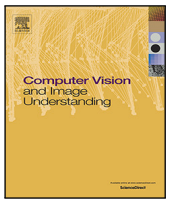
Parametric kernels for artifact mitigation in patch-based image aggregation using generative models

Computer Vision and Image Understanding 260 (2025) 104457

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu



# Parametric kernels for artifact mitigation in patch-based image aggregation using generative models

Nicola Michielli [a], Francesco Marzola [a], Francesco Branciforti [a], Kristen M. Meiburger [a], Alessandro Gambella [b], Massimo Salvi [a,*]

[a] Biolab, PoliTo[BIO]Med Lab, Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
[b] Pathology Unit, Department of Surgical Sciences and Integrated Diagnostics (DISC), University of Genoa, Genoa, Italy

ARTICLE INFO

Communicated by Concetto Spampinato

*Keywords:*
Checkerboard artifacts
Generative models
Gigapixel
Medical imaging
Patch aggregation

ABSTRACT

Recent advancements in artificial intelligence applications have highlighted the effectiveness of generative models for domain transfer, image enhancement and simulation. However, when applied to large-scale gigapixel images, the use of traditional patch-based image aggregation methods introduces checkerboard or blocking artifacts, which compromises image quality and fidelity. In this paper, we propose a parametric kernel that is specifically designed to target the underlying grid structure to mitigate these artifacts. With the use of adjustable zero-padding and linear-padding parameters, our kernel provides fine control over the fusion process by combining a central area of constant weights with border regions of gradually decreasing weights. The proposed method was validated using three medical imaging modalities (digital pathology, fluorescence microscopy and ultrasound imaging) for different generative model tasks. The results showed statistically significant improvements ($p < 0.0001$) in artifact removal when compared to state-of-the-art methods. Quantitative analysis revealed improvements in fusion quality measures for digital pathology, fluorescence microscopy and ultrasound imaging of 7.1%, 27.4% and 20.0%, respectively. Additionally, expert evaluators confirmed superior visual quality and reduced artifacts in blind assessments of reconstructed images, with our method achieving significantly higher scores across all modalities. Our method is versatile, compatible with various generative models and can be easily adjusted by modifying kernel parameters. This kernel-based approach significantly advances the quality of synthesized medical images, directly supporting more reliable clinical assessment and automated analysis.

## 1. Introduction

Rapid advancements have been made recently in artificial intelligence applications, with generative models prominently emerging as effective tools for tasks like domain transfer (Guan and Liu, 2022), image enhancement (Armanious et al., 2020; Salvi et al., 2022), and simulation (Dalmaz et al., 2022). These models have demonstrated their versatility and efficacy in generating realistic and high-quality images, enabling advancements in medical research and clinical practice. However, the computational limitations associated with the input size of generative models present a substantial challenge, particularly when dealing with large-scale images.

To address this issue, image division into smaller patches becomes necessary during the training and inference stages. During training, the input image is partitioned into a number of patches, allowing the generative model to be trained on smaller input sizes. In the inference stage, the image is also divided into patches, and the model is applied using a sliding window approach (Xu et al., 2022). The resulting patches generated by the model are then aggregated to reconstruct the final synthetic image. While this patch-based approach enables the processing of large images, it can potentially introduce a significant challenge that undermines the final result: generated artifacts, commonly known as checkerboard or image blocking artifacts (Wu, 2019), often become visible as illustrated in Fig. 1.

These artifacts arise due to the inherent grid structure resulting from the patch division process. The grid pattern becomes apparent in the synthesized/reconstructed image, compromising the overall quality and visual fidelity. Such artifacts greatly diminish the applicability and utility of generative models not only in medical imaging applications but also in other domains where high-fidelity image generation is crucial, such as creative content generation and satellite imagery analysis.

This problem has gained increasing attention, especially with the widespread adoption of diffusion models for image generation tasks.

* Correspondence to: Biolab, Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129, Turin, Italy.
*E-mail address:* massimo.salvi@polito.it (M. Salvi).

https://doi.org/10.1016/j.cviu.2025.104457
Received 14 November 2024; Received in revised form 28 May 2025; Accepted 29 July 2025
Available online 5 August 2025

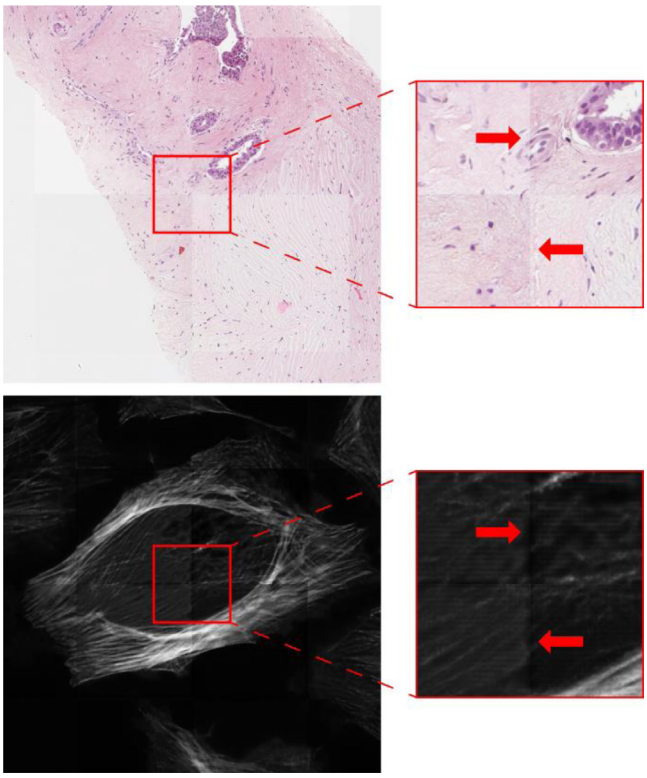


**Fig. 1.** Effects of checkerboard or blocking artifacts after patch aggregation on a RGB image (top) and a grayscale image (bottom), with a zoomed-in view highlighting the artifacts.

Several recent studies have proposed custom architectural modifications to mitigate boundary artifacts, for example through patch-aware attention mechanisms (Haji-Ali et al., 2024; He et al., 2024; Liu et al., 2025). In medical imaging, Jewsbury et al. (2024) demonstrated how such artifacts can severely degrade the quality of gigapixel histology images. The challenge becomes particularly relevant when generating high-resolution outputs where maintaining global coherence is essential.

In this paper, we propose a new parametric kernel-based method that smoothly blends image patches while preserving image details to reduce or eliminate checkerboard artifacts. The key innovation is the use of parametric kernels to weigh the intensity of the generated image predictions during aggregation inference. These kernels target the underlying grid structure of the images, improving the visual quality and fidelity of the synthesized images compared to existing approaches. The main contributions of this paper are the following:

- We provide a precise mathematical definition of a parametric kernel designed specifically to reduce checkerboard artifacts during patch-based image reconstruction.
- The proposed kernel offers a versatile and general formulation that can be applied regardless of the generative model used.
- We extensively evaluate the proposed method using quantitative metrics to assess performance and compare computational times against state-of-the-art approaches employing different models and tasks. In addition, a qualitative analysis was carried out by involving domain experts who scored the visual quality of the reconstructed images.
- The parametric kernels defined here demonstrate potential for future applications, including unpaired tasks, semantic segmentation, classification networks and other large field-of-view imaging applications.

The proposed method shows potential for reducing artifacts, improving visual quality, and enhancing the applicability of generative models in medical imaging. The rest of this paper is organized as follows: Section 2 summarizes the related works; Section 3 provides an exhaustive description of the proposed method; experimental results are reported in Section 4 and Section 5 discusses the work as a whole. In Section 6, the key findings and their significance from the experimental results are briefly summarized.

## 2. Related works

The standard approach involves a sliding window with no overlap, where patches are fed into the generative model and the output patches are stitched together to form the reconstructed image. However, this stitched image presents checkerboard artifacts since the boundary information of each patch is inconsistent after reconstruction. One naïve approach to directly avoid this issue is to simply resize the image to fit the model's input size. While this method bypasses the patch aggregation issue, it produces low-resolution images with interpolation artifacts. This approach is particularly unsuitable for very large images, such as microscopy images, where crucial information would be lost due to the sub-sampling process.

Another suboptimal solution involves training generative models on smaller patches and then directly performing inference on full resolution images through rescaling. However, this process causes unreliable predictions since the network encounters structures at magnification it has never seen during training. Moreover, this approach becomes computationally unfeasible for large-scale images due to large memory demands.

Therefore, several strategies have been proposed to mitigate the patch aggregation challenge over recent years, but they all come with their own limitations. In their study, Müller et al. (Müller and Kramer, 2021) restore the original medical image shape by computing an average value in the overlapping part of two adjacent patches. However, this result is a high blurring effect similar to a lowpass image filter application. Another approach is the "center-crop" strategy, which involves using a sliding window with overlap and retaining only the central portion of the network's prediction (Salvi et al., 2021). However, this approach exponentially increases computational time based on the crop size, and it does not guarantee effective artifact reduction. Another strategy, known as "Patchify", has been proposed by Wu (2021). This method covers the overlapping region between patches with the subsequent patch. However, Patchify only supports image reconstruction with a uniform step size that must be an integer value for all patches. This poses a significant limitation on the choice of step value, especially for images of different sizes. A more recent approach addressing the patch-based reconstruction artifacts is Gaussian reconstruction (Salvi et al., 2023; Xu et al., 2022). This method employs a Gaussian function to smooth the borders of each patch during the aggregation process. This method has two major drawbacks: increased computational time and potential loss of original image dynamic range at fusion points, which can lead to over- or under-saturation issues. While this might not significantly impact classification and segmentation networks, it becomes crucial for generative models where unrealistic medical image details may be generated due to intensity saturation.

The proposed kernel-based approach, as detailed in the following section, addresses the drawbacks of these existing methods. In addition, the proposed method provides a versatile and general formulation that can also reproduce state-of-the-art techniques by varying kernel parameters.

**Table 1**
Overall dataset composition and data splitting into Train and Test subsets for each image task.

| Dataset | Task | Subset | # Images | Tile size (pixels) | Patch size (pixels) |
|---|---|---|---|---|---|
| Digital Pathology (Salvi et al., 2024) | Stain normalization | Train | 500 | 2048 × 2048 | 1024 × 1024 |
| | | Test (Breast) | 100 | 4096 × 4096 | |
| | | Test (Gastric) | 50 | 5632 × 5632 | |
| | | Test (Lung) | 290 | 2800 × 2800 | |
| Fluorescence Microscopy (Hagen et al., 2021) | Image enhancement | Train | 523 | 2048 × 2048 | 512 × 512 |
| | | Test | 200 | | |
| Ultrasound CCA (Meiburger et al., 2022, 2021) | Semantic image synthesis | Train | 2079 | 480 × 480 | 128 × 128 |
| | | Test | 100 | | |

## 3. Materials and methods

### 3.1. Datasets

In this study, three datasets from different medical imaging modalities and scales were utilized. The datasets used for validation were selected to represent a diverse range of medical imaging applications and modalities (Table 1).

The digital pathology dataset comprises histological images from multiple tissue types (breast, gastric, and lung) with different staining protocols. Because of its complex tissue structures and large image sizes (ranging from 2048 × 2048 to 5632 × 5632 pixels), this dataset poses unique challenges for patch aggregation. These images are also a perfect test case for assessing stain normalization methods because they show notable variations in staining patterns and intensity (Salvi et al., 2024).

The fluorescence microscopy dataset consists of 723 images with 2048 × 2048 pixel size. These images capture cellular structures with varying fluorescence intensities and noise levels. The dataset enables a comprehensive evaluation of image enhancement techniques because it contains both high-quality acquisitions and low-signal images with low contrast (Hagen et al., 2021). This dataset is especially well-suited for testing patch aggregation in situations where maintaining fine structural details is essential due to the variety of image quality and cellular patterns.

The ultrasound dataset focuses on the common carotid artery (CCA) and comprises 2179 B-mode ultrasound images. These images were acquired from multiple clinical centers using different ultrasound systems (Meiburger et al., 2022, 2021), introducing variability in image appearance and quality. The dataset was created especially for semantic image synthesis tasks, where it is essential to preserve anatomical consistency across patch boundaries.

### 3.2. Generative models

For each dataset, three different generative models were trained and tested, taking image patches as input as specified in the last column of Table 1. All three models were trained using paired data, which allowed the calculation of quantitative full-reference metrics between the model output and the ground truth results.

#### 3.2.1. Pix2Pix GAN

The first model is the pixel-to-pixel (p2p) GAN (Isola et al., 2017). The model consists of a generator (UNet architecture with 24 layers) and a discriminator (patchGAN with 96 layers). The generator employs skip connections and uses ReLU activations in the encoder and LeakyReLU in the decoder. The discriminator processes 70 × 70 overlapping image patches to determine if they are real or generated. Let $x \in$ X represent real images, z ∈ Z represent random noise, $y \in$ Y represent target images. The discriminator D aims to distinguish between real and fake images, while the generator G aims to create realistic fake images. The least squares loss function for the discriminator is (Mao et al., 2017):

$$\min_D V_{LSGAN}(D) = \frac{1}{2} E_{[x\in X, y\in Y]}\left[D(x) - b^2\right] + \frac{1}{2} E_{[x\in X, z\in Z]}\left[(D(x, G(z)) - a)^2\right] \quad (1)$$

where a = 0 (label for fake data), b = 1 (label for real data). The objective function for the generator is:

$$\min_G V_{LSGAN}(G) = \frac{1}{2} E_{[x\in X, z\in Z]}\left[(D(x, G(x, z)) - c)^2\right] \quad (2)$$

where c = 1 (value for generator to classify data as real). The GAN was optimized via simultaneous gradient descent on D and G, with alternating updates to the discriminator and generator networks. The model was trained using the Adam optimizer with learning rate 2e−4 ($\beta 1$ = 0.5, $\beta 2$ = 0.999) and batch size 4. Training required approximately 24 h on our hardware setup (Ge Force RTX 3090) for 150 epochs.

#### 3.2.2. Pyramid GAN

The second generative model employed is the pyramid GAN (Liu et al., 2022), which also utilizes the least squares objective function. The generator is an attention UNet (Oktay et al., 2018) with 64 layers, and the discriminator is a patchGAN with 64 layers. The generator incorporates self-attention modules after each encoder block to capture long-range dependencies. The model employs the $L_1$ loss, consistent with the p2p GAN and an additional multi-scale loss ($L_{\text{multi-scale}}$). A Gaussian kernel smooths each layer of the generator to achieve the multi-scale pyramid effect. All other parameters were set to their default values as specified in the original paper (Liu et al., 2022). The objective function is defined as:

$$G = arg \max_G \max_D L_{GAN}(G, D) + \lambda_1 L_1 + L_{multi-scale} \quad (3)$$

Training used the Adam optimizer with learning rate 1e−4 and batch size 8. The model completed 150 epochs in approximately 30 h.

#### 3.2.3. Diffusion model

The third and last generative model incorporates a diffusion model based on Denoising Diffusion Probabilistic Models (DDPMs) (Wang et al., 2022). This model with parameters $\theta$ takes as input the semantic layout and the noisy image and tries to learn the distribution $p_\theta(y_0)$ approximating the real distribution $q(y_0)$. The training process of this model optimizes $L_{\text{simple}}$ and $L_{\text{VLB}}$, which are described hereafter. $L_{\text{simple}}$ is the denoising loss with the original image as a reference and is defined as:

$$L_{simple} = \mathbb{E}_{t,y,\epsilon}\left[\left\| y - y_\theta\left(\sqrt{\alpha_t} y + \sqrt{1-\alpha_t}\epsilon, x, t\right)\right\|^2\right] \quad (4)$$

where $\mathbb{E}_{t,y,\epsilon}$ denotes the expectation over three variables – $t$ (the timestep), $y$ (the reference output image), and $\epsilon$ which is noise randomly sampled from the standard Gaussian distribution. $\alpha_t$ is a predefined noise schedule parameter that changes with $t$. The total number

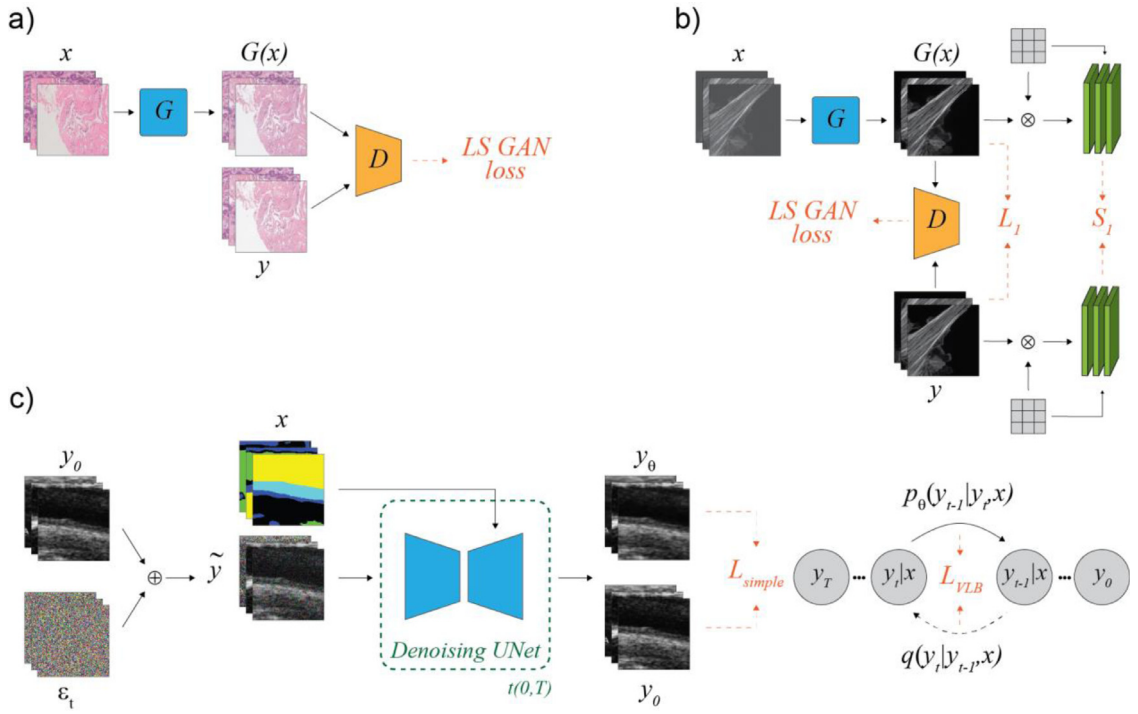



**Fig. 2.** Schematics of the training stage for each of the three applications. For GAN-based applications: G and D are Generator and Discriminator networks, $x$ is the original image, $G(x)$ is the output of the generator, and $y$ is the target image. (a) Pix2Pix architecture for stain normalization. (b) Pyramid GAN for image enhancement. The gray square represents the low-pass Gaussian kernel used to subsample the image; $L_1$ and $S_1$ are the traditional and multiscale loss functions, respectively. (c) Diffusion model for semantic image synthesis. $y_0$ is the original image, $\epsilon_t$ is the added noise at the timestep $t$, $\tilde{y}$ is the image corrupted with Gaussian noise, $x$ is the semantic map conditioning the decoder part of the denoising UNet.

of diffusion steps was set to 100 using DDIM sampling and a cosine noise schedule (Song et al., 2021). The second objective $L_{\mathrm{VLB}}$ is defined as the KL divergence between the estimated distribution $p_\theta\left(y_{t-1} \mid y_t, x\right)$ and diffusion process posterior $q\left(y_t \mid y_{t-1}, x\right)$. The total loss is then calculated as $L_{total} = L_{simple} + \lambda L_{VLB}$, setting lambda to 0.001 as implemented in Stojanovski et al. (2023). Training used the AdamW optimizer with a learning rate of 1e−4. The model required approximately 36 h to complete 150 epochs with batch size 4. During inference, the diffusion process was started from the same noise input for each patch extracted from an image.

### 3.2.4. Model training and evaluation

The three models (Pix2Pix GAN, Pyramid GAN, Diffusion model) were each trained on the three datasets, resulting in a total of 9 models. Training was done for 150 epochs on each model. The training process was conducted using a dedicated workstation equipped with a GeForce RTX3090, a 16-core CPU, and 64 GB of RAM. The PyTorch framework (v1.13.1) was used for both training and testing.

For each application, the generative model that minimized the Fréchet Inception Distance (FID) during training was chosen. Fig. 2 provides examples of the generative tasks performed by these selected models. Specifically, the Pix2Pix GAN performed best for stain normalization in digital pathology images. The Pyramid GAN achieved the best performance for image enhancement in fluorescence microscopy. And the Diffusion model was optimal for semantic image synthesis in ultrasound CCA images.

## 3.3. Parametric kernels for patch aggregation

Here we introduce our novel mathematical kernel formulation for patch aggregation. The proposed kernel has the following characteristics:

- Parametric: the kernel's hyperparameters can be adjusted to define different kernel configurations.
- Square and separable: it can be factored into horizontal and vertical components.
- Adaptive: the kernel can be applied to reconstruct input images of various sizes, and it is independent of the generative model.

Our parametric kernel is designed to smoothly blend adjacent patches by weighting the pixel intensities based on their distance from the patch center. The kernel consists of a central region with constant weights and a peripheral region where the weights gradually decrease in a linear manner. We chose a linear decrease to preserve the original image dynamic range. For image reconstruction, we employ a 2D configuration, as shown in Fig. 3(a). The kernel size is represented as $L \times L$ pixels, and the central region consists of an all-ones matrix. Additionally, a 1D function is introduced, which is a piecewise linear symmetric trapezoidal function. This function allows adjustment of the kernel size ($L$), zero-padding width ($pZ$), and linear padding width ($pL$). Fig. 3(b) visually demonstrates the kernel sliding along one dimension to cover an input length of size $N$. One important computed parameter is the kernel stride ($s$), which defines the step size at which the kernel moves during the aggregation process and is determined as:

$$s = \begin{cases} L - 2pZ & \text{if } pL = 0 \\ L - 2pZ - pL - 1 & \text{otherwise} \end{cases} \tag{5}$$

As shown in Fig. 3(b), kernel stride dictates the movement of the kernel across the input data. For a given input length of size $N$, the number of kernel passes $K$ is computed as:

$$K = \begin{cases} N/s & \text{if } pL = 0 \\ (N + pL - 1)/s & \text{otherwise} \end{cases} \tag{6}$$

In order to obtain an integer number of kernel passes, extra borders have to be computed to fully cover the input length. The extra borders on the left and right, denoted as $extra\_bL$ and $extra\_bR$, respectively, are expressed as:

$$extra\_bL = \begin{cases} pZ & \text{if } pL = 0 \\ pZ + pL & \text{otherwise} \end{cases} \tag{7}$$

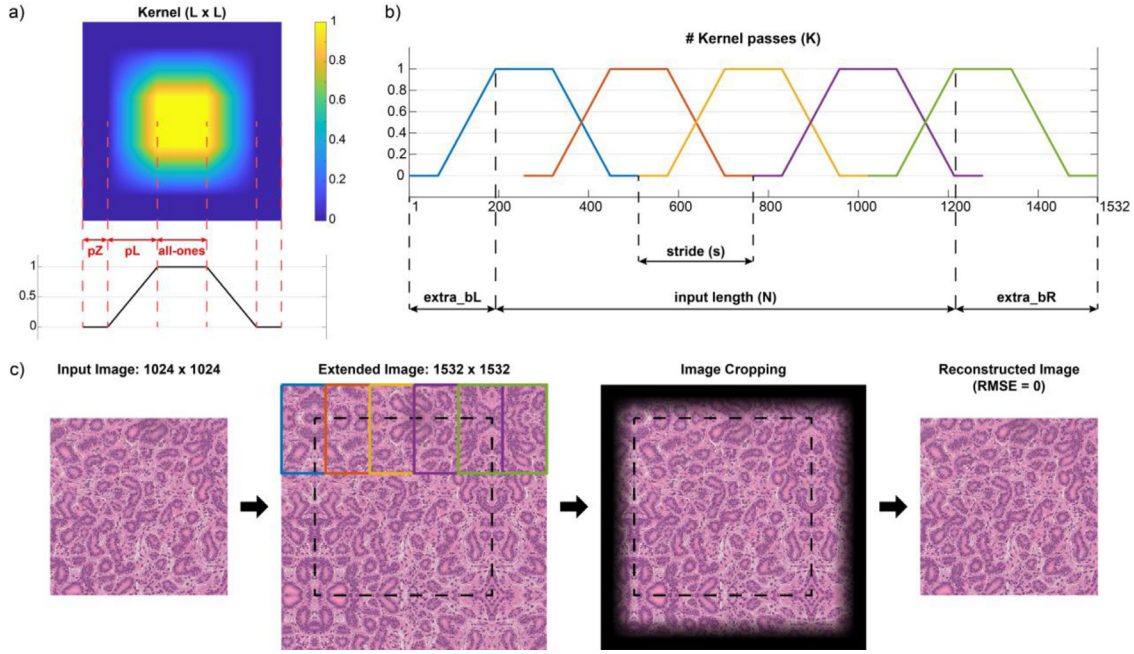



**Fig. 3.** Schematic of parametric kernel for patch aggregation. (a) 1D (bottom) and 2D (top) views of the proposed kernel. The kernel consists of a central constant region and a peripheral linearly decreasing region, with adjustable sizes determined by the hyperparameters. (b) The kernel is applied in a sliding window manner to blend adjacent patches. (c) The complete workflow of kernel-based patch aggregation for 2D images, including image extension, patch extraction, kernel multiplication and output stitching.

$$extra\_bR = \begin{cases} \lceil K \rceil \cdot s - N + pZ & \text{if } pL = 0 \\ \lceil K \rceil \cdot s + 1 - N + pZ & \text{otherwise} \end{cases} \tag{8}$$

where $\lceil \cdot \rceil$ denotes the ceiling operator. It can be noted that the $extra\_bL$ value is immediately defined by kernel hyperparameters, while the $extra\_bR$ also depends on the number of kernel passes.

This process can be extended to a 2D image by also computing top and bottom extra borders with the same rationale. Fig. 3(c) presents an example of an original input image, without showing checkerboard artifacts, to prove how the dynamic range is maintained. Given an input image of size $N \times N$ pixels, an extended image is obtained by mirroring the input image around the boundaries and the extended image size is computed using the extra borders on the left and right and analogously on top and bottom. The kernel passes are highlighted with different colors: each patch extracted from the extended image is multiplied element-wise with the kernel to obtain the output prediction. Finally, the aggregated image is obtained after image cropping to recover the original input size. It is worth noting that the root-mean-square error (RMSE) between the input image and the aggregated image is exactly zero, showing that no over- or under-saturation issues are introduced in the process.

The patch aggregation method is a result of integrating the proposed kernel into the generative model inference. The kernel size is directly determined by the input size of the generative model. The first step involves automatically computing the extended image by adding extra borders according to formulas (7) and (8) on the left and right and analogously on top and bottom. The extended image is then split into $K \times K$ input patches using a stride $s$. Each input patch is fed into the generative model and the predicted output patch is multiplied element-wise with the kernel coefficients; mathematically, the output prediction is estimated as follows:

$$P_{i,j} = P_{i,j} + \tilde{Y}_{i,j} * W \tag{9}$$

where $P_{i,j}$ denotes the output prediction, $\tilde{Y}_{i,j}$ is the output patch predicted by the generative model, $W$ is the 2D parametric kernel and the operator $*$ denotes the Hadamard (element-wise) multiplication. Finally, the output prediction is cropped and rescaled to 0–255 dynamic range to obtain the final 8-bit reconstructed image. The proposed methodology is described in detail in Algorithm 1.

The proposed kernel definition allows for easy implementation of state-of-the-art aggregation approaches, such as the NOcrop and center-crop strategies, by adjusting the hyperparameters. The NOcrop strategy involves a sliding window with no overlap, which can be achieved by setting $pZ = 0$ and $pL = 0$. On the other hand, the center-crop strategy (Salvi et al., 2021), also referred to as the ZEROX configuration in this study, can be obtained by setting $pZ = X$ and $pL = 0$. In summary, the following kernel configurations can be achieved by setting the corresponding hyperparameters:

- LINY ($pZ = 0, pL = Y$): trapezoidal kernel with linear padding only;
- ZEROX+LINY ($pZ = X, pL = Y$): full configuration of kernel;
- NOcrop ($pZ = 0, pL = 0$): non-overlapping sliding window (the all-ones area is 100%);
- ZEROX ($pZ = X, pL = 0$): sliding window with center-crop.

Fig. 4 illustrates the four main configurations of one-dimensional kernel overlapping functions.

### 3.4. Performance metrics

The evaluation of the effective fusion ability of the proposed kernel-based strategy requires the use of image quality assessment metrics. Both full-reference metrics (e.g., Structural Similarity Index, Peak Signal-to-Noise Ratio, Visual Information Fidelity (Sheikh and Bovik, 2006)) as well as no-reference metrics (e.g., Blind/Referenceless Image Spatial Quality Evaluator (Mittal et al., 2012), Natural Image Quality Evaluator (Mittal et al., 2013), Perception-based Image Quality Evaluator (Venkatanath et al., 2015)) are inadequate for detecting checkerboard artifacts. These metrics assess image quality globally, while checkerboard artifacts only occur locally at patch boundaries

**Algorithm 1: Parametric kernels for patch aggregation**

**Input:**

- Input image ($N \times N$)
- Kernel size ($L \times L$)
- Zero-padding width ($pZ$)
- Linear padding width ($pL$)
- Generative model

**Output:** Prediction of the generative model after patch aggregation.

**STEP 1: Kernel definition and extended image computation**

1. *Definition of kernel $W$ with $pZ$ and $pL$*
2. *Kernel stride* $\rightarrow s = \begin{cases} L - 2pZ & if\ pL = 0 \\ L - 2pZ - pL - 1 & otherwise \end{cases}$
3. *Number of kernel passes* $\rightarrow K = \begin{cases} N/s & if\ pL = 0 \\ (N + pL - 1)/s & otherwise \end{cases}$
4. *Extra border on the left* $\rightarrow extra\_bL = \begin{cases} pZ & if\ pL = 0 \\ pZ + pL & otherwise \end{cases}$
5. *Extra border on the right* $\rightarrow extra\_bR = \begin{cases} \lceil K \rceil \cdot s - N + pZ & if\ pL = 0 \\ \lceil K \rceil \cdot s + 1 - N + pZ & otherwise \end{cases}$
6. *Extended image* ($extra\_bL + N + extra\_bR$) *obtained by mirroring the input image.*

**STEP 2: Generative model inference with patch aggregation**

1. *Initialize output prediction $P$*
2. ***for*** *each row index* ($i$) ***do***
3. ***for*** *each column index* ($j$) ***do***
4. *Generative model inference* $\rightarrow \tilde{Y}_{i,j}$
5. $P_{i,j} = P_{i,j} + \tilde{Y}_{i,j} * W$
6. ***end for***
7. ***end for***
8. *Prediction cropping and rescaling to obtain the final output image.*

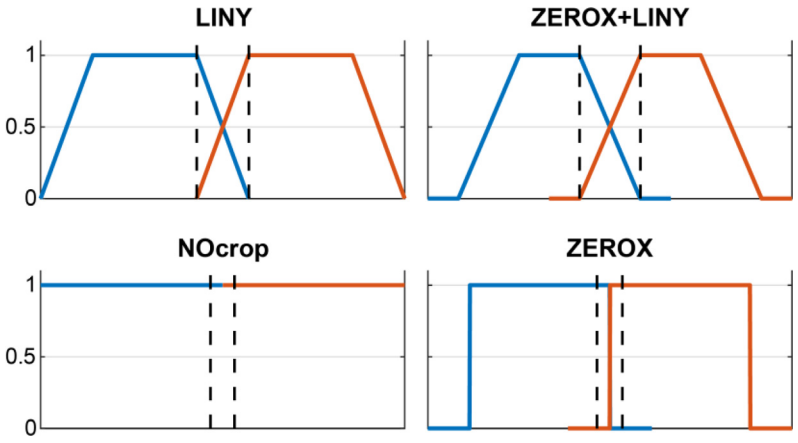



**Fig. 4.** Comparison of different kernel configurations for patch aggregation. The proposed trapezoidal (LINY) and full (ZEROX+LINY) configurations provide a smooth transition between adjacent patches, while the non-overlapping (NOcrop) and center-crop (ZEROX) strategies result in abrupt changes.

during fusion (Jiang et al., 2020; Xu et al., 2022). To specifically target these artifacts, we computed local full-reference metrics focused on the fusion points from patch aggregation. As shown in Fig. 5, this procedure allows a precise assessment of the checkerboard artifacts by evaluating metrics only at the patch boundaries, i.e., fusion points during patch aggregation.

The proposed kernel strategy facilitates easy identification of fusion edges for different kernel configurations. For non-overlapping and center-crop sliding windows (NOcrop and ZEROX configurations), we define artifact regions for metric computation using $M$ pixels on both sides of the fusion edge. In the trapezoidal (LINY) and full (ZEROX+LINY) kernel configurations, the fusion region is determined by the width of the linear padding ($pL$), as shown with dashed black vertical lines in Fig. 4.

For quantitative performance comparison, we compute full-reference metrics such as Pearson correlation coefficient (PCC) and root-mean-square error (RMSE) exclusively within the fusion regions. In the case of digital pathology, RGB images are first converted to grayscale before metric computation. Additionally, all RMSE values are computed using the 0–255 image dynamic range. Higher PCC values combined with lower RMSE values indicate better fusion ability in the mitigation of checkerboard artifacts. To avoid optimizing PCC and RMSE values in opposite directions simultaneously, we proposed a distance metric (DIST) that considers both full-reference metrics for the

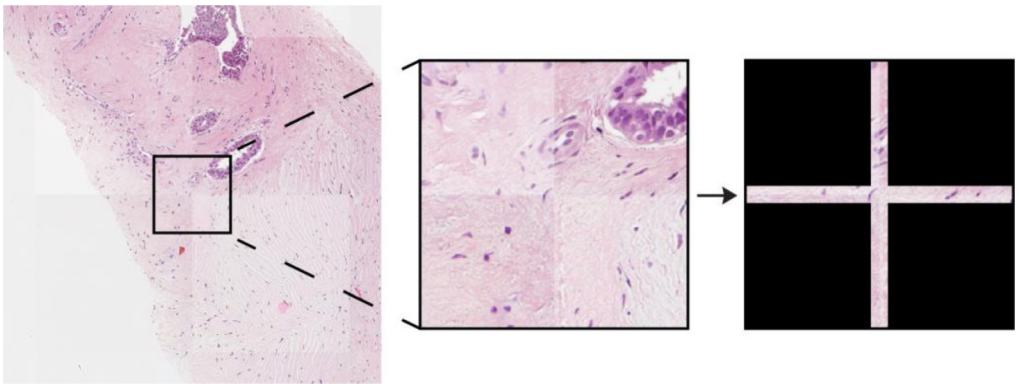


**Fig. 5.** Two-dimensional view of fusion regions retained for local full-reference metrics computation. Blacked out image regions do not contribute to metric calculation.

**Table 2**
Kernel configurations with zero (ZEROX) and linear (LINY) padding for test image data of each generation task. Each column represents a kernel configuration. X represents the number of pixels of parameter pZ, while Y refers to the parameter pL. For instance, ZERO16+LIN48 is a kernel with pZ equal to 16 pixels and pL equal to 48 pixels.

| Subset | Params | Kernel configurations | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | NOcrop | ZEROX | | LINY | | ZEROX+LINY | | | |
| Digital Pathology (Test - Breast) | *pZ* | 0 | 64 | 128 | 0 | 0 | 32 | 64 | 16 | 32 |
| | *pL* | 0 | 0 | 0 | 64 | 128 | 32 | 64 | 48 | 96 |
| | *(KxK)* | 16 | 25 | 36 | 25 | 25 | 25 | 36 | 25 | 25 |
| Digital Pathology (Test - Gastric) | *pZ* | 0 | 64 | 128 | 0 | 0 | 32 | 64 | 16 | 32 |
| | *pL* | 0 | 0 | 0 | 64 | 128 | 32 | 64 | 48 | 96 |
| | *(KxK)* | 36 | 49 | 64 | 36 | 49 | 49 | 49 | 49 | 49 |
| Digital Pathology (Test - Lung) | *pZ* | 0 | 64 | 128 | 0 | 0 | 32 | 64 | 16 | 32 |
| | *pL* | 0 | 0 | 0 | 64 | 128 | 32 | 64 | 48 | 96 |
| | *(KxK)* | 9 | 16 | 16 | 9 | 16 | 16 | 16 | 16 | 16 |
| Fluorescence Microscopy (Test) | *pZ* | 0 | 32 | 64 | 0 | 0 | 16 | 32 | 8 | 16 |
| | *pL* | 0 | 0 | 0 | 32 | 64 | 16 | 32 | 24 | 48 |
| | *(KxK)* | 16 | 25 | 36 | 25 | 25 | 25 | 36 | 25 | 25 |
| Ultrasound CCA (Test) | *pZ* | 0 | 8 | 16 | 0 | 0 | 4 | 8 | 2 | 4 |
| | *pL* | 0 | 0 | 0 | 8 | 16 | 4 | 8 | 6 | 12 |
| | *(KxK)* | 16 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| all-ones area (%) | | 100 | 76.56 | 56.25 | 76.56 | 56.25 | 76.56 | 56.25 | 76.56 | 56.25 |

evaluation of different patch aggregation methods:

$$DIST = \sqrt{(100 \cdot (1 - PCC))^2 + (RMSE)^2} \tag{10}$$

The minimization of this metric results in a higher performance of a kernel configuration in reducing checkerboard or image blocking artifacts.

## 4. Results

### 4.1. Kernel configurations and parameters

Different kernel configurations were tested to evaluate the performance of patch aggregation during generative model inference. Table 2 provides the details for each kernel configuration, including the zero-padding width ($pZ$), linear padding width ($pL$), and the number of kernel passes ($K \times K$) for all test subsets. The kernel size ($L \times L$) is set to match the input size of each generative model: 1024 × 1024 for the digital pathology image task, 512 × 512 for the fluorescence microscopy task, and 128 × 128 for the ultrasound CCA task.

### 4.2. Quantitative performance comparison

For each image task, quantitative performance evaluations are conducted for all proposed kernel configurations and state-of-the-art approaches by computing local full-reference metrics (PCC, RMSE and DIST) within the fusion regions (Fig. 5). The regions of artifacts for metric computation are defined by using $M$ pixels on both the right and left sides of the fusion edge. For digital pathology and fluorescence microscopy images, the value of $M$ was set to 10, while for the ultrasound images, which are smaller in size (Table 1), a value of 4 was used. In the trapezoidal (LIN) and full (ZERO+LIN) kernel configurations, the fusion region is precisely determined by the width of the linear padding ($pL$).

Ablation study is reported by comparing the NOcrop strategy (which simply stitches the generative model output with no overlap) with the application of different kernel configurations (in this case the generative model outputs are aggregated by using the proposed kernel-based formulation). Average values for digital pathology, fluorescence microscopy, and ultrasound CCA image datasets are presented in Tables 3, 4, and 5, respectively. As reported in Table 2, two parameter sets are tested for state-of-the-art center-crop approach (ZEROX), two parameter sets are evaluated for trapezoidal kernel configuration (LINY) and four parameter sets are tested for full configuration of kernel (ZEROX+LINY). The best configurations for the same kernel type in terms of lowest DIST metric are reported in Tables 3, 4 and 5.

We also compared our parametric kernels to the Patchify method (Wu, 2021). To ensure an equal comparison, we set the step size to achieve the same number of passes as the best kernel configuration

**Table 3**
Average quantitative performance metrics on digital pathology test images. The best values for each metric are shown in bold.

| Subset | Metric | NOcrop | Patchify | ZEROX | LINY | ZEROX+LINY |
|---|---|---|---|---|---|---|
| Breast | PCC | 0.9891 | 0.9887 | 0.9904 | **0.9919** | 0.9917 |
| | RMSE | 6.10 | 5.96 | 5.76 | **5.56** | 5.62 |
| | DIST | 6.25 | 6.19 | 5.91 | **5.66** | 5.72 |
| | Time (s) | **1.07** | 1.53 | 2.02 | 1.53 | 1.53 |
| Gastric | PCC | 0.9896 | 0.9905 | 0.9911 | **0.9923** | 0.9922 |
| | RMSE | 5.79 | 5.58 | 5.50 | 5.30 | **5.23** |
| | DIST | 5.89 | 5.67 | 5.59 | 5.37 | **5.30** |
| | Time (s) | **2.30** | 2.92 | 3.57 | 2.91 | 2.86 |
| Lung | PCC | 0.9895 | 0.9906 | 0.9913 | 0.9922 | **0.9924** |
| | RMSE | 7.57 | 7.47 | 7.36 | 7.12 | **7.11** |
| | DIST | 7.67 | 7.56 | 7.43 | 7.18 | **7.17** |
| | Time (s) | **0.57** | 0.93 | 0.92 | 0.94 | 0.95 |
| Overall | PCC | 0.9894 | 0.9902 | 0.9909 | **0.9921** | **0.9921** |
| | RMSE | 7.03 | 6.91 | 6.81 | 6.58 | **6.57** |
| | DIST | 7.15 | 7.03 | 6.90 | 6.65 | **6.64** |
| | Time (s) | **0.88** | 1.29 | 1.27 | 1.29 | 1.30 |

**Table 4**
Average quantitative performance metrics on fluorescence microscopy test images. The best values for each metric are shown in bold.

| Metric | NOcrop | Patchify | ZEROX | LINY | ZEROX+LINY |
|---|---|---|---|---|---|
| PCC | 0.9542 | 0.9616 | 0.9670 | **0.9728** | 0.9726 |
| RMSE | 5.13 | 4.98 | 4.45 | **4.17** | 4.21 |
| DIST | 7.25 | 6.62 | 5.83 | **5.26** | 5.32 |
| Time (s) | **0.38** | 0.63 | 0.77 | 0.56 | 0.56 |

**Table 5**
Average quantitative performance metrics on ultrasound CCA test images. The best values for each metric are shown in bold.

| Metric | NOcrop | Patchify | ZEROX | LINY | ZEROX+LINY |
|---|---|---|---|---|---|
| PCC | 0.8180 | 0.8395 | 0.8488 | **0.8745** | 0.8718 |
| RMSE | 23.73 | 25.86 | 23.53 | 20.88 | **20.54** |
| DIST | 30.17 | 30.63 | 28.16 | **24.51** | 24.56 |
| Time (s) | **68.16** | 106.60 | 106.52 | 106.27 | 106.43 |

of our method. With $K \times K$ passes, the step size of Patchify can be computed as follows:

$$step = \frac{(N - L)}{(K - 1)} \tag{11}$$

where $N \times N$ denotes the input image size and $L \times L$ is the patch size. According to this definition, step values equal to 768, 768, 592, 384, 88 pixels are set for breast (digital pathology), gastric (digital pathology), lung (digital pathology), fluorescence microscopy and ultrasound CCA images, respectively.

When comparing computational times, the non-overlapping kernel configuration (NOcrop) consistently demonstrates the best performance since it requires the fewest number of kernel passes. In terms of quantitative image quality metrics (PCC, RMSE, and DIST), the trapezoidal (LINY) and the full (ZEROX+LINY) kernel configurations yield the best results, merging effectively adjacent patches during the aggregation process. More specifically, the best configurations in terms of parameter set ($pZ$ and $pL$) are the ZERO32+LIN32 for digital pathology image task, LIN32 for fluorescence microscopy and LIN8 for ultrasound CCA.

Fig. 6(a) provides a summary of the quantitative performances, in terms of DIST metric, comparing the state-of-the-art configurations (NOcrop, Patchify and ZEROX — also known as center-crop) with the proposed kernel configurations (LINY and ZEROX+LINY). Paired-sample t-tests are conducted to compare the best and worst configurations in terms of DIST value for each dataset (digital pathology, fluorescence microscopy, and ultrasound CCA). These tests reveal statistically significant differences ($p < 0.0001$). In Fig. 6(b), the computational overhead introduced by the proposed parametric kernel is reported by comparing the computational times to reconstruct tile images with increasing size. The best configurations for trapezoidal kernel (LINY) and full configuration of the kernel (ZEROX+LINY) are reported for each image task. The computational overhead is expressed as the ratio of the time spent processing a single tile image with the proposed kernel with respect to the time spent processing the same tile image with the state-of-the-art direct patch stitching approach (NOcrop).

Across the three imaging modalities, our analysis revealed distinct optimal configurations. For digital pathology images, the ZEROX+LINY kernel configuration achieved the best performance with an average DIST value of 6.64 ± 2.71, significantly outperforming the state-of-the-art NOcrop approach, which showed the worst performance (DIST = 7.15 ± 2.67). In fluorescence microscopy images, the LINY kernel configuration performed best (DIST = 5.26 ± 1.91), showing substantial improvement over the NOcrop approach, which yielded the poorest results (DIST = 7.25 ± 1.92). For ultrasound CCA images, the LINY kernel configuration again demonstrated superior performance (DIST = 24.51 ± 7.00), while the Patchify approach showed the worst results (DIST = 30.63 ± 8.37).

Fig. 7 provides a visual comparison of these results, with a zoomed-in view showing the checkerboard artifacts at fusion points. Corresponding performance metrics are also included for each patch aggregation method.

### 4.3. Qualitative assessment of reconstructed image

To complement quantitative assessment, three domain experts evaluated the visual quality of the reconstructed images: a pathologist with 8 years of experience in digital pathology (A.G.), a computer scientist specialized in fluorescence microscopy analysis (M.S.), and a professor with extensive experience in vascular ultrasound imaging (K.M.M.). Each expert independently rated their own unique set of images, and no two evaluators rated the same image.

Based on the quantitative results from Section 4.2, we selected the best-performing kernel configuration for each modality: ZEROX+LINY for digital pathology and LINY for both fluorescence microscopy and ultrasound images. The evaluation was conducted using a blind, randomized approach comparing three reconstruction methods: the standard NOcrop approach, the Patchify method (Wu, 2021), and the proposed kernel-based configuration. For each imaging modality, 50 reconstructed images were randomly selected from the test set. These

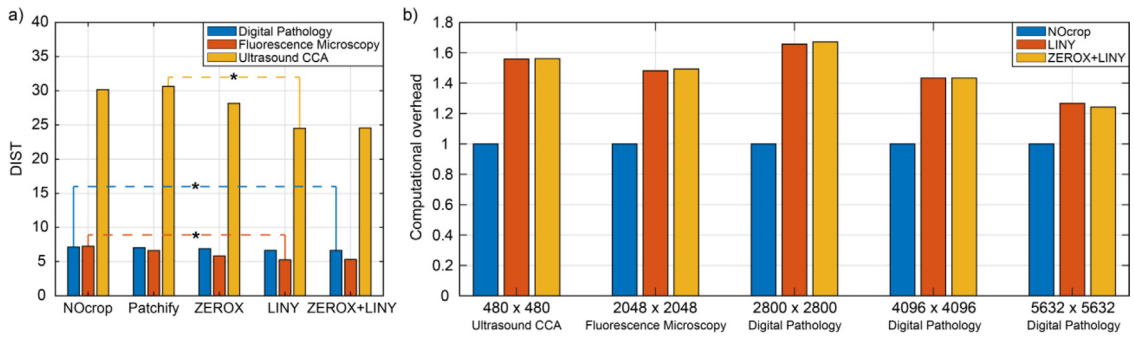



**Fig. 6.** (a) Average DIST values of state-of-the-art (NOcrop, Patchify and ZEROX — also known as center-crop) vs. proposed kernel configurations (trapezoidal – LINY – and full – ZEROX+LINY – configurations) for digital pathology (blue), fluorescence microscopy (red) and ultrasound CCA (orange) reconstructed images. The asterisk (*) denotes a statistically significant difference ($p<0.0001$) between worst and best configurations, highlighted with dashed lines for each image task. (b) Computational overhead introduced by the proposed kernel formulation with respect to the NOcrop approach for patch aggregation in tile images with increasing size.

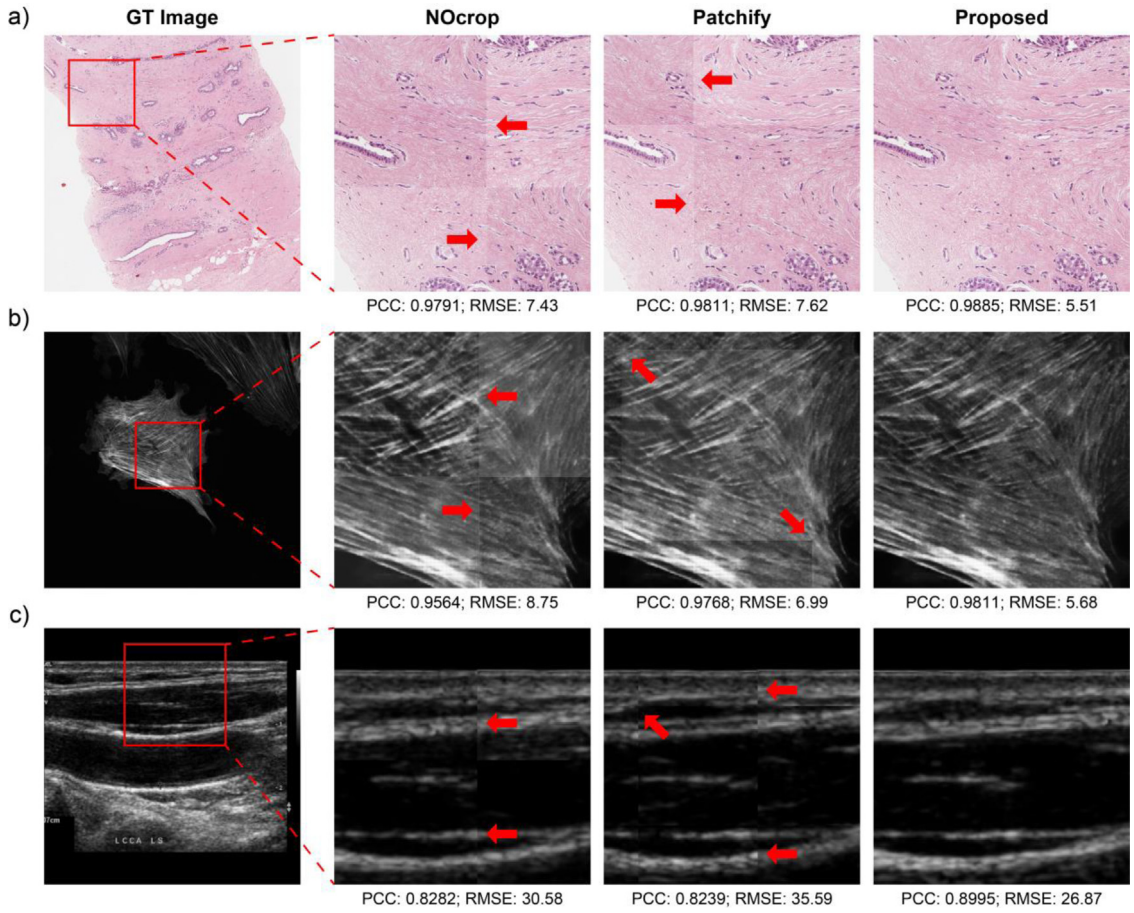



**Fig. 7.** Sample images of digital pathology (a), fluorescence microscopy (b) and ultrasound CCA (c) dataset, are shown for state-of-the-art visual comparison with performance metrics (PCC and RMSE). Ground truth (GT) images are reported in the first column, zoomed-in views of state-of-the-art approaches are shown in the second and third columns, and the proposed kernel-based reconstructed images are reported in the last column. Checkerboard artifacts at the fusion points are highlighted with red arrows.

images were presented to the respective expert, with each set containing the same image reconstructed using the three different methods in random order. The experts were unaware of which reconstruction method was used for each image.

The evaluation protocol required experts to rate each reconstructed image on a 5-point qualitative scale based on the absence of patch boundary artifacts (where 1 indicates prominent artifacts and 5 indicates complete absence of artifacts). For digital pathology, our method achieved a mean score of $4.06 \pm 0.96$, outperforming both NOcrop ($3.08 \pm 1.43$) and Patchify ($3.38 \pm 1.31$) approaches, with a statistically significant improvement ($p < 0.0001$). In fluorescence microscopy analysis, our kernel-based configuration received a mean score of $4.56 \pm 0.58$, compared to NOcrop ($3.00 \pm 0.93$) and Patchify ($2.70 \pm 0.81$), with a statistically significant difference ($p < 0.0001$). For ultrasound CCA images, our method scored $3.56 \pm 0.88$, while NOcrop and Patchify methods scored $1.18 \pm 0.44$ and $1.10 \pm 0.36$ respectively, with a statistically significant difference ($p < 0.0001$). Results are summarized in Table 6.

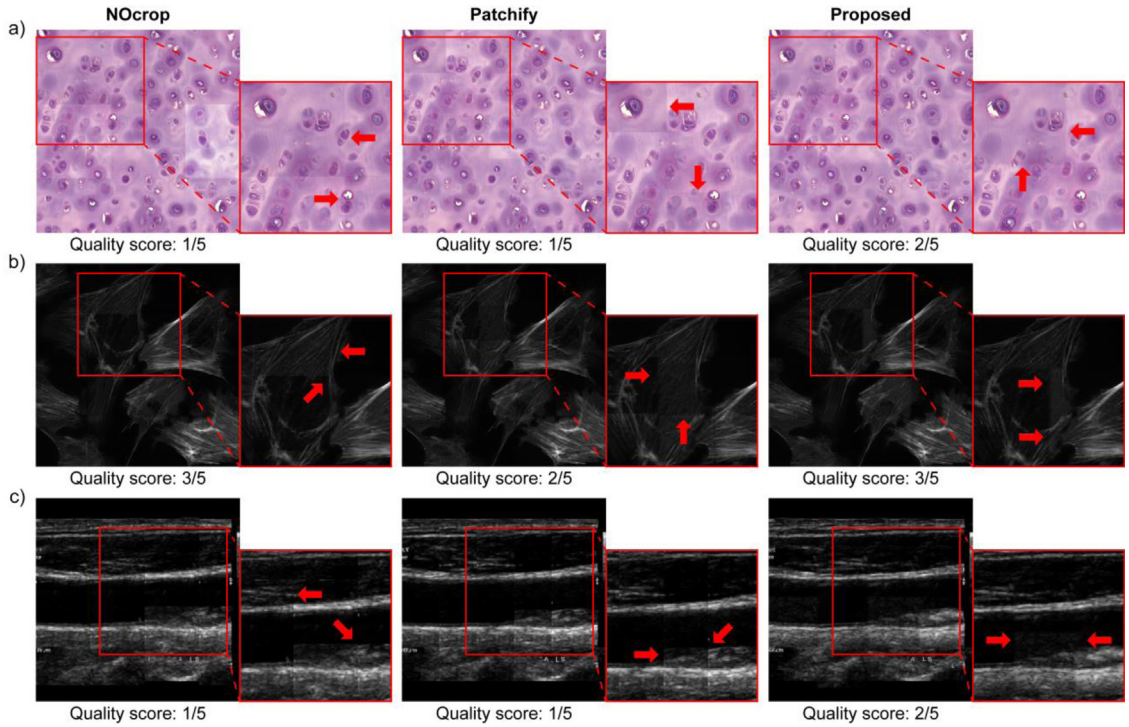



**Fig. 8.** Failure cases for (a) digital pathology, (b) fluorescence microscopy, and (c) ultrasound CCA image task. State-of-the-art approaches (NOcrop and Patchify) are compared with the proposed kernel-based strategy in the worst-case scenario. Checkerboard artifacts at the fusion points are highlighted with red arrows in zoomed-in views. Visual quality scores evaluated by experts are reported for each evaluated image.

**Table 6**
Expert assessment results comparing three reconstruction methods across different imaging modalities. Scores are reported as mean ±standard deviation (n=50 images per modality). Best results are highlighted in bold.

| Dataset | NOcrop | Patchify | Proposed kernel |
|---|---|---|---|
| Digital Pathology | 3.08 ± 1.43 | 3.38 ± 1.31 | **4.06*** ± 0.96 |
| Fluorescence Microscopy | 3.00 ± 0.93 | 2.70 ± 0.81 | **4.56*** ± 0.58 |
| Ultrasound CCA | 1.18 ± 0.44 | 1.10 ± 0.36 | **3.56*** ± 0.88 |

* Denotes a statistically significant improvement ($p < 0.0001$).

Fig. 8 provides examples of failure cases with the lowest visual quality score evaluated by experts, where the proposed kernel does not effectively remove checkerboard artifacts, for each medical imaging modality.

## 5. Discussion

The task of patch-based image aggregation is crucial in high-resolution imaging and gigapixel analysis. In this paper, we present a novel approach that addresses the challenge of patch aggregation artifacts commonly encountered when employing AI generative models in medical imaging applications, such as digital pathology gigapixel imaging, fluorescence microscopy and ultrasound imaging. Our proposed method introduces a mathematical formulation of parametric kernels that offers several novel aspects, including the generalization of existing patch aggregation techniques and the ability to preserve image dynamics at the fusion points. These kernels provide a versatile solution that can be applied to various generative models without the need for retraining or customization.

The validation process posed certain difficulties due to the lack of specific metrics optimized for evaluating the quality of images with aggregation artifacts. We evaluated our method using local full-reference metrics by comparing the synthesized images to ground truth images, only at the boundary regions between adjacent patches, as indicated in Fig. 5. Specifically, we utilized quantitative metrics such as PCC, RMSE, DIST metric and computational times. Notably, configurations incorporating the linear component of the kernel consistently outperformed other configurations. Specifically, the trapezoidal (LINY) and full (ZEROX+LINY) kernel configurations demonstrated superior results in terms of image quality, as evaluated by the selected metrics. As summarized in Fig. 6(a), it can be observed a statistically significant (paired-sample t-test, $p < 0.0001$) reduction in terms of DIST metric of 7.1%, 27.4% and 20.0% for digital pathology, fluorescence microscopy and ultrasound CCA image task, respectively, between the proposed kernel-based strategy and the state-of-the art. The linear region of the kernel played a pivotal role in the fusion process, effectively addressing the issue of the underlying grid structure in the aggregated output image. This can be visually observed in Fig. 7, where the artifacts resulting from the patch-based aggregation are noticeably resolved by our approach.

Regarding computational efficiency, Fig. 6(b) illustrates that while our method introduces some computational overhead compared to the NOcrop approach, this overhead decreases as image size increases – from approximately 60% for smaller images to around 30% for larger images (5632 × 5632 pixels). This indicates favorable scalability for large-scale medical images, which is particularly relevant for gigapixel imaging applications. Additionally, both kernel configurations (LINY and ZEROX+LINY) exhibit comparable computational costs, allowing flexibility in kernel selection without significant performance penalties.

These quantitative improvements in artifact reduction are further reinforced by expert evaluations, where our method achieved significantly higher scores across all imaging modalities. In digital pathology, our method scored 4.06 ± 0.96 compared to NOcrop (3.08 ± 1.43) and Patchify (3.38 ± 1.31). Even more pronounced improvements were observed in fluorescence microscopy (4.56 ± 0.58 vs. 3.00 ± 0.93 and 2.70 ± 0.81) and ultrasound CCA imaging (3.56 ± 0.88 vs. 1.18 ± 0.44 and 1.10 ± 0.36). The statistical significance ($p < 0.0001$) of these improvements across all modalities provides strong validation of our method's effectiveness.

Comparing our method to existing algorithms, we identified several advantages. Firstly, the proposed kernel-based approach can be adapted to different input sizes of the generative models and can be employed to patch together images of different sizes, without any constraints. On the contrary, the Patchify method (Wu, 2021), is restricted to a uniform step size which must satisfy the required condition, according to formula (11), for success: this poses a big limitation on the choice of step value, especially for images with different sizes. Additionally, our approach generalizes current patch aggregation techniques, surpassing their performance in terms of the evaluated metrics. The proposed method ensures the preservation of image dynamics at the fusion points, resulting in a more accurate representation of the original image. The versatility of our approach allows for various configurations by adjusting the contributions of the linear and zero parts of the kernel, making it adaptable to different generative models. The proposed kernel functions equally well with grayscale or RGB images and can be easily extended to reconstruct matrices with a generic number of channels, such as softmax probability maps with a different number of predicted classes.

In probabilistic models such as diffusion models, the application of methods like the non-overlapping sliding window (NOcrop) can lead to abrupt transitions between patches. This may result in visible artifacts and inaccurate representation of the underlying structure. Additionally, these models have significant computational demands. By minimizing the width of fusion regions, inference times can be substantially reduced. Fine-tuning the linear and zero sections of the kernel is a potential solution for decreasing computational load while retaining optimal reconstruction quality. This improvement could extend the applicability of such models to near-real time applications.

It is important to note that while our kernel-based approach significantly mitigates checkerboard artifacts, its effectiveness ultimately depends on the quality and stability of the underlying generative model's output. As illustrated in Fig. 8, when the generative model produces inconsistent or unstable results – particularly when operating in out-of-domain scenarios or on challenging image regions – our method can reduce but not completely eliminate artifacts. This limitation stems from the fact that our kernel operates on the generated patches and cannot compensate for fundamental inconsistencies in the generative model's output across adjacent patches. Nevertheless, even in these worst-case scenarios, our approach still provides substantial improvements over standard methods, as evidenced by the expert visual quality scores reported in Fig. 8.

While our study demonstrates the potential of parametric kernels for artifact mitigation in patch-based image aggregation, there are several avenues for future research that could enhance the impact of this work. Developing dedicated no-reference metrics specifically designed to quantify the presence and severity of patch aggregation artifacts would enable more accurate and automated evaluation of different methods. Future research could explore adaptive kernel configurations that automatically optimize parameters based on image content and model characteristics, potentially further improving the quality of patch-based image reconstruction. The proposed kernel configuration was presented for a square input patch size. However, in future developments, it can be extended to accommodate rectangular patch sizes by incorporating varying zero- and linear-padding values along the rows or columns.

Our kernel-based approach holds potential for expansion beyond medical imaging applications into other domains that employ generative models and patch-based image reconstruction. To validate this broader applicability, we conducted additional experiments using the FloodNet satellite imagery dataset (Rahnemoonfar et al., 2021), where we applied our method to generate synthetic satellite images from semantic annotations. As illustrated in Fig. 9, our parametric kernel configurations successfully mitigated boundary artifacts in these non-medical images as well.

Importantly, patch-based processing is not limited to generative models but is also intrinsic to certain acquisition schemes—such as fluorescence microscopy or tiled scanning systems—where images are natively captured in patches. In these contexts, checkerboard artifacts can still occur and compromise visual quality, making our method a valuable tool for enhancing image fidelity across a wider range of use cases.

Beyond the demonstrated scenarios, our approach is also well-suited for super-resolution and artistic image synthesis tasks (e.g., style transfer), where preserving coherence across patch boundaries is critical. The theoretical foundation of our method supports its application across any patch-based imaging pipeline where boundary artifacts pose a challenge.

## 6. Conclusion

In conclusion, our proposed method presents a novel solution for reducing checkerboard and blocking artifacts in patch-based image aggregation. The key innovation is the introduction of parametric kernels that can be tuned to target these artifacts. Our approach is versatile, as the kernels are compatible with various generative models, providing a robust framework for artifact reduction. We demonstrated the capability of the method for three different medical imaging applications, achieving significant improvements in fusion quality measures for digital pathology (7.1%), fluorescence microscopy (27.4%), and ultrasound imaging (20.0%). These quantitative improvements were further validated through expert evaluations, which showed statistically significant ($p < 0.0001$) preference for our method across all modalities. The mathematical formulation of the parametric kernels demonstrates advantages over existing techniques, and the comprehensive validation process supports its effectiveness. The flexibility of our kernel formulation creates opportunities for expansion to other applications involving patch-based image reconstruction and synthesis.

## CRediT authorship contribution statement

**Nicola Michielli:** Software, Conceptualization, Writing – original draft, Methodology, Visualization, Formal analysis. **Francesco Marzola:** Writing – original draft, Visualization, Software, Data curation, Methodology. **Francesco Branciforti:** Writing – review & editing, Visualization, Data curation. **Kristen M. Meiburger:** Writing – review & editing, Resources, Visualization, Data curation, Supervision. **Alessandro Gambella:** Visualization, Writing – review & editing, Supervision. **Massimo Salvi:** Writing – original draft, Supervision, Visualization, Conceptualization, Validation, Resources, Formal analysis, Software, Data curation, Methodology.

## Declaration of competing 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.

## Data availability

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

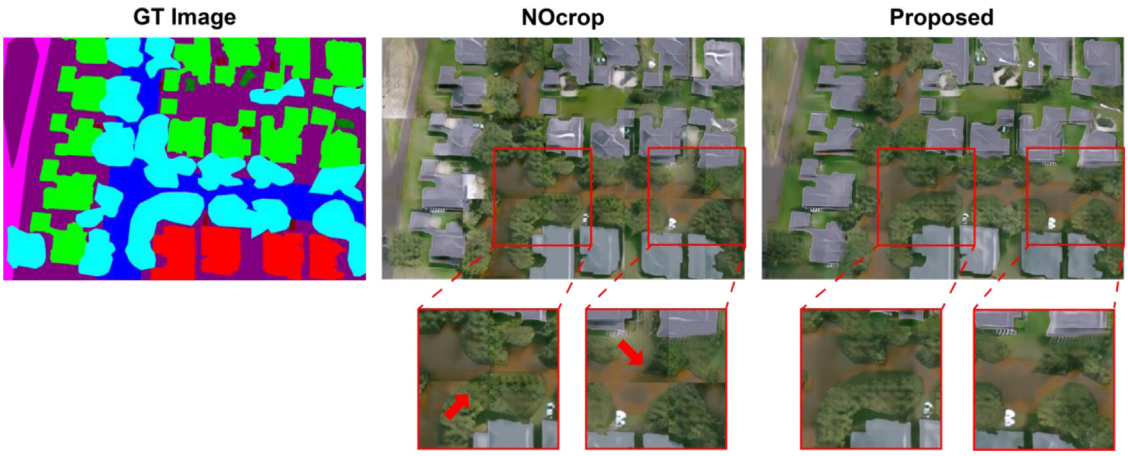



**Fig. 9.** Application of the proposed parametric kernel to satellite imagery synthesis. Input semantic annotation map from the FloodNet dataset (Legend: blue — road flooded; purple — road non flooded; red — building flooded; green — building non flooded; cyan: tree). NOcrop strategy exhibits visible checkerboard artifacts at patch boundaries (red arrows), while our proposed kernel configuration (ZEROX+LINY) effectively eliminates these artifacts while maintaining structural integrity.

## References


Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B., 2020. MedGAN: Medical image translation using GANs. Comput. Med. Imaging Graph. 79, 101684. http://dx.doi.org/10.1016/j.compmedimag.2019.101684.

Dalmaz, O., Yurt, M., Cukur, T., 2022. ResViT: Residual vision transformers for multimodal medical image synthesis. IEEE Trans. Med. Imaging 41, 2598–2614. http://dx.doi.org/10.1109/TMI.2022.3167808.

Guan, H., Liu, M., 2022. Domain adaptation for medical image analysis: A survey. IEEE Trans. Biomed. Eng. 69, 1173–1185. http://dx.doi.org/10.1109/TBME.2021.3117407.

Hagen, G.M., Bendesky, J., Machado, R., Nguyen, T.A., Kumar, T., Ventura, J., 2021. Fluorescence microscopy datasets for training deep neural networks. Gigascience 10, http://dx.doi.org/10.1093/gigascience/giab032.

Haji-Ali, M., Balakrishnan, G., Ordonez, V., 2024. Elasticdiffusion: Training-free arbitrary size image generation through global-local content separation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6603–6612.

He, Y., Liu, Z., Qi, M., Ding, S., Zhang, P., Song, F., Ma, C., Wu, H., Cai, R., Feng, Y., Zhang, H., Zhang, T., Zhang, G., 2024. PST-diff: Achieving high-consistency stain transfer by diffusion models with pathological and structural constraints. IEEE Trans. Med. Imaging 43, 3634–3647. http://dx.doi.org/10.1109/TMI.2024.3430825.

Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings -30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. Institute of Electrical and Electronics Engineers Inc., pp. 5967–5976.

Jewsbury, R., Wang, R., Bhalerao, A., Rajpoot, N., Vu, Q.D., 2024. StainFuser: Controlling diffusion for faster neural style transfer in multi-gigapixel histology images. arXiv Prepr. arXiv:2403.09302.

Jiang, J., Prodduturi, N., Chen, D., Gu, Q., Flotte, T., Feng, Q., Hart, S., 2020. Image-to-image translation for automatic ink removal in whole slide images. J. Med. Imaging 7, 057502. http://dx.doi.org/10.1117/1.jmi.7.5.057502.

Liu, S., Liu, K., Margolis, S., Wu, W., Knezevich, S.R., Elder, D.E., Eguchi, M.M., Elmore, J.G., Shapiro, L.G., 2025. Generating seamless virtual immunohistochemical whole slide images with content and color consistency. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging. ISBI, pp. 1–5. http://dx.doi.org/10.1109/ISBI60581.2025.10981306.

Liu, S., Zhu, C., Xu, F., Jia, X., Shi, Z., Jin, M., 2022. BCI: Breast cancer immunohistochemical image generation through pyramid Pix2pix. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. pp. 1814–1823.

Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z., Smolley, S.P., 2017. Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. Institute of Electrical and Electronics Engineers Inc., pp. 2813–2821. http://dx.doi.org/10.1109/ICCV.2017.304.

Meiburger, K.M., Marzola, F., Zahnd, G., Faita, F., Loizou, C.P., Lainé, N., Carvalho, C., Steinman, D.A., Gibello, L., Bruno, R.M., Clarenbach, R., Francesconi, M., Nicolaides, A.N., Liebgott, H., Campilho, A., Ghotbi, R., Kyriacou, E., Navab, N., Griffin, M., Panayiotou, A.G., Gherardini, R., Varetto, G., Bianchini, E., Pattichis, C.S., Ghiadoni, L., Rouco, J., Orkisz, M., Molinari, F., 2022. Carotid ultrasound boundary study (CUBS): Technical considerations on an open multicenter analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal B-mode ultrasound scans. Comput. Biol. Med. 144, 105333. http://dx.doi.org/10.1016/j.compbiomed.2022.105333.

Meiburger, K.M., Zahnd, G., Faita, F., Loizou, C.P., Carvalho, C., Steinman, D.A., Gibello, L., Bruno, R.M., Marzola, F., Clarenbach, R., Francesconi, M., Nicolaides, A.N., Campilho, A., Ghotbi, R., Kyriacou, E., Navab, N., Griffin, M., Panayiotou, A.G., Gherardini, R., Varetto, G., Bianchini, E., Pattichis, C.S., Ghiadoni, L., Rouco, J., Molinari, F., 2021. Carotid ultrasound boundary study (CUBS): An open multicenter analysis of computerized intima–media thickness measurement systems and their clinical impact. Ultrasound Med. Biol. 47, 2442–2455. http://dx.doi.org/10.1016/j.ultrasmedbio.2021.03.022.

Mittal, A., Moorthy, A.K., Bovik, A.C., 2012. No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21, 4695–4708. http://dx.doi.org/10.1109/TIP.2012.2214050.

Mittal, A., Soundararajan, R., Bovik, A.C., 2013. Making a completely blind image quality analyzer. IEEE Signal Process. Lett. 20, 209–212. http://dx.doi.org/10.1109/LSP.2012.2227726.

Müller, D., Kramer, F., 2021. MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. BMC Med. Imaging 21, 12. http://dx.doi.org/10.1186/s12880-020-00543-7.

Oktay, O., Schlemper, J., Folgoc, L. Le, Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D., 2018. Attention U-net: Learning where to look for the pancreas. arXiv Prepr. arXiv:1804.03999.

Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., Murphy, R.R., 2021. FloodNet: A high resolution aerial imagery dataset for post flood scene understanding. IEEE Access 9, 89644–89654. http://dx.doi.org/10.1109/ACCESS.2021.3090981.

Salvi, M., Branciforti, F., Molinari, F., Meiburger, K.M., 2024. Generative models for color normalization in digital pathology and dermatology: Advancing the learning paradigm. Expert Syst. Appl. 245, 123105. http://dx.doi.org/10.1016/j.eswa.2023.123105.

Salvi, M., Branciforti, F., Veronese, F., Zavattaro, E., Tarantino, V., Savoia, P., Meiburger, K.M., 2022. DermoCC-GAN: A new approach for standardizing dermatological images using generative adversarial networks. Comput. Methods Programs Biomed. 225, 107040. http://dx.doi.org/10.1016/j.cmpb.2022.107040.

Salvi, M., Manini, C., López, J.I., Fenoglio, D., Molinari, F., 2023. Deep learning approach for accurate prostate cancer identification and stratification using combined immunostaining of cytokeratin, p63, and racemase. Comput. Med. Imaging Graph. 109, 102288. http://dx.doi.org/10.1016/j.compmedimag.2023.102288.

Salvi, M., Mogetta, A., Gambella, A., Molinaro, L., Barreca, A., Papotti, M., Molinari, F., 2021. Automated assessment of glomerulosclerosis and tubular atrophy using deep learning. Comput. Med. Imaging Graph. 90, 101930. http://dx.doi.org/10.1016/j.compmedimag.2021.101930.

Sheikh, H.R., Bovik, A.C., 2006. Image information and visual quality. IEEE Trans. Image Process. 15, 430–444. http://dx.doi.org/10.1109/TIP.2005.859378.

Song, J., Meng, C., Ermon, S., 2021. Denoising diffusion implicit models. In: ICLR 2021-9th International Conference on Learning Representations. International Conference on Learning Representations, ICLR.

Stojanovski, D., Hermida, U., Lamata, P., Beqiri, A., Gomez, A., 2023. Echo from noise: Synthetic ultrasound image generation using diffusion models for real image segmentation. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Science and Business Media Deutschland GmbH, pp. 34–43. http://dx.doi.org/10.1007/978-3-031-44521-7_4.

Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S., 2015. Blind image quality evaluation using perception based features. In: 2015 Twenty First National Conference on Communications. NCC, pp. 1–6. http://dx.doi.org/10.1109/NCC.2015.7084843.

Wang, W., Bao, J., Zhou, W., Chen, Dongdong, Chen, Dong, Yuan, L., Li, H., 2022. Semantic image synthesis via diffusion models. arXiv Prepr. arXiv:2207.00050.

Wu, M., 2019. Wavelet transform based on meyer algorithm for image edge and blocking artifact reduction. Inf. Sci. (Ny) 474, 125–135. http://dx.doi.org/10.1016/j.ins.2018.09.058.

Wu, W., 2021. Patchify: A library that helps you split image into small, overlappable patches, and merge patches into original image.

Xu, Y., Hu, S., Du, Y., 2022. Research on optimization scheme for blocking artifacts after patch-based medical image reconstruction. Comput. Math. Methods Med. 2022, http://dx.doi.org/10.1155/2022/2177159.