regions where large portions of the images were covered by clouds (highlighted in yellow). The analysis demonstrates that the proposed method achieves remarkable results across different datasets, effectively inpainting and restoring missing content while preserving the visual quality and texture of the original images.

6.3.4 Implementation of Bilateral Convolution Inpainting Model on the Different Masking Ratio

Since the unstructured masks are more difficult and resemble real-world applications, they are utilized for training and testing. The free form masking is majorly adopted by inpainting approaches [165]. All images are reduced to 255x255 for training and evaluation to reduce the computational time during the inpainting process. The Place 2 dataset is used for this analysis**.** The existing work square or rectangle region is masked from the ground truth images and then applied for the inpainting. For this model, the area is masked with the holes of arbitrary shapes on the ground truth images. Two sets of masked photos are utilized: 20,000 for testing and 55,000 for training. The masked images are trained and tested with a size of 255x255. The images are masked with different ratios, like 0-10%, 10-20%, 20-30%, 30-40%, and 40-50%, as shown in Figure 6.12. The model used the Adam optimizer with the learning rate of 0.00005**.** The proposed B.Conv method and P.Conv exhibit promising performance when dealing with a 10-20% masking ratio. In this range, both methods effectively remove the specified region from the images and replace it with background information, all while preserving the overall image quality. This ability to pinpoint missing parts without noticeable quality degradation is essential for maintaining the integrity and usefulness of the medical images. However, as the masking ratio increases to

20-30%, 30-40%, and 40-50%, the effectiveness of the methods starts to differ—the "*Proposed B.Conv*" method continues to demonstrate satisfactory results in filling the masked region despite its increasing size. It can handle larger masked areas while maintaining the coherence of the background and ensuring the appearance of a seamless inpainting process.

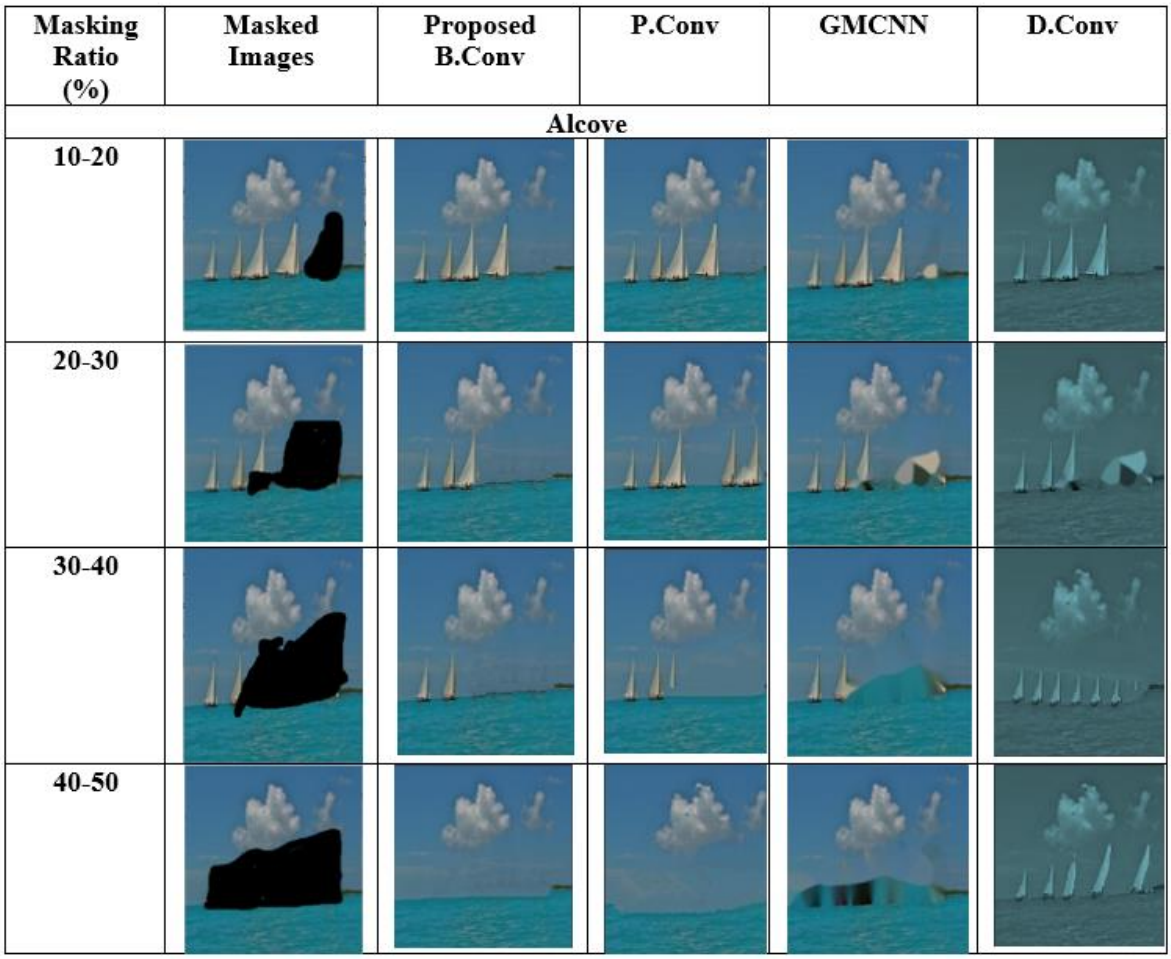


Figure 6.12 Qualitative Analysis of the Bilateral Convolutional Inpainting Model on Different Masking Ratio

The P.Conv method faces some limitations as the masking ratio becomes larger. When the masked region becomes more substantial, the P.Conv method struggles with accurately filling the area, resulting in less realistic inpainted images. The quality of the inpainting becomes compromised, and the filled part might appear less natural or contain noticeable artifacts. The quantitative analysis of the proposed method is represented in Table 6.3 with the different masking ratios. In comparison, the GMCNN and D.Conv methods face challenges in effectively handling larger masking ratios. As the masked area increases, these methods struggle to remove the region entirely from the images. Residual

traces of the masked region might remain, and the inpainting process may appear incomplete or flawed, affecting the overall image quality. Additionally, for the GMCNN and D.Conv methods, the texture of the inpainted region seems different from the surrounding image content. This difference in texture is visually apparent and might raise concerns about the accuracy and reliability of the inpainting process, particularly in medical image analysis, where preserving texture details are critical for accurate diagnosis. In summary, the "*BConv*" method and P.Conv perform well in inpainting smaller regions (10-20% masking ratio) with minimal quality degradation.

Table 6.3 Quantitative Analysis of the Bilateral Convolutional Inpainting Model on Different Masking Ratio

Metrics	Mask (%)	B.Conv (proposed)		P.Conv		GMCNN		D.Conv	
		Place 2	Open Logo	Place 2	Open Logo	Place 2	Open Logo	Place 2	Open Logo
L^1Loss ↓	1-10	**0.55**	**0.62**	0.68	0.72	0.77	0.71	0.82	0.84
	10-20	**1.19**	**1.69**	1.28	1.72	1.34	1.36	1.47	1.92
	20-30	**2.11**	**2.41**	2.42	2.68	2.56	2.41	2.47	2.97
	30-40	**3.20**	**3.26**	3.43	3.67	3.74	3.79	3.89	3.97
	40-50	**4.51**	**4.49**	4.62	4.63	4.78	4.46	4.62	4.92
PSNR↑	1-10	**34.79**	**30.43**	34.04	31.04	29.82	28.14	30.06	27.04
	10-20	**29.49**	**27.52**	28.75	26.75	27.26	26.87	28.75	26.75
	20-30	**26.03**	**25.47**	25.59	24.59	24.12	23.19	22.70	24.59
	30-40	**23.58**	**23.75**	23.40	18.40	20.41	18.78	23.58	18.12
	40-50	**21.65**	**26.79**	21.56	20.56	20.12	20.01	18.56	18.19
SSIM ↑	1-10	**0.97**	**0.96**	0.82	0.81	0.71	0.86	0.73	0.72
	10-20	**0.94**	**0.90**	0.81	0.89	0.73	0.76	0.69	0.70
	20-30	**0.89**	**0.84**	0.74	0.72	0.64	0.68	0.62	0.51
	30-40	**0.89**	**0.78**	0.78	0.70	0.60	0.54	0.78	0.49
	40-50	**0.73**	**0.72**	0.71	0.70	0.58	0.51	0.72	0.36
L2 Loss↓	1-10	**0.20**	**0.98**	1.30	1.36	1.30	1.76	1.80	2.30
	10-20	**0.61**	**1.33**	1.72	1.78	1.92	3.82	2.42	3.33
	20-30	**1.57**	**1.36**	1.76	1.92	2.36	3.96	4.02	4.36
	30-40	**3.38**	**3.23**	3.92	4.03	2.71	4.02	4.23	4.79
	40-50	**6.89**	**5.89**	7.79	7.89	2.89	4.89	4.26	4.83

However, as the masking ratio increases, the "*Proposed B.Conv*" method maintains its effectiveness in filling more significant regions, while the P.Conv needs to fill up the missing portions. The GMCNN and D.Conv methods need help effectively removing and inpainting larger areas, leading to incomplete and visually different texture appearance in the inpainted images. Understanding these performance characteristics is vital for choosing the most suitable inpainting method based on the specific masking ratio and the desired level of image quality preservation in medical image processing tasks.

6.3.5 Implementation of Bilateral Convolutional Inpainting Model using Different Loss Functions

It targets the reconstruction of per pixels and how smoothly the hole values from the surrounding pixels are predicted. The input with the hole is taken as I_{in} with the binary mask image M (i.e. Representing 0 for holes and 1 for valid pixels), and the predicted output with $I_{out}$and the ground truth images is I_g. The perceptual loss, total variation loss, style loss are calculated along with the hole and valid pixel of the digital images. The per-pixel reconstruction accuracy is improvised using the pixel loss. The pixel loss for the hole region is represented in (6.16), and the valid region is represented in equation (6.17).

$$L_{\square ole} = ||(1 - M) \odot (I_{out} - I_{gt})|| \qquad (6.16)$$

$$L_{valid} = ||M \odot (I_{out} - I_{gt}) \qquad (6.17)$$

The perceptual loss is calculated as in the equation 6.18.

$$L_{perceptual} = \sum_{n=0}^{N-1} ||\psi_n(I_{out}) - \psi_n(I_{gt})| \quad |_1 + \sum_{n=0}^{N-1} ||\psi_n(I_{comp}) - \psi_n(I_{gt})| \quad |_1 \qquad (6.18)$$

It calculates the L^1 distance between I_{out}, I_{comp} and ground truth I_{gt} images. It is also known as feature reconstruction loss, used to validate the similar feature representation of the predicted and ground truth images. The I_{comp} represent the raw output images with non-hole pixels on the images. The ψn defines the activation function for the n^{th} number of selected layers, and each layer has the pooling layer P_n. Similarly, for the style loss I calculated, initially, autocorrelation is applied on each feature map before applying the L^1 Loss. The style loss includes the raw output$L_{styleout}$, as shown in equation (6.19), and the composited output of the model $L_{stylecomp}$, as shown in equation (6.20). The K_n represent

the normalization of $1/C_nH_nK_n$ where n is the selected layer, and C_n is the autocorrelation calculation of the images.

$$L_{styleout} = \sum_{n=0}^{N-1} ||K_n\big((\psi_n)I_{out}\big)^T(\psi_n(I_{out})) - \big(\psi_n(I_{gt})\big)^T\psi_n(I_{gt}))| \quad |_1 \tag{6.19}$$

$$L_{stylecomp} = \sum_{n=0}^{N-1} ||K_n\left((\psi_n)I_{comp}\right)^T(\psi_n(I_{comp})) - \big(\psi_n(I_{gt})\big)^T\psi_n(I_{gt}))| \quad |_1 \tag{6.20}$$

The other loss function is the total variation loss L_{tv}, shown in equation (6.21), in which P represents the area of 1- dilation of pixel of the empty region. It is responsible for the spatial continuity and smoothness of the restored images.

$$L_{tv} = \sum_{(i,j)\epsilon P,(i,j+1)\epsilon P} ||I_{comp}^{i,j+1} - I_{comp}^{i,j}| \quad |_1 + \sum_{(i,j)\epsilon P,(i,j+1)\epsilon P} ||I_{comp}^{i+1,j} - I_{comp}^{i,j}| \quad |_1 \tag{6.21}$$

The total loss L_{total} is represented in the equation (6.22)

$$L_{total} = L_{valid} + 6L_{\square ole} + 0.05L_{perceptual} + 120\big(L_{styleout} + L_{stylecomp}\big) + 0.1L_{tv} \tag{6.22}$$

An ablation study is a crucial analysis carried out to understand the individual contributions of different components or techniques within a model. The ablation study in suggested inpainting attempts to explore the effects of different loss functions when handling large masked regions. The study begins by applying inpainting without any specific masking function. It likely results in incomplete filling of the missing parts since no particular technique is employed to handle the masked areas. This baseline assesses the improvements achieved when incorporating different loss functions. The proposed method is enhanced by introducing various loss functions: perceptual loss, style loss, total variation loss, and pixel loss. These loss functions are combined to predict and fill the large masked regions. Each loss function serves a specific purpose:

- **Perceptual Loss:** It measures the similarity between the predicted inpainted and original images regarding high-level features extracted from deep neural networks

[166]. It helps to ensure that the overall content and structure of the inpainted region match the original image.

- **Style Loss:** It captures the texture and artistic characteristics of the image. By incorporating style loss, the inpainted region is encouraged to retain the texture and style of the surrounding pixels, enhancing visual quality [166].
- **Total Variation Loss:** It promotes smoothness in the inpainted regions, reducing pixel artifacts and preserving overall image coherence.
- **Pixel Loss:** It scales the pixel-wise discrepancy in between the inpainted and original regions. It ensures a closer match between the predicted and actual pixel values, preserving finer details [167].

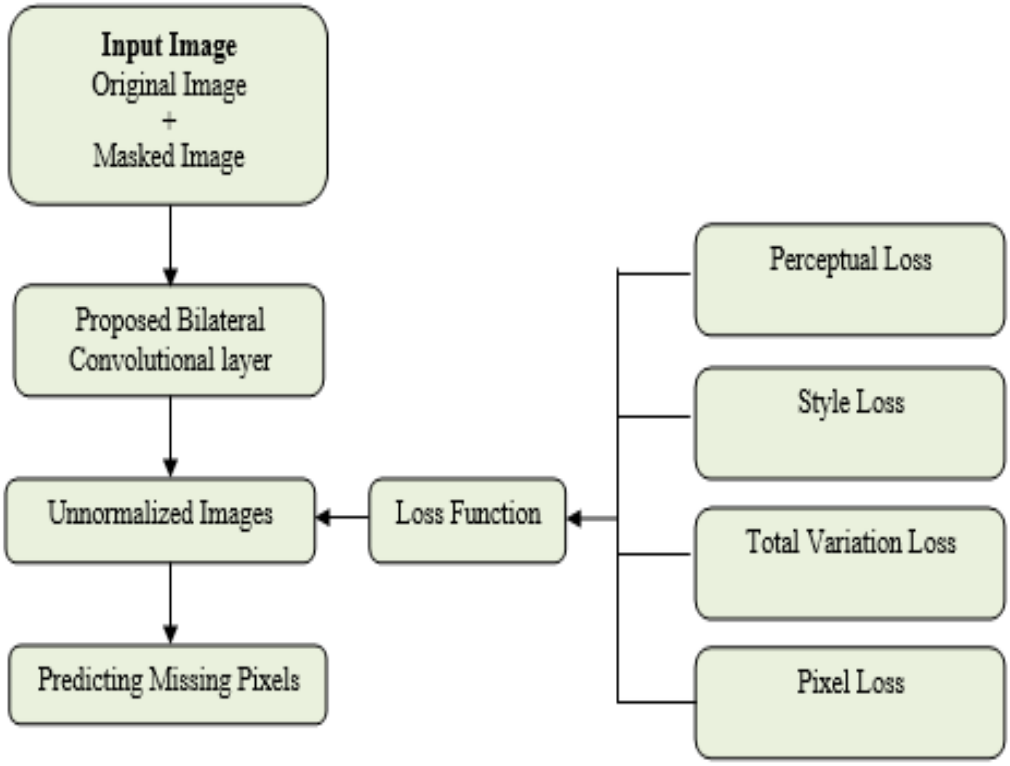


Figure 6.13 Proposed Bilateral Convolutional Inpainting Model on the Different Loss Function

The ablation study systematically assesses the effect of using each loss function independently and in combination with the proposed method. By comparing the results of different setups, the study sheds light on the significance of each loss function and how they contribute to the overall performance of the inpainting model. The proposed methods with different loss function are illustrated in Figure 6.13.

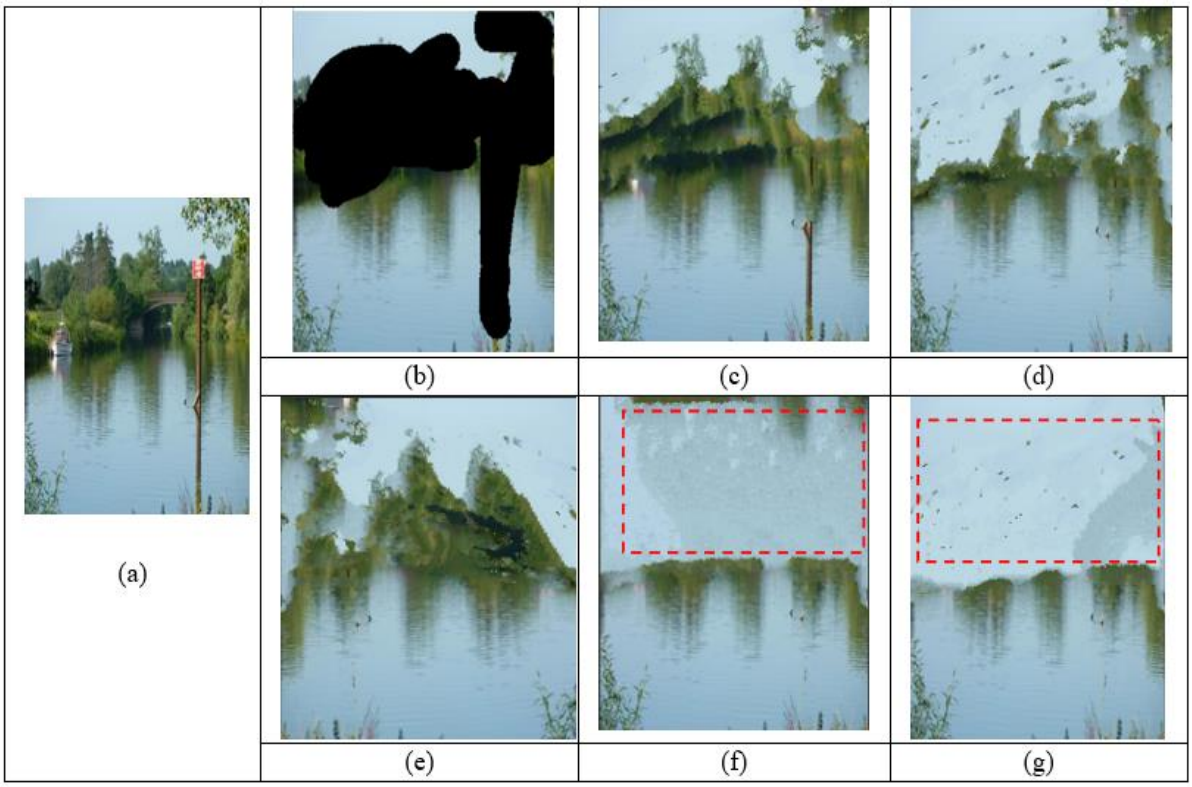


Figure 6.14 Inpainting Digital Images with Different Loss Functions (a) original Images (b) Masked images with a ratio of 40-50%. (c) Inpainted image without loss function. (d) Inpainted image with a style loss function. (e) Inpainted image with a perceptual loss function. (f) Inpainted image with total variation loss function (g) Inpainted image with pixel loss function

The images collected from the place 2 datasets are masked with the ratio of 40-50% to check the loss function to reduce the artefacts during the larger region inpainting process. Inpainting the images without any specific loss function resulted in incomplete region filling. This method was not able to fully restore the missing content. The style loss function also failed to fill the digital images texture. The inpainted regions lacked the full richness and details of the original texture. Similar to the style loss, the perceptual loss function also failed to achieve complete filling of the surface in the digital images. The inpainted regions lacked the original texture's finer details. The total variation loss function successfully filled the regions without leaving any gaps. However, the inpainted image appeared to be manipulated and lost some of its original look. The pixel loss function provided better quality for higher masking ratios in the Place 2 dataset than other loss functions. Although it didn't fully restore the texture, the quality was relatively better than the style and perceptual loss functions as shown in the Figure 6.14.

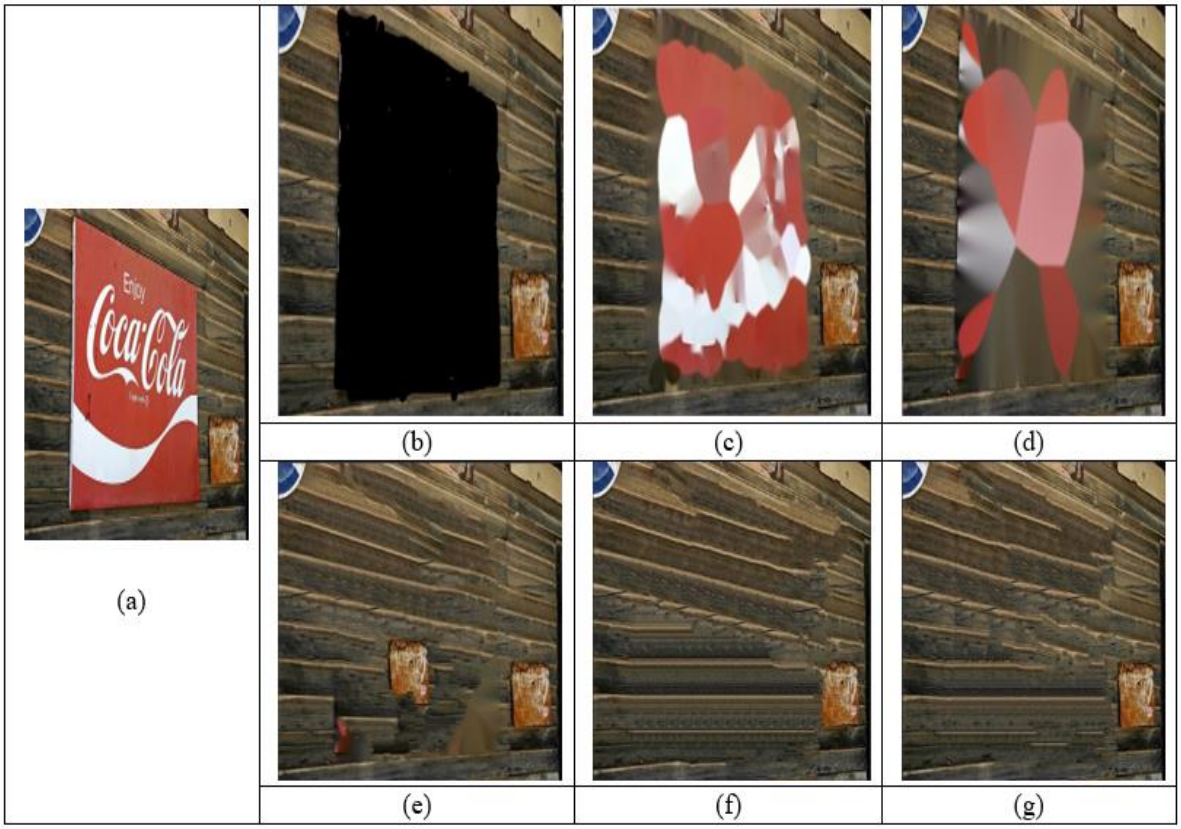


Figure 6.15 Inpainting Digital Images with Different Loss Functions. (a) Original Images (b) Masked images with a 40-50% ratio. (c) Inpainted image without loss function. (d) Inpainted image with a style loss function. (e) Inpainted image with a perceptual loss function. (f) Inpainted image with total variation loss function (g) Inpainted image with pixel loss function

Based on the analysis, it was observed that inpainting without using any loss function resulted in incomplete filling of the regions. The style and perceptual loss functions couldn't fully restore the texture in the digital images. The total variation loss function filled the regions but affected the original appearance. The pixel loss function showed better quality for higher masking ratios, although it didn't fully restore the texture. Similarly, the images are implemented on the masking ratio of 40-50% in the open logo dataset. Based on the analysis, it was observed that inpainting without using any loss function resulted in incomplete filling of the regions as shown in the Figure 6.15. Neither the style nor perceptual loss functions could fully restore the texture in the digital images. The total variation loss function filled the parts but affected the original appearance. The pixel loss function showed better quality for higher masking ratios, although it didn't fully restore the texture.

6.4 Summary

The proposed bilateral convolutional inpainting model effectively removed glare regions from smart colposcopy images and other medical images. It seamlessly eliminated the glare regions while filling them with high-quality pixels. Additionally, it was applied to other datasets, such as Place 2, the Open Logo dataset, and satellite images, to remove desired objects from the images. Across these datasets, the proposed method consistently demonstrated effective performance, removing specified objects and filling the regions with neighboring high-quality pixels. To ensure the robustness of the inpainting method, the proposed model was tested on different masking ratios. The results indicated that the suggested method outperformed previous deep learning inpainting techniques in predicting missing pixel values in empty regions, especially for the largest masked areas. Different loss functions were applied to the proposed method to enhance the inpainting quality in more significant masked regions. An ablation study revealed that combining total variation loss and pixel loss with the proposed method led to better predictions of missing pixel values in more significant masked regions with higher quality. The proposed bilateral convolutional inpainting model remarkably removed glare regions and desired objects from various images. It achieved high-quality inpainting results by leveraging suitable loss functions, particularly in more significant masked areas. The method showed great potential for enhancing image restoration tasks and could be applied effectively in different domains, including medical imaging and satellite imagery.