Engineering Applications of Artificial Intelligence 159 (2025) 111763

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Generative adversarial network-based prediction of microhole profile drilled with high-energy electron beam on silicon wafer

Hyunmin Park [a,b], Jun Goo Kang [c], Jin Seok Kim [c], Eun Goo Kang [c], Seung-Kyum Choi [d], Hyung Wook Park [a,*]

[a] Department of Mechanical Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan, 44919, Republic of Korea
[b] School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, 47907, IN, USA
[c] Digital Transformation R&D Department, Korea Institute of Industrial Technology, 143 Hanggaulro, Sangrok-gu, Ansan-si, Gyeonggi-do, 15588, Republic of Korea
[d] George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, North Avenue, Atlanta, GA, 30332, USA

ARTICLE INFO

Keywords:
High-energy electron beam drilling
Microhole on silicon wafer
Generative adversarial network
Embedding a self-attention mechanism with multiple parallel heads

ABSTRACT

Drilling with a high-energy electron beam on a semiconductive ceramic substrate is emerging as an effective solution for creating high-aspect-ratio microholes. This method effortlessly surpasses band gaps and facilitates machining with continuous irradiation. However, the inherent brittleness and crystallinity of the semiconductive ceramic substrate hinder handling for quality analysis of the drilled substrate. From experimental drilling, the unique deformation history of the microhole with a high-energy electron beam hints at the potential for predictive modeling for non-destructive analysis of the microhole. In this study, we proposed a conditional generative adversarial network model for the non-destructive analysis of drilled microholes. We collected a limited number of images of hole inlets and cross-sectional holes for training and testing the network model. To effectively build the predictive model with our limited dataset, we introduced a generative adversarial network architecture with embedding a self-attention mechanism with multiple parallel heads. This architecture combines the advantages of the convolutional neural networks and the self-attention mechanism. The proposed architecture showed improvements in training loss and evaluation for image generation compared to the original convolutional neural network. The predictive precision for the inlet diameter, hole straightness, and drilled depth was enhanced by 11.6 %, 8.3 %, and 15 %, respectively. The maximum improvement in predictive accuracy for the inlet diameter, hole straightness, and drilled depth was 36.2 %, 22.98 %, and 58.6 %, respectively. These results indicate that the proposed model not only generated the geometrical profile of the cross-sectional hole but also accurately predicted geometrical dimensions.

1. Introduction

High-purity ceramic materials have become indispensable for the production of electronic devices and sensors in the 21st century. Among these materials, silicon carbide (SiC) wafers are particularly valuable due to their diverse applications, including chips, memory devices, transistors, and transducers based on piezoelectric, electrostatic, and electromagnetic energy sources (Shen et al., 2016), (Xu et al., 2021). Efficient implementation of SiC applications necessitates microcircuit patterning and enhanced circuit integration, with micro-hole drilling being a fundamental process (Khuat et al., 2014), (Liu et al., 2016). Specifically, Through Silicon Via (TSV) manufacturing, a key process in chip packaging, requires optimized microhole arrays with high aspect ratios (Kumar et al., 2021; Guo et al., 2022; Wang, 2019). However, conventional drilling processes have shown limited potential in microhole manufacturing due to the high hardness and brittleness of SiC wafers. Traditional removal methods on silicon wafers often result in chipping and crack propagation at the drilling site. To mitigate these issues, nonconventional drilling techniques have emerged as novel solutions for microhole drilling on silicon wafers. Techniques such as plasma drilling (Ozgur and Huff, 2017), vibration-assisted drilling (Kurniawan et al., 2024), electrical discharge machining (Guan and Zhao, 2022), and laser drilling have been utilized (Kar et al., 2024).

However, even laser drilling, the most commercialized and

* Corresponding author. Department of Mechanical Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulasn, 44919, Republic of Korea.
E-mail address: hwpark@unist.ac.kr (H.W. Park).

https://doi.org/10.1016/j.engappai.2025.111763
Received 9 March 2025; Received in revised form 4 June 2025; Accepted 10 July 2025
Available online 21 July 2025

developed method among nonconventional drilling technologies, has its drawbacks despite its advantages. Laser drilling on silicon wafers requires high-energy infrared wavelengths to overcome the bandgap of the silicon wafer surface (Nasrollahi et al., 2020), (Singh and Samuel, 2023). Because the single-crystal, mirror-like surface of silicon wafers can easily reflect and dissipate photons from lasers with short wavelengths, ultrashort pulsed lasers with high energy intensity are necessary to effectively drill through silicon wafers (Li et al., 2021). Moreover, the implementation of high-energy and repetitive penetration using ultrashort pulses can lead to increased setup costs for equipment configuration and higher electrical energy consumption for drilling a single hole.

Electron beam (EB) drilling has been recognized as an innovative technology that offers superior performance compared to existing nonconventional microhole drilling methods, potentially serving as an advanced alternative (Kim et al., 2016), (Howitt et al., 2008). The high-energy EB can ablate and penetrate industrial ceramic materials irrespective of their electrical properties and light reflectance. Compared to ultrashort pulsed lasers, the high-energy EB achieves greater penetration depth in both metals and ceramic materials. High-aspect-ratio microholes can be easily obtained not only in conductive metals but also in single-crystal wafers with semiconductive properties. Additionally, the penetration capability of the EB enables the drilling of up to 100,000 holes per second (Kim et al., 2016), (Leitz et al., 2012). Furthermore, the formation of a vacuum atmosphere is essential in EB drilling, as the substrate is drilled within a vacuum chamber. This environment prevents oxidation or contamination on the surface and inside the hole, highlighting the potential of EB drilling as a green manufacturing process. Previous studies on EB drilling for semiconductive materials have shown that effective drilling can be achieved with long pulse durations or continuous irradiation (Howitt et al., 2008), (Kizuka and Tanaka, 1997). It has been demonstrated that the penetration depth of accelerated electrons is excellent even in semiconductor materials with a wide bandgap, and the mechanisms of EB penetration into the insulating layer of silicon wafers have been identified (Howitt et al., 2008). EB irradiation can traverse from the conduction band to the valence band, reaching the bottom of the material with ease. Due to its inherent penetration capabilities, EB drilling operates effectively without requiring precise pulse timing (Kim et al., 2016), (Leitz et al., 2012).

Even with advanced high-aspect-ratio drilling technologies, microhole analysis in SiC wafers continues to pose significant challenges. Destructive hole analysis, the most general and definitive method to verify microhole quality, necessitates cutting the cross-section of the hole. The destructive analysis of SiC wafers always encounters difficulties due to the material's characteristics. The specific crystal orientation of SiC makes it tough to cut selectively in the desired direction (Sekhar et al., 2020). Additionally, high-purity silicon wafers are considerably more fragile and susceptible to breakage compared to other ceramic materials (Popovich et al., 2013). As the thickness of the SiC wafer decreases, handling it before and after EB drilling requires increased caution. Direct analysis of drilled silicon wafers by human hands or physical tools, even if successful in acquiring cross-sections, can easily lead to unintended defects such as secondary cracks or damage (Cook, 2006), (Liu et al., 2022).

To replace destructive analysis for vulnerable and fragile substrates, non-destructive analysis methods have been preferred, and related studies are continually ongoing. In conventional non-destructive analysis of silicon wafers, non-contact methods such as infrared laser scanning (Cela et al., 2023), X-ray inspection (Chen et al., 2023a), and scanning electron microscopy (SEM) are used (Chen et al., 2022). For non-destructive internal analysis of silicon wafers, SEM analysis is most preferred. SEM imaging not only offers the highest resolution and sharpness but also provides more intuitive graphical information compared to other methods. However, in the case of non-destructive analysis based on SEM images, examinations are often limited to surface analysis, and verifying the quality of microholes ultimately requires disassembling the drilled substrate. Non-destructive analysis methods typically involve much higher setup costs compared to destructive methods, and the time required for measurement is substantial. Additionally, precise and accurate measurements often require the assistance of skilled operators.

Recently, Generative Adversarial Network (GAN) models, based on deep learning of image data, have rapidly emerged as a novel methodology with the potential to serve as a fast and accurate non-destructive analysis model. GANs excel in generating both image and numerical data and can produce reliable results even with sparse datasets that are challenging to collect (Liu et al., 2021; Mumbelli et al., 2023; Zhang et al., 2021). Given the industrial trend of steadily increasing demand and scale for silicon wafer-based applications, there are compelling reasons to consider adopting GAN-based analysis.

When the generative images from GAN models are established as grayscale SEM images by learning the causal relationships in silicon wafer drilling, GAN models can be effectively utilized as non-destructive analysis tools. Although some studies have demonstrated successful GAN-based analysis using SEM images, these studies have typically employed thousands of SEM images (Wang et al., 2021), (Du and Shi, 2020). Training and optimizing GAN models with a limited amount of image data remains challenging due to the time-consuming nature of data collection and preparation for deep learning models. Furthermore, no AI-based studies have specifically focused on analyzing microhole quality, highlighting a gap in the application of GAN models for non-destructive analysis.

On the other hand, while GAN models have been used for quality analysis in some cases of laser machining on silicon wafers, most of these efforts have not focused on generating SEM images (Mills et al., 2018; Heath et al., 2018; McDonnell et al., 2021). In contrast, there has been limited research applying GAN models to the analysis of electron beam (EB) drilling. Furthermore, no studies have yet employed SEM image-based GAN models for analyzing microholes created by high-energy EB drilling.

This study proposes and implements predictive modeling based on a conditional GAN (c-GAN) for assessing the geometrical qualities of microholes drilled with high-energy electron beams (EB). Through drilling experiments involving high-energy EB irradiation with key process parameters, the geometric profile and characteristics of the hole inlet and cross-sections were analyzed. Notably, the hole inlet exhibits deformation history due to the vacuum atmosphere during drilling. Colorful optical images of the hole inlet are assumed to provide rich features for deep learning architectures used in c-GANs. Consequently, using the Pix2Pix model, a well-established c-GAN framework, a predictive model for non-destructive analysis was developed. The objective of the proposed model is to predict SEM images of the hole cross-section using only the hole inlet image as input. To achieve this, a Multi-head Attention module was integrated into the U-Net architecture, resulting in a GAN model with U-Net and Embedded Multi-head Attention (UNEMA). The UNEMA GAN was optimized by balancing the learning rates of the generator and discriminator to maximize training and test performance. After testing the model, experimental validations were conducted not only to evaluate generative performance but also to assess geometric hole qualities.

2. High-energy electron beam drilling on thin SiC wafer

High-energy electron beams drilled SiC wafers by concentrating accelerated electrons onto a microscale spot. To leverage the benefits of high-energy electron beam (EB) drilling—such as full substrate penetration, high-aspect ratio drilling, and ultra-high machining speed—advanced systems for EB irradiation and parameter configuration were required. Using equipment capable of focused irradiation with high-energy-density electrons, experimental drilling of microholes was conducted under various key process parameters. Additionally, strategies for preparing a training dataset for developing a non-destructive

analysis model were proposed.

2.1. Experimental setup of high-energy electron beam drilling

For the setup of experimental conditions, three key process parameters closely related to the irradiation mechanism and energy intensity of the high-energy electron beam were considered: beam current, solenoid coil (SC) current, and irradiation time (Kim et al., 2016), (Kang et al., 2023). Thermally emitted electrons are released from the cathode filament in random directions. The acceleration voltage causes these electrons to propagate along the cathode-anode axis, with the number of emitted electrons being proportional to the beam current. Higher acceleration voltage and beam current result in increased beam intensity of the irradiated electron beam (Kim et al., 2018).

Fig. 1 illustrates the schematics of high-energy electron beam (EB) irradiation within the gun. The electron beam is emitted from the cathode filament and initially focused by the upper solenoid coil, which aligns the beam. After passing through the upper solenoid coil, the Rogowski coil provides an indirect measurement of the electron beam's stability and linearity in response to the irradiation sequence (Moreno et al., 2017). The stigmator, located below the Rogowski coil, corrects for any astigmatism and enhances the circularity of the EB spot on the substrate (Goswami, 2020). The geometrical profile of the high-energy EB follows a Gaussian distribution.

When the EB reaches the lower solenoid coil, this coil directs the accelerated electrons to form the EB spot on the material's surface. The effective radius refers to the radius of the EB spot on the substrate's top surface. Minimizing the effective radius is crucial for enhancing beam intensity (Park et al., 2022). Under constant beam current and acceleration voltage, a narrower effective radius increases the intensity of the EB spot, which in turn improves the penetration of the electron beam (Derby and Olbert, 2010; Bakeev et al., 2019; Egerton, 2005). The lower solenoid coil functions more effectively as an electromagnetic lens for the EB compared to the upper solenoid coil. Consequently, the experimental EB drilling focused specifically on adjusting the current of the lower solenoid coil.

Additionally, it was important to consider the time required for complete penetration when the energy intensity is transferred to the substrate. The EB irradiation time was identified as a crucial parameter

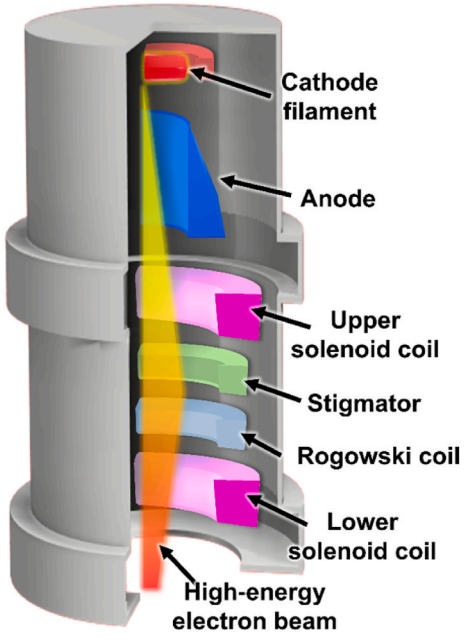

Fig. 1. Schematics of gun structure and irradiation of high-energy electron beam.

in EB drilling. This irradiation time, measured in microseconds, refers to the duration of the continuous beam, not the pulsed beam. Table 1 outlines the major experimental parameters for SiC wafer drilling. In Table 1, solenoid coil refers specifically to the lower solenoid coil. The current of the solenoid coil was adjusted by increasing or decreasing it relative to the reference current. When ΔI_{SC} was set to zero, I_{SC} matched the reference current of the solenoid coil.

A total of 60 experimental conditions were applied to the high-energy electron beam (EB) drilling by combining five ΔI_{SC} conditions, four I_{beam} conditions, and three t conditions. For each experimental condition, a 4 × 15-hole array was drilled, resulting in a total of 60 EB irradiations. The hole pitch, or the distance between consecutive irradiations, was set to 1 mm, which was deemed sufficient to prevent thermal interference between adjacent irradiations.

For the EB drilling, a 0.8 mm thick silicon carbide (SiC) wafer was used as the substrate. The SiC wafer primarily consists of over 99 % silicon and is passivated with silicon nitride (SiN). The surface of the SiC wafer is mirror-like, with high photon reflectivity and insulating properties. The high-energy EB can penetrate the insulating SiN layer through collisions driven by the kinetic energy of the accelerated electrons. Once the EB passes through the insulating layer via atomic displacement mechanisms (Howitt et al., 2008), it readily reaches the conductive layer. In the conductive layer, the EB acts as a heat source, sufficient to induce phase changes in the silicon wafer based on its melting and vaporization points. Table 2 provides the material properties of the SiC wafer.

2.2. Image preparation for hole inlet and cross-sectional hole

The deep learning model for non-destructive analysis aimed at image translation from hole inlet as input image to cross-sectional hole as output image. From the experimental results for the high-energy EB drilling, the image data of the hole inlet and cross-sectional hole drilled under 60 experimental conditions were individually obtained and preprocessed.

During EB penetration on substrate, recoil and Marangoni pressures are activated between path of EB and the substrate around the EB. Pressure-affected melts (or gas) eject in the opposite direction of the irradiation. Immediately, gravitational force causes the ejected melts (or gas) to sink around the entrance of the EB on the substrate (Park et al., 2022). Such ejection induced depositions in the form of Si_xN_y around the hole inlet, showing a variety of colors. The color around the hole inlet was changed by degree of deposition thickness on the substrate surface. The experimental results showed that the surface changed the reflectance of visible light depending on the thickness of Si_xN_y deposition. It was estimated that Si_xN_y created by EB drilling was deposited in a range of about 200–400 nm (Henrie et al., 2004). Drilled hole inlets were obtained by optical microscopy (OM, VHX-7000, Keyence) as 2180 x 2660 x 3 RGB images. It was justified to use an OM to capture the degree of coloring on the SiC surface. There was no post processing to remove the deposited colors for investigating the trend of color deposition. Figs. 2(c) and 3(a) demonstrates the hole inlet images depending on the experimental conditions.

In contrast to acquisition of the hole inlet images, obtaining cross-sectional hole images from the drilled SiC wafer was a much more difficult task. To obtain the images of cross-sectional hole without

Table 1
Experimental conditions based on major process parameters.

Experimental parameter	Value
Acceleration voltage (V_{acc}) [keV]	130 (fixed)
Reference current of solenoid coil [mA]	2740 (fixed)
Change of solenoid coil current (ΔI_{SC}) [mA]	−30/−15/0/+15/+30
Beam current (I_{beam}) [mA]	20/30/40/50
Irradiation time (t) [μs]	20/40/80

Table 2
Material properties of silicon carbide wafer.

Material properties	Value (or range)
Density [g/cm^3]	2.33-3.1
Specific heat [J/kg/K]	700–750
Melting temperature [°C]	1410
Vaporization temperature [°C]	2355–2900
Thermal conductivity [W/m/K]	130–150

disassembling the 0.8 T SiC wafer, manual cutting with a diamond cutter was performed, and manual polishing was conducted to distinguish between the drilled area by the EB and the other areas which were not removed in the SiC wafer. Since the removed (or penetrated) area by EB irradiation composed with a re-solidificated Si layer without Si_xN_y deposition, the inner wall of the hole had a high light reflectivity. Observing a very gleaming surface of the cross-sectional hole with an OM was insufficient. Accordingly, cross sections of the drilled hole were captured by Scanning Electron Microscopy (SEM, Nano230 FE-SEM, Thermofisher). Obtained images were saved as 1024 x 960 x 3 RGB images.

2.3. Preprocessing of training/test data for deep learning model

To train/test a non-destructive analysis model that can predict the corresponding shape of a cross-sectional hole using only hole inlet, image data obtained by OM and SEM were organized into a training

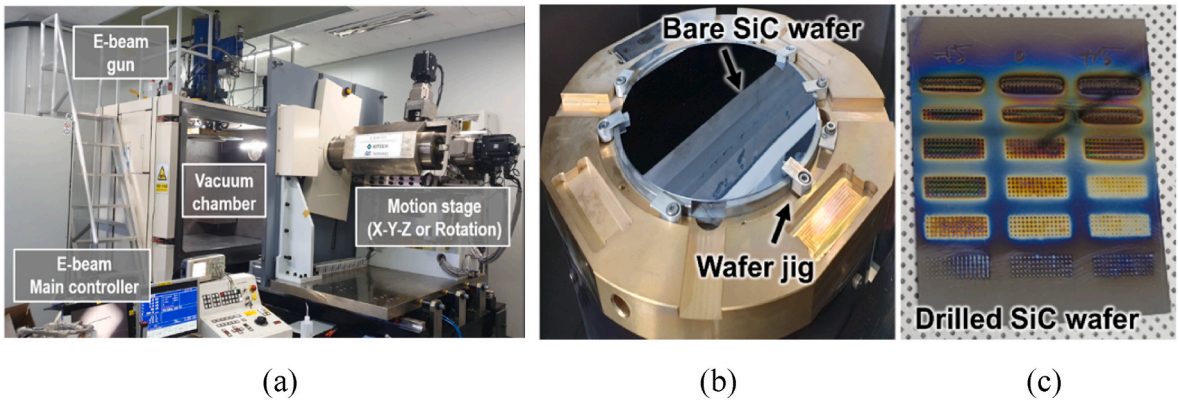

Fig. 2. (a) equipment of high-energy electron beam, (b) experimental setup, and (c) silicon wafer substrate drilled with high-energy electron beam.

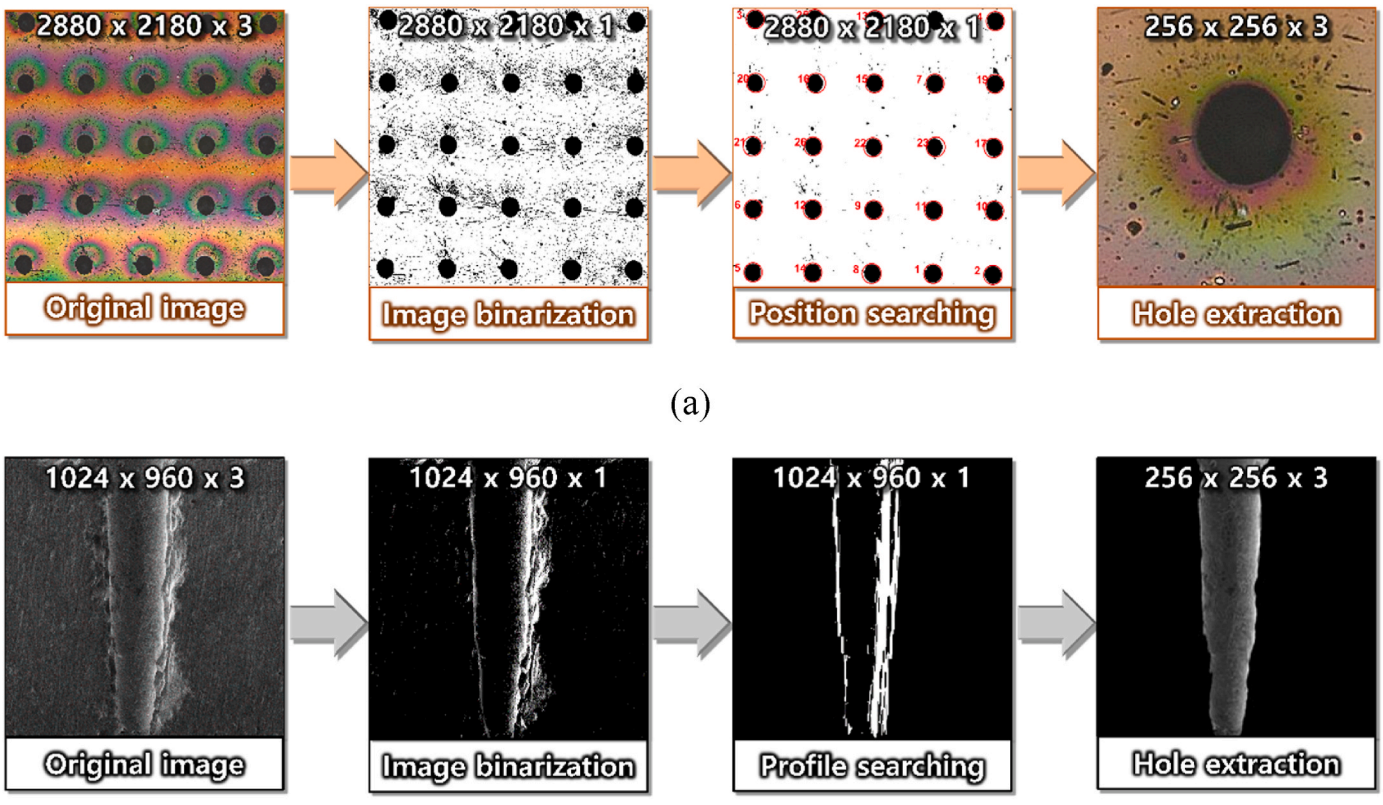

Fig. 3. Gathering and preprocessing of image for (a) hole inlet, and (b) cross-sectional hole.

dataset and a test dataset. From the perspective of the deep learning model, the hole inlet image was set as the input image and the cross-sectional hole image as the target image. The inlet and cross-sectional images for four holes per experimental condition were collected. For 60 experimental conditions, 240 holes inlet images and 240 cross-sectional hole images were prepared. Finally, a total of 480 image data were composed in dataset. Each hole inlet image was paired and labeled 1:1 with its corresponding cross-sectional hole image drilled under the same EB drilling conditions. Thus, a total 240 paired dataset of input and target images were obtained. It was difficult to say that 240 paired image data were sufficiently large in building a deep learning model. In addition, each pair of image data had a lot of noise, and the image size of the hole inlet (2180 x 2660 x 3) and the cross-section of hole (1024 x 960 x 3) was different, "which could negatively affect training and test. To promote training/test of the deep learning model more efficiently, individual preprocessing was carried out on the images of hole inlet and cross-sectional hole.

Fig. 3(a) and (b) represent the preprocessing of input and target images, respectively. The original input images included hole inlet array. The hole inlet array was split into single hole inlets and cropped to 256 x 256 x 3 RGB images. At that point, a circular Hough transformation mechanism, which implemented with MATLAB software (Caya et al., 2019), was used to catch the center point of the circular hole inlet and to position center point of the hole on central pixel indices (128,128) of the cropped image. All hole-inlet images were processed to include hole inlet area (black pixel) and a circumferential colored deposition simultaneously within a size of 256 x 256. On the other hand, in the image of cross-sectional hole, only the inner wall of the drilled hole was set as Area Of Interest (AOI). The rationale of the AOI setting was not only to provide the model users with geometrical hole profile, but also to prevent other areas outside of drilled area from being recognized as noises or unnecessary features to the deep learning architecture. Edge detection of the inner surface of the drilled hole was performed using the Prewitt algorithm, which was also conducted in MATLAB (Sinitca et al., 2023). And all pixels outside of the AOI were converted with black colored pixel. The cross-sectional hole images were resized as 256 x 256 x 3 RGB images to match the resized inlet images. Finally, in all preprocessed target image, the AOI was re-aligned so that the centerline of the AOI was placed in the 128th pixel column of the image. The preprocessing task for target image made the centerline of drilled hole located on the center of resized target images.

3. Predictive modeling based on conditional generative adversarial network

The images of the hole inlet and cross-sectional hole acquired experimentally, when considered as a pixel-based data matrix, were composed of visual features represented by a combination of numerous numerical data. These features involved the distribution of colors, shapes, brightness, and contrast. For efficiently processing the image data with such complicated pixelwise features, deep learning techniques for the construction of objective function can be an appropriate choice. Furthermore, the deposited color distribution around the hole inlet was considered and assumed that it can provide meaningful features to the deep learning model. Based on learning the correlation between the unique features and the drilled hole profile, a conditional Generative Adversarial Network (c-GAN) model that facilitates the conversion from hole inlet to cross-sectional hole was devised.

3.1. Pix2Pix model for prediction of cross-sectional-hole

Among various conditional GANs that perform image-to-image conversion, specific GAN models exist that do not merely alter the style or frame of the input image but rather transform it into an entirely different output image. As a deep learning model capable of executing extreme image translation from a multicolored circular hole inlet to a SEM-captured cross-sectional hole, it was confidently determined that Pix2Pix is the most suitable GAN model for this purpose. The Pix2pix model is one of the representative adversarial deep learning models in which generators and discriminator compete (Isola et al., 2017). The framework of Pix2Pix has been still used and applied due to its high flexibility and various application field.

With the paired images of input and target, generator of Pix2Pix serves to create a new generative cross-sectional image, indicated as fake image (Fig. 4). The generator received the input which was adjusted with size of 256 x 256 x 3 in the preprocessing step. The architecture on the Pix2Pix generator generally included U-Net architecture. The U-Net architecture is a deep learning structure in which represents a symmetrical structure through integration of encoder and decoder. In the encoder part of U-Net, convolution operations are performed continuously, and at the same time, each convolution layer is equipped with a skip connection that links a shortcut to the decoder. Such arrangement of skip connections preserves and references features from unconvolved images, which enhances the learning performance by maintaining important features that may be lost during the convolution process. U-Net architecture is known to draw high performance even in learning environments with a small amount of training data, and in the case of image generation, the U-Net has the advantage for bringing out pixel details in local pixel areas (Ronneberger et al., 2015). Detailed structure of the U-Net is shown in Fig. 5(a).

As a main structure of discriminator in Pix2Pix model, patch GAN was adopted, which divides the image into a patch unit smaller than the whole image scale and determine the authenticity between the patch and patch. The approach of patch GAN focuses on the local and spatial features of image pixels, making it effective for accurately detecting and mimicking local image details. The discriminator with the patch GAN calculated the distance between the target and generative image at the patch-unit level, and then evaluated the authenticity whether the given target image is real or fake (Chen et al., 2023b). It was anticipated that evaluating on a patch-by-patch basis would directly contribute to enhancing image-to-image conversion performance of the generator and the texture quality of the generative images.

Loss function between the generator and discriminator was adopted the Binary Cross Entropy (BCE) loss function, which is specialized in determining the cross entropy for the binary classes (real and fake) among the various type of loss functions related with cross entropy. The BCE loss function is a representative function that contributes to inducing the discriminator to judge the generative image as a real image. Furthermore, it is known that BCE has a high performance in preventing gradient vanishing at the construction of GAN model (Isola et al., 2017), (Hurtik et al., 2022). Additionally, the L1 loss function was adopted together to minimize the qualitative difference between the generative image and the target image. The L1 loss was calculated by summing the absolute value of the pixel difference between the generative image and the target image. It is known that L1 loss function, which linearly calculates pixel error values, is more specialized in modifying the image blurring of the generative image than L2 loss function (Isola et al., 2017), (Janocha and Czarnecki, 2017). On the L1 loss function, a regulation parameter (λ) for adjusting the importance of the L1 loss was multiplied. As the λ value increases, the generative image is more effectively induced to mimic the target image. At the step of parameter optimization for generator and discriminator, the ADAM optimizer was utilized. The ADAM optimizer showed the merit of maintaining the gradient in noise and nonlinearity in arithmetic operation with image data with high-dimensional pixels (Kingma and Ba, 2014). Detail specifications for training and test of Pix2Pix model are shown in Table 3.

3.2. Modification of U-net architecture by embedding multi-head attention

The U-Net architecture-based Pix2Pix model is well-suited for hole-inlet to hole-section image conversion given its intended purpose and

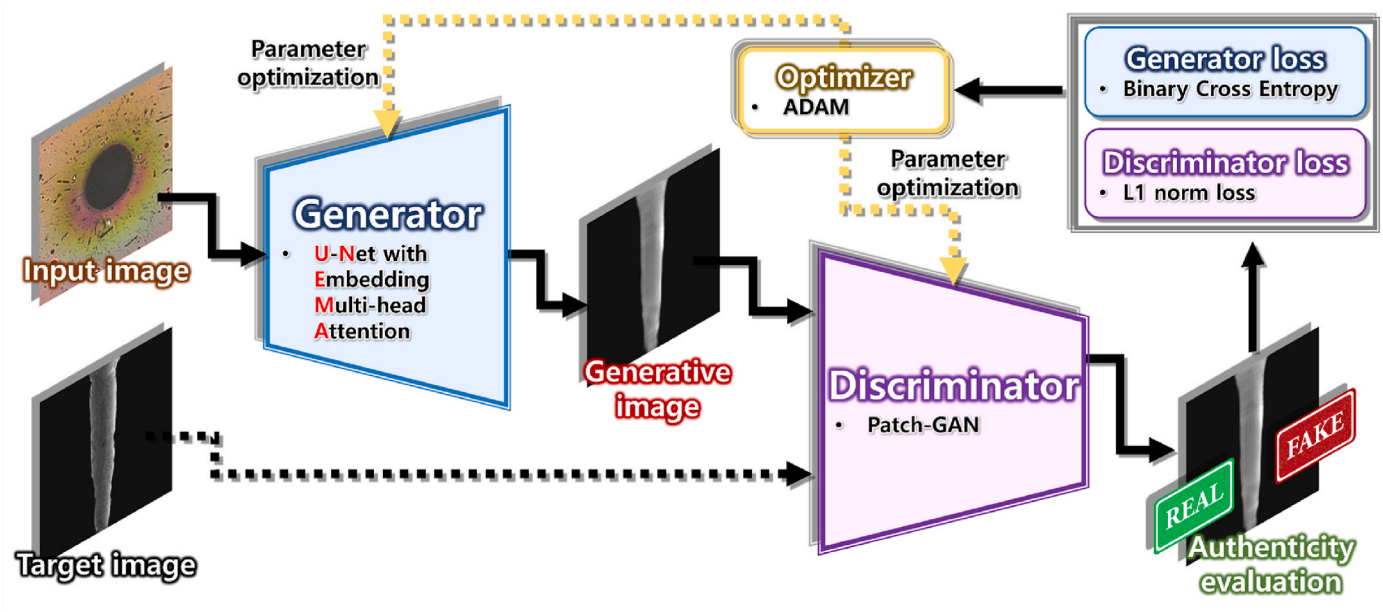

Fig. 4. Pix2Pix modeling for the prediction of cross-sectional hole.

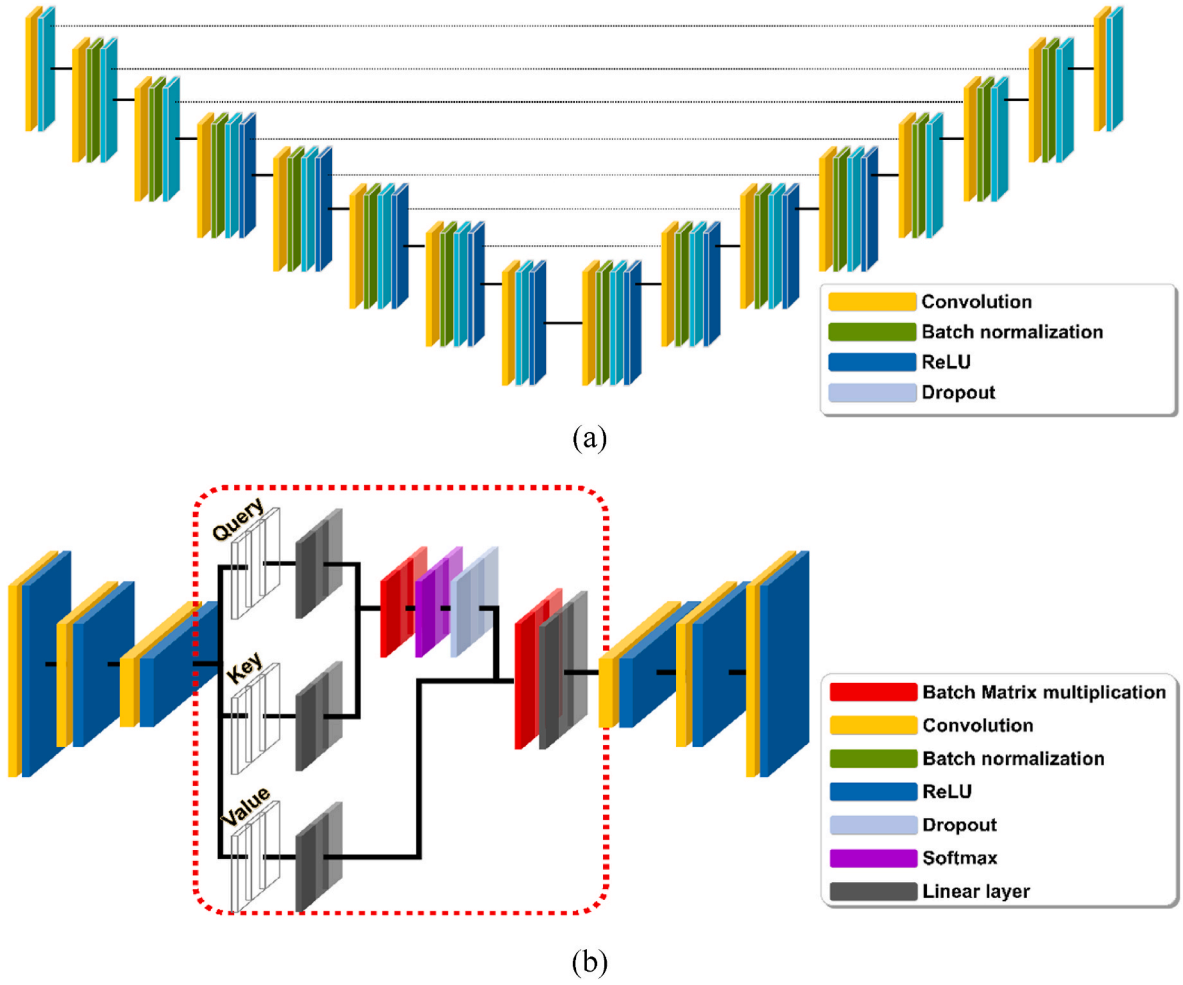

Fig. 5. The architecture of (a) U-Net (Ronneberger et al., 2015) and (b) multi-head attention (Vaswani et al., 2017).

Table 3
PC specifications for c-GAN training and test.

Component	Specification
CPU	Intel(R) Core (TM) i7-8700K @ 3.70 GHz
GPU	NVIDIA GeForce GTX 1050
RAM	16.0 GB
OS	Microsoft Window 10, 64-bit
Programming language	Python 3.8.18
Framework	Pytorch 2.2.1

benefits. However, inherent limitations of the model also exist. Although it has advantages in implementing local pixel details of generative image, it is vulnerable to reproducing the overall context and large-scale features (Sahli et al., 2023). When generating the geometric shape of the cross-sectional hole, the model may compromise the inferior reliability by generating output images that are dimensionally excessive or insufficient. Furthermore, since the prepared dataset included fewer than 500 images, these shortcomings could be further exacerbated. On such training/test environment of deep learning, an architecture of U-Net with Embedding Multi-head Attention (UNEMA) was devised and applied on the Pix2Pix model. The multi-head attention module is an innovative transformer structure, which is an application mechanism designed based on a self-attention mechanism (Vaswani et al., 2017). It has an excellent ability to grasp long-range dependencies and specializes in processing time-dependent sequence data or text data. Since the AOI of the microhole contains pixel-based features and has a vertically long pillar shape, it was assumed that the multi-head attention module can leverage its operational strength on learning over long-distance pixel area. And it was expected to perform simultaneous catching of feature tensors due to parallel iterations.

The individual self-attention head has three gates: query (Q), key (K), and value (V). The query represents the importance and dependency of the input's pixel-based features. The key captures the features of each pixel, enabling the calculation of their correlations. Subsequently, the key and query are used to measure their similarity. Inner product with the query and key derives the attention score based on that similarity. The value serves as the gate to ultimately transform the features of the input tensor and generate the final output tensor through a weighted sum based on the attention scores (Vaswani et al., 2017).

Fig. 5(b) represents the structure of a multi-head attention. The role of the linear layer (colored in dark gray) is to feature vectors and assists in the dimensional transformation of the tensor. It is essential when creating a tensor structure suitable for parallel operations. Batch matrix multiplication (BMM), originating from query and key, converts into an attention score matrix representing the similarity between query and key. The attention score is then normalized in the SoftMax layer (colored in purple). Finally, the value matrix and the normalized attention score are dot-produced to derive the self-attention value. By arranging a string of single module architecture multiplied as a parallel structure, which is shown like red-dot rectangle at Fig. 5(b), a multi-head attention architecture was embedded into the bottleneck part of U-Net architecture. Depending on the number of self-attention modules, the number of multi-heads can be determined, and parallel self-attention mechanism can be accordingly implemented. In the bottleneck part of UNEMA architecture, the feature tensor with the smallest dimension fed into a multi-head attention module. The multi-head attention module transformed the input feature tensor into a tensor that preserves important features, focusing on feature locations based on long-range pixel dependencies and relationships. After the feature transformation, the feature tensor was passed to the decoder section of the U-Net, which then output the final generative image.

4. Results and discussions

The typical C-GAN for image data requires over thousands of levels in training/test images. However, the UNEMA GAN was exposed to only 240 image data pairs, and the situation with a scarcity of data like this study can be commonly happened in real manufacturing field. Previous other studies about the GAN have already introduced several data augmentation techniques to compensate for the insufficient image dataset. However, none of the data augmentation techniques in the proposed modeling was applied. Before investigating the performance of proposed UNEMA GAN, characteristics of hole inlet and cross-sectional hole from experimental results were discussed. From the experimental results, training the Pix2Pix model including the UNEMA GAN was conducted, and generative images were obtained from test dataset. both a qualitative and quantitative evaluation of the generative image were conducted with test dataset.

4.1. Experimental results of high-energy electron beam drilling

The high-energy EB drilling produced distinctive results on the SiC wafer surface when microhole was drilled. In general, the lower beam current caused the depositions to be clearer blue and mustard yellow, while the higher beam current resulted in darker purple, pink, and orange colors (Henrie et al., 2004). When the value of beam current and irradiation time was constant, it was confirmed that distribution of surface coloring varied depending on the change of SC current and irradiation time. Similarly, the diameter of the hole inlet also varied depending on the current change of the solenoid coil, even if a constant beam current and irradiation time were adopted. Fig. 6 provides the geometrical changes of drilled hole from the OM images of hole inlet and SEM images of cross-sectional hole. The experimental results indicated that the manipulation of the solenoid coil current brings a significant difference in the geometrical hole shape. As the tendency of change in inlet diameter in Fig. 7(a), (b), and (c), the minimal diameter of hole inlet was found when the appropriate values of beam current and SC current were applied, rather than simply adopting the lower beam current or SC current values. Moreover, it can be identified that thermal cracks on the inner surface of drilled hole were minimized when EB irradiation was performed under EB conditions inducing the minimal diameter of hole inlet in Fig. 6.

Similarly to the trend of inlet diameter, the deep and full penetration of hole did not progress by simply increasing the beam current or the irradiation time. As shown in Fig. 6, not only the beam current and irradiation time, but also the SC current affect the geometrical shape of the drilled hole. When the beam current strengthened, initial focal point was placed deeper into the substrate. Controlling the SC current higher, focal length from lower solenoid coil to the focal point can be shortened, which means that a positive change in the SC current can make the focal point shift above the substrate surface (Kang et al., 2023). Additionally, when the EB with a higher beam current was irradiated, the focal point could move upward with only a small increment of SC current. As the focal point was placed on the top of the substrate with proper beam current and SC current, the deep penetration was conducted, resulting in a high-aspect ratio hole with the fine straightness. Based on the experimental trends of the hole inlet (Fig. 7(a–c)) and drilled depth (Fig. 7 (d–f)), it was determined that deep hole penetration occurs with the narrow hole inlet diameter. Over the experimental results, the minimum diameter of full-penetration hole was shown as 223 μm and corresponding aspect ratio of the hole was calculated 3.58.

4.2. Performance evaluation of training and test for predictive model

To investigate the learning performance of UNEMA GAN, various training parameters were adjusted, and the results were investigated. Among the training parameters, the major learning parameters that brought significant changes in results were the generator's learning rate and regularization factor (λ) respectively. While the Discriminator's Learning Rate (DLR) was fixed as 1e-7, the learning performance of each training trial was identified by multiplying (or dividing) the Generator's

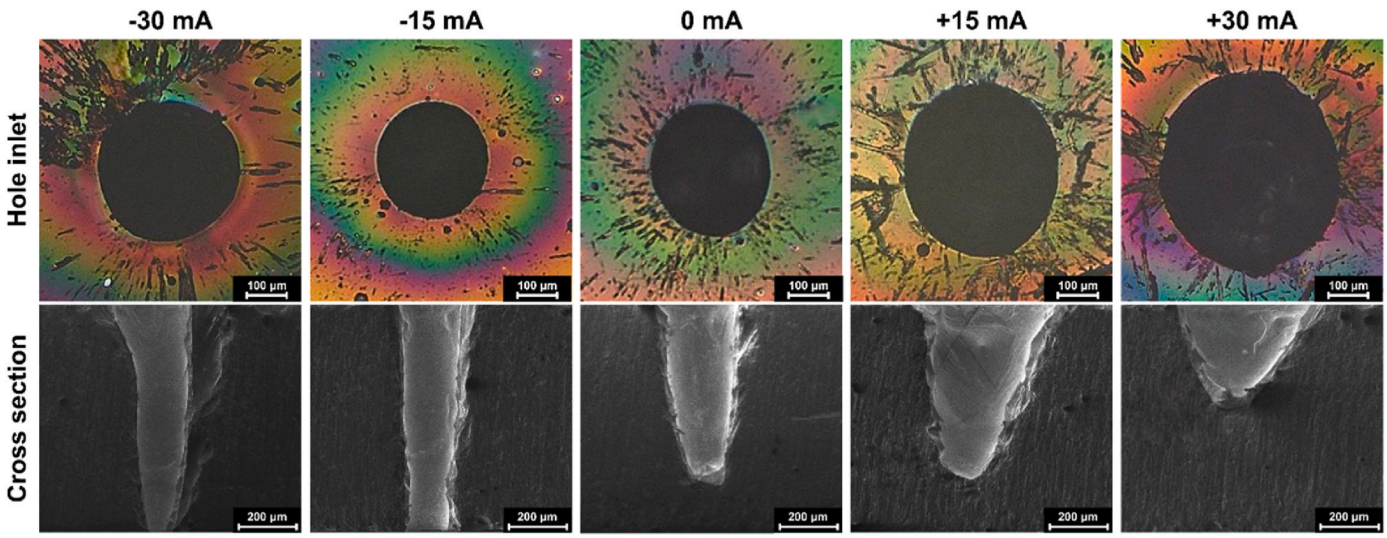

Fig. 6. Hole inlet and cross-sectional hole as changed by solenoid coil current in 50 mA beam current and 40 μs irradiation time.

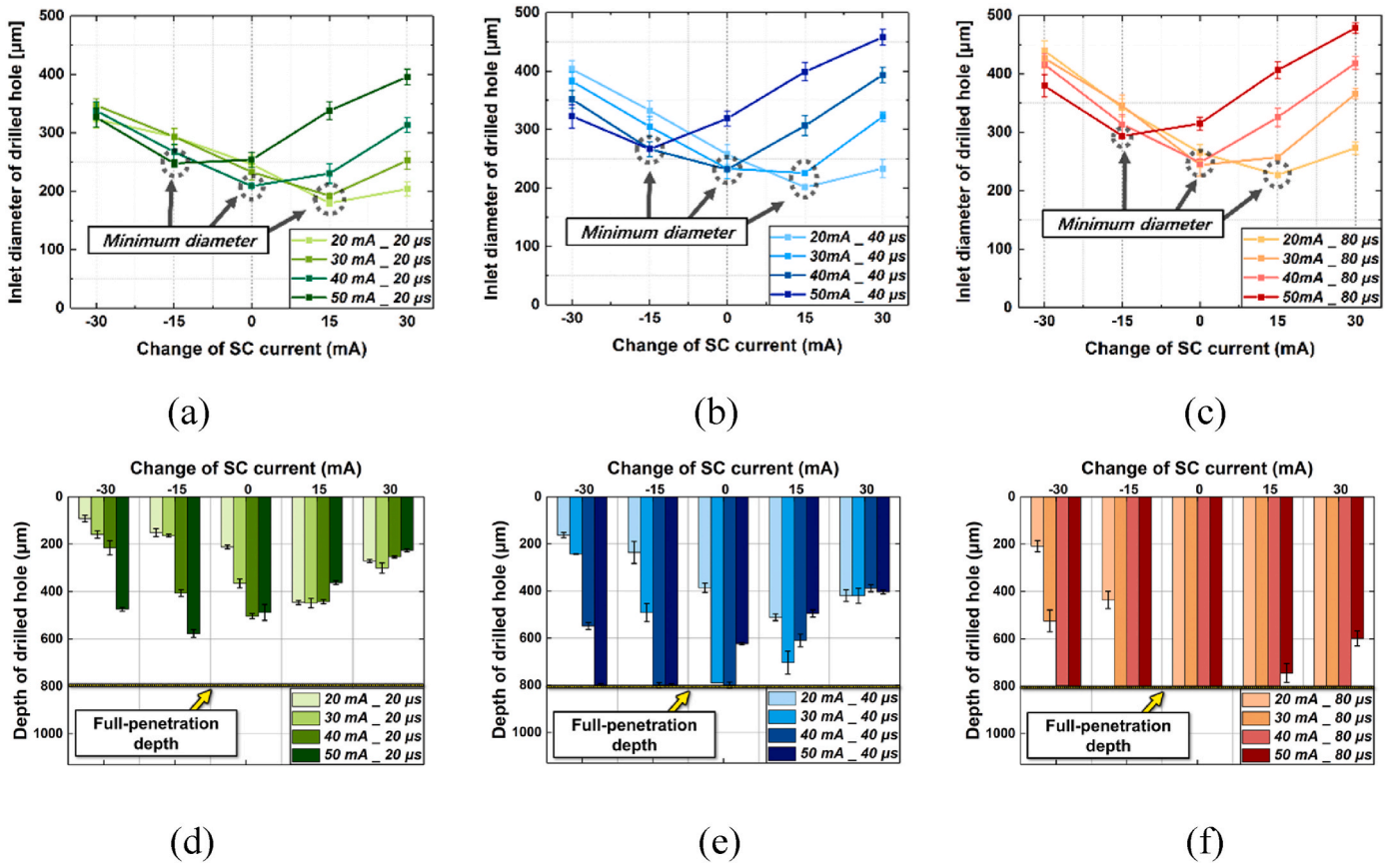

Fig. 7. Inlet diameter and depth of hole drilled in 40 μs irradiation time.

Learning Rate (GLR) by ten. The details of training parameters and conditions used in the training can be found in Table 4. All learning rates were applied as constant values for 50 epochs, without any extra scheduling or planning for the learning rate change. The 240 pre-processed image data were split per experimental condition at a 75:25 ratio to ensure balanced coverage, resulting in 180 image pairs for the training set and 60 for the test set. A condition-wise stratified splitting strategy was applied during the data partitioning process. Since each experimental condition yielded four inlet and cross-sectional images of the hole, this splitting ratio ensured that three samples were used for training and one for testing per condition. The training and test set included all experimental cases conducted in this study, and no image shuffling was applied in the dataset. During the 50 training epochs, the weight parameters of the epoch in which the best loss of BCE occurred were saved. In other words, the weights when the BCE loss was most optimized were finally adopted by the generator.

Based on the specific training parameters, it was quantitatively evaluated how significantly the performance of UNEMA GAN

Table 4
Training parameters for the Pix2Pix training.

Parameter	Generator	Discriminator
Architecture	UNEMA	Patch GAN
Learning rate	1e-2/1e-3/1e-4/1e-5	1e-7
Patch size	N/A	16 x 16 x 1
Number of attention head	8	N/A
Regularization factor (λ)	100/125/150	
Optimizer	ADAM	
First-order moment coefficient (β_1)	0.5	
Second-order moment coefficient (β_2)	0.999	

underwent modification in training/test compared to the Pix2Pix model with original U-Net architecture as a baseline model. Table 5 displays the best training loss and scores of evaluation criteria based on the training with regularization factors of 100, 125, and 150, respectively. The BCE loss in Table 5 means the minimum loss achieved during 50 epochs of training. Fréchet Inception Distance (FID) is a prominent evaluation criterion of image similarity that utilizes the InceptionV3 layer (Martin et al., 2018). The lower the FID, the more similar the generative image was to the target image. SSIM, Structural Similarity Index Measure, serves as another criterion indicating how structurally similar a target image is to a generative image. Higher values of SSIM denote better performance of image generation (Setiadi, 2021). PSNR stands for Peak Signal-to-Noise Ratio and is useful for evaluating the similarity of visual quality aspects such as image sharpness, distortion, and noise, as it represents the ratio of signal strength to noise level. Like SSIM, a higher PSNR score means that the c-GAN has better performance in image generation (Setiadi, 2021).

Except for the BCE loss, the values of each criterion are the average values of evaluations based on 60 target images in the test dataset. In the case of BCE loss, UNEMA GAN showed the better training performance at all λ conditions than original U-Net architecture. The FID, SSIM, and PSNR were all optimized at a 1e-3 and 1e-4 GLR. Considering only training performance of c-GAN, higher GLRs led to better loss values. Nevertheless, drawing insights from the conditions under which other visual evaluation criteria achieve their maximum performance, maintaining an appropriate balance between the GLR and DLR proved more meaningful in enhancing the quality of generative images. Specifically, the UNEMA GAN achieved a higher improvement in FID score than other evaluation criteria, indicating that embedding the multi-head attention into the bottleneck of U-Net had a positive effect on improving image qualities.

An investigation into the impact of the number of attention heads was conducted. As shown in Table 4, all experiments were conducted with the number of attention heads fixed at eight. Generally, increasing the number of attention heads allows for more precise capture of spatial relationships between fine-grained features and pixel-level patterns, while reducing the number of heads tends to improve training speed. This trade-off necessitates a balanced approach. In this result, using eight attention heads provides an optimal compromise between performance and computational efficiency (Vaswani et al., 2017).

Table 6 presents the results of additional experiments conducted with varying numbers of attention heads. For each configuration, training performance and image quality evaluation criteria, including FID, SSIM, and PSNR were compared. While qualitative assessment through visual inspection revealed no clearly distinguishable differences in image quality, the quantitative evaluation criteria demonstrated that the configuration with eight attention heads consistently yielded the best performance. Furthermore, the lowest BCE Loss was observed under the eight-head configuration, confirming it as the most stable and effective setting in terms of both training stability and predictive accuracy. Based on these results of investigation, eight-head configuration was adopted for all model training processes, as it provided the most favorable results across both loss and image quality evaluations.

In addition to comparing the proposed UNEMA model with the conventional vanilla U-Net architecture, this study further includes experiments involving more recent architecture derived from vanilla U-Net. These comparisons aimed to comprehensively evaluate the relative positioning and practical utility of the proposed model. For this purpose,

Table 5
Evaluation of model performance depending on regularization factor, learning rate, and architecture type of generator.

Regularization factor(λ)	Learning rate	Generator	Evaluation criteria			
		Architecture	BCE Loss	FID	SSIM	PSNR
100	1e-2	U-Net	1.7979			
		UNEMA	1.7863			
	1e-3	U-Net	2.2881	18.0004	0.8731	18.0287
		UNEMA	**1.7332**	**10.2466**	**0.8896**	**18.8696**
	1e-4	U-Net	2.6198	17.0124	0.8577	18.6905
		UNEMA	**1.9162**	**13.3339**	**0.8999**	**21.3518**
	1e-5	U-Net	4.5236	21.2662	0.5281	16.0701
		UNEMA	4.4671	22.2165	0.5455	16.7316
125	1e-2	U-Net	1.2773			
		UNEMA	1.1380			
	1e-3	U-Net	2.2177	15.5839	0.9003	19.9203
		UNEMA	**1.7481**	**13.0220**	**0.9038**	**21.0279**
	1e-4	U-Net	2.2773	15.2708	0.8815	18.0158
		UNEMA	**1.8283**	**12.7264**	**0.8900**	**20.0656**
	1e-5	U-Net	4.5831	20.4213	0.5281	16.7585
		UNEMA	4.0596	20.2930	0.5455	16.7837
150	1e-2	U-Net	1.6268			
		UNEMA	1.5115			
	1e-3	U-Net	2.5731	23.3766	0.8968	19.4293
		UNEMA	**1.5444**	**18.9932**	**0.9038**	**19.9773**
	1e-4	U-Net	2.6557	21.0149	0.8911	18.7639
		UNEMA	**1.7508**	**14.4597**	**0.9072**	**22.1215**
	1e-5	U-Net	5.9286	21.8120	0.6106	17.3636
		UNEMA	5.5670	20.060	0.7477	17.4586

Table 6
Evaluation of model performance depending on number of attention head in UNEMA generator.

Regularization factor(λ)	Learning rate	Number of attention head	Evaluation criteria			
			BCE Loss	FID	SSIM	PSNR
150	1e-4	4	2.3421	20.5612	0.8959	19.8579
		8	**1.7508**	**14.4597**	**0.9072**	**22.1215**
		16	2.8300	20.4552	0.8948	18.9276

U-Net++ (Chen et al., 2021) and TransUNet (Zhou et al., 2018) were selected as benchmark architectures. These two models were developed three and six years, respectively after the vanilla U-Net, and represent notable attempts to overcome its limitations. U-Net++, also known as Nested U-Net, extends the skip connection design between the encoder and decoder by introducing a finely nested skip pathway structure. This architecture enhances feature recovery by mitigating information loss and has been shown to perform well in tasks where precise boundary delineation is essential (Chen et al., 2021). TransUNet is a hybrid architecture that combines Vision Transformer (ViT) components with the original U-Net structure. It is designed to simultaneously learn global contextual information and preserve structural consistency. Its ability to effectively model long-range dependencies makes it particularly suitable for high-resolution image reconstruction and generation tasks (Zhou et al., 2018).

In our comparative experiments, the discriminator architecture was kept constant, while the generator component in UNEMA was replaced with either U-Net++ or TransUNet. All models were trained under identical conditions, using a learning rate of 1e-4 and a regularization factor λ of 150, as optimized in our study. Table 7 summarizes the image quality evaluation criteria for the three models. Interestingly, the use of advanced architectures such as U-Net++ and TransUNet did not necessarily result in superior performance compared to UNEMA. While U-Net++ demonstrated comparable results under certain conditions, it did not consistently outperform the proposed model.

These results suggest that under very limited training conditions, such as a batch size of 2, a dataset consisting of only 240 images, and a computational environment using a GTX 1050 GPU with 2 GB of memory (Table 3), complex architectures may be more prone to overfitting and unstable training performance. In contrast, simpler architectural designs can offer more stable training outcomes in low-data regimes. It is important to note that with access to larger datasets, potentially numbering in the thousands, or more advanced hardware that supports higher batch sizes, enhanced architectures such as U-Net++ or TransUNet are likely to outperform simpler models. However, in the context of manufacturing and processing domains where data availability and computational resources remain limited, lightweight and robust designs such as UNEMA can still serve as highly practical and effective alternatives.

Proving the performance of the proposed GAN model should be evaluated not only in quantitative aspects but also in qualitative aspects. It is because the comparison of similarity between two images with human eyes contains subjective elements which are too ambiguous to be expressed numerically and theoretically. The qualitative evaluation was also conducted following the quantitative evaluation of the generative image confirmed by manipulating the GLR and regularization factor. The generative images from UNEMA GAN resulted in the better performance compared to those applying U-Net architecture. Furthermore, considering the results of Table 5 and generative performance, a condition with 1e-4 GLR and 150 λ was selected as the best training option, which demonstrated outstanding results in both qualitative and quantitative evaluations. Fig. 8 organizes the representative generative images with the test dataset, from the UNEMA GAN which trained with 1e-4 of GLR and 150 of regularization factor. Fig. 9 presents a comparison of images generated by U-Net and UNEMA. Overall, improvements in dimensional accuracy and feature representation are evident in the outputs of the UNEMA GAN. Under the −30 mA ΔI_{SC} (Fig. 9(b)) and the +15 mA ΔI_{SC} (Fig. 9(c)), the generative images produced by UNEMA show a significant enhancement in shape quality compared to those generated using U-Net alone as the generator architecture.

As a notable phenomenon, model collapse was observed in conditions where GLR was excessively high at 1e-2 and excessively low at 1e-5. The model collapse refers to a phenomenon where the probabilistic mode that the generator may output is unified into one, resulting in the same output being always displayed for any different input (Durall et al., 2020). When the GLR was adopted over 1e-2 in the training, only the black-colored image was output. In Table 5, UNEMA GAN trained with 1e-2 didn't show any evaluated value due to the black-colored images. And when the GLR was 1e-5 or less, the only identical cross-sectional hole was observed regardless of the input image. As with the trend observed in quantitative evaluations, a higher GLR did not necessarily result in better generative performance. Additionally, the lower BCE loss value did not always output a higher-quality generative image.

4.3. Validation of predictive accuracy and precision for generative cross-sectional hole

It was simultaneously demonstrated that the potential for successful training of Pix2Pix model with limited data and the efficacy of adopting UNEMA are both promising. However, to facilitate the goal of predictive model in real manufacturing field, it is essential to not only provide visually high-quality images but also accurately provide dimensional information of micro-holes. In other words, the generative images from the proposed model should accurately predict the geometrical

Table 7
Comparison of training and generation performance for U-Net++, TransUNet, and UNEMA generators.

Regularization factor(λ)	Learning rate	Generator Architecture	Evaluation criteria			
			BCE Loss	FID	SSIM	PSNR
150	1e-3	UNEMA	**1.5444**	**18.9932**	**0.9038**	**19.9773**
		U-Net++	3.4633	22.5112	0.5812	16.4861
		TransUnet	2.5090	19.4884	0.0298	3.0517
	1e-4	UNEMA	**1.7508**	**14.4597**	**0.9072**	**22.1215**
		U-Net++	2.4692	15.5704	0.8834	18.2219
		TransUnet	5.8048	18.8605	0.0655	16.2165
	1e-5	UNEMA	**5.5670**	**20.060**	**0.7477**	**17.4586**
		U-Net++	21.1031	15.6948	0.0585	16.7820
		TransUnet	65.8201	21.2537	0.0056	6.94031

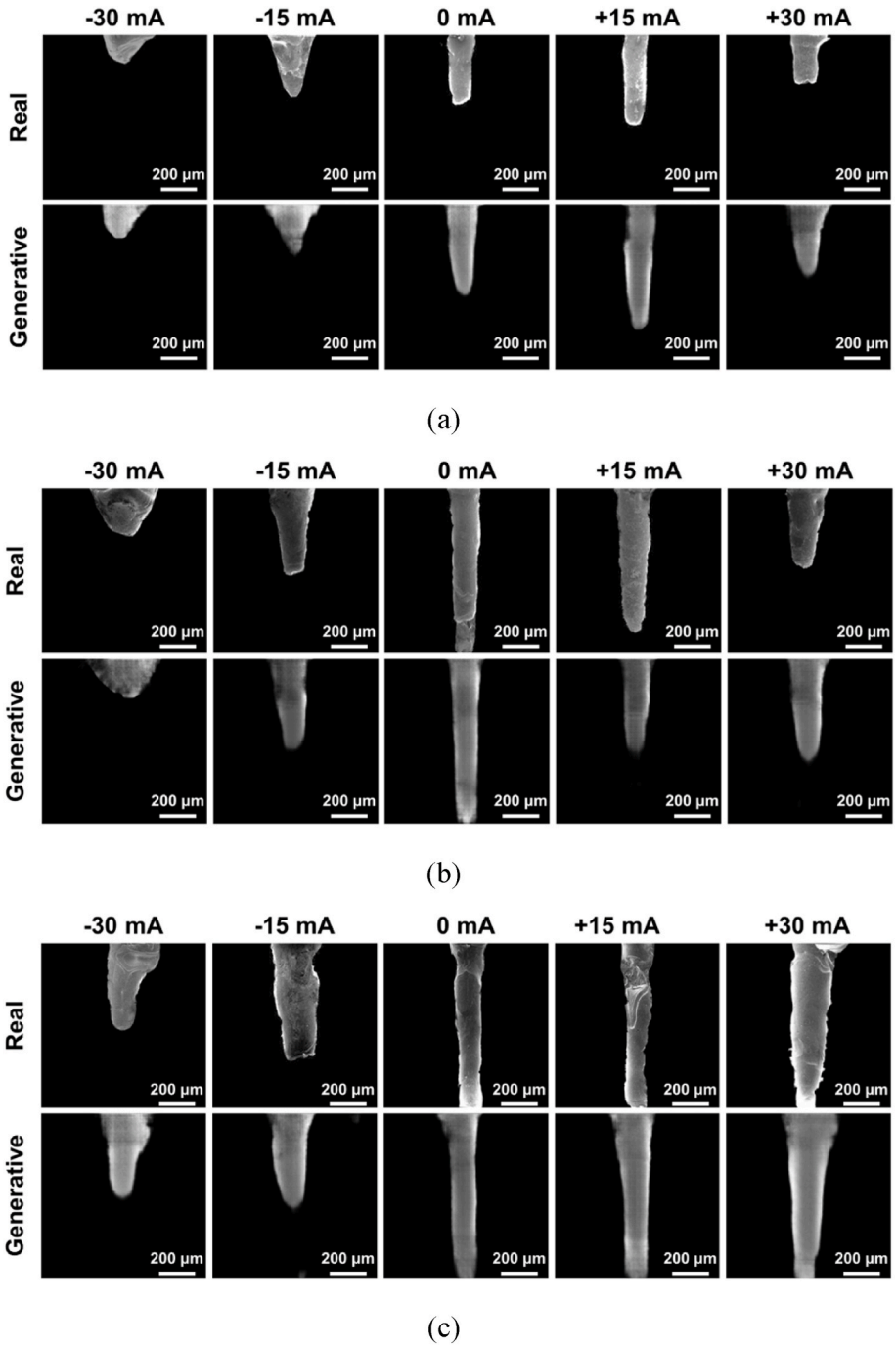

Fig. 8. Generative image of cross-sectional hole drilled with 40 mA beam current in (a) 20, (b) 40, and (c) 80 μs irradiation time.

dimensions of the holes drilled by electron beam irradiation. Certainly, while the SSIM evaluation between the target image and the generative image did reflect some degree of dimensional comparisons, validating how well the proposed model predicts the geometrical dimensions of drilled hole for actual operators in the manufacturing field can be a meaningful task. Experimental validations of inlet diameter, straightness, and penetration depth of the hole were further evaluated between the target image and the generative images. The images in the test dataset inherently contained geometrical dimensions as part of the experimental results, and therefore, it was determined that images for test can be utilized as reference of experimental validation. Fig. 9 depicts plots comparing geometric dimensions between experimental and predictive (generative) cross-sectional images of drilled holes. Since the images of the 60 cross-sectional holes in the test dataset were drilled under individually different experimental conditions, each image had different inlet diameter, depth, and straightness of hole. The distribution of geometrical dimensions can be identified according to 60 experimental conditions. The x-axis label of each plot in Fig. 9 refers to the descending ranks for each measured geometrical criterion. The black dots represent the experimental results from the target cross-sectional hole images in the test dataset, and other colored dots shows the dimensions of generative images based on individual architectures. The red dash-dot lines represent an error boundaries of ±20 % related to the experimental results. It was considered that a predictive performance with a relative error of less than 20 % was sufficient to determine the trend of process results in the process industry. The predictive accuracy

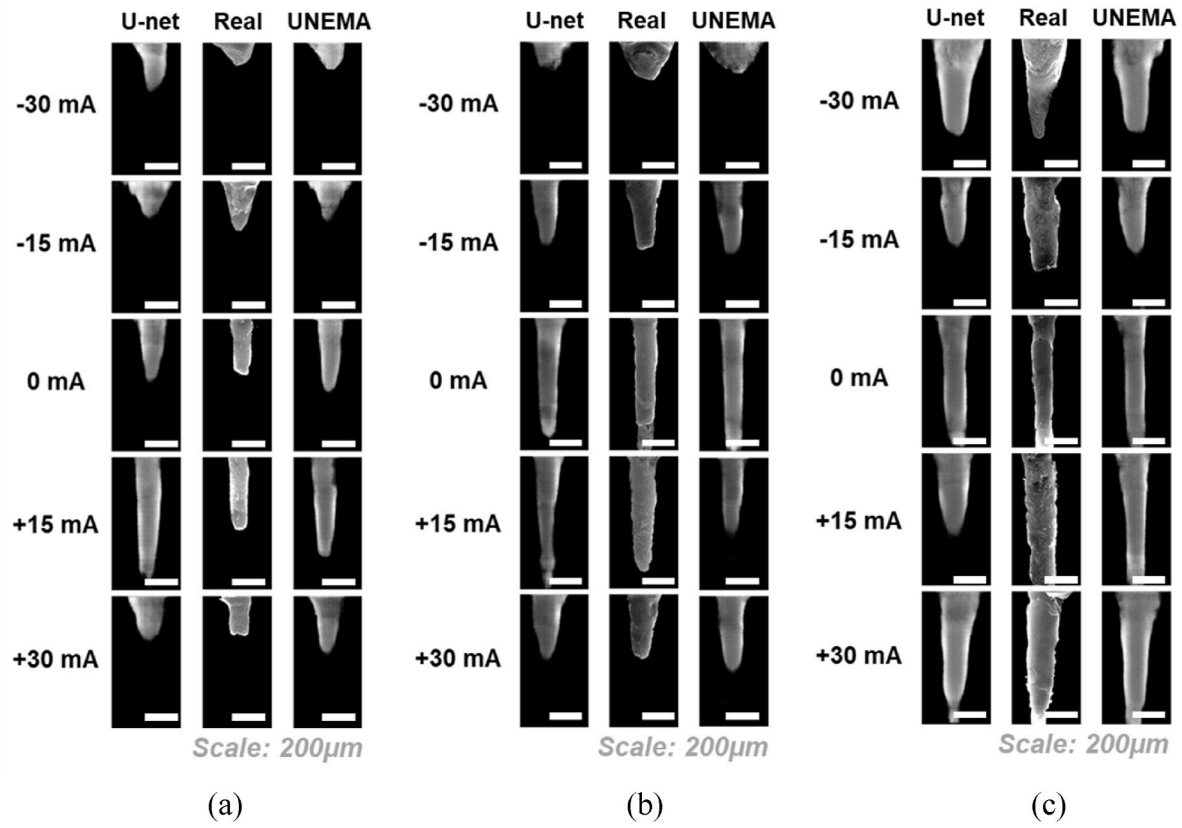

Fig. 9. Comparison of images generated by U-Net and UNEMA based on 30 mA beam current during (a) 20 μs, (b) 40 μs, and (c) 80 μs.

was calculated the relative error rate of geometrical dimensions between the experimental and generative results. And the predictive precision was determined by the number of generative results that predicted within the error boundaries, out of all 60 predictions. The specific experimental conditions of EB irradiation such as 20 mA of beam current, −30 mA of SC current change, and 20 μs irradiation time, failed to ablate and drill the substrate, resulting in the 60th image showing a value of zero (Fig. 10).

In Fig. 10(a), the inlet diameters of the generative images closely mirror the distribution and trends observed in the experimental inlet diameters. The UNEMA GAN-generative images increased the predictive accuracy compared to U-Net generative images and were placed inside the 20 % error boundaries. Specifically, the accuracy for the 53rd experimental result improved by 36.2 % on the prediction of UNEMA GAN. And the UNEMA GAN conducted an approximate 12 % increment in precision compared to U-Net-based generative images. The diameter predictions from the 57th to the 60th experimental results indicated that both the U-Net and UNEMA architectures performed with reduced efficacy. Nevertheless, the outputs from the UNEMA GAN demonstrated smaller difference from the experimental results than those generated by the model employing U-Net architecture.

In the case of straightness as shown in Fig. 10(b), the closer the angle is to 90°, the straighter and more vertical the microhole was drilled (Park et al., 2022) (Yilbas and Aleem, 2004). Experimental results of hole straightness varied from nearly vertical straight holes above 80° to sharply inclined tapered holes at 45°. Notably, the UNEMA GAN accurately reflected the trend of experimental hole straightness. The UNEMA generative images increased 8.3 % of the predictive precision of predictive model than original Pix2Pix model with U-Net architecture. The predictive precision of the proposed model was close to 99 %. And the UNEMA GAN improved the predictive accuracy up to 22.98 % in respect with the hole straightness. Moreover, the UNEMA generative image expressed the lower angle value for the 60th image than the U-Net generative image.

The most challenging aspect of the generative images from the c-GAN model for predicting geometric hole quality was the penetration (drilling) depth. The Pix2Pix model did not predict drilled depth based on the equation of physical heat transfer or heat diffusion. Instead, it purely learns the features of pixel images and updates through loss functions and optimizers to mimic the target image. Therefore, accurately predicting the generation of vertical pixel area with limited training data was quite challenging. The UNEMA-generative images were more adjacent to the experimental depth line than those based on the U-Net architecture. By embedding the Multi-Head Attention module (MHA) into the U-Net, the vertical predictive performance has been enhanced. All generative images achieving a depth of 800 μm exhibited high predictive accuracy (Fig. 10(c)). However, the partially penetrated holes from the U-Net generation displayed significant deviations from the established error boundaries. Consequently, the images predicted by the Pix2Pix model employing U-Net architecture demonstrated inferior depth prediction, achieving a precision of 66.7 %. The UNEMA GAN model ameliorated the extent of these deviations, demonstrating that the generative outputs were closer to the error boundaries. The proposed model exhibited a 15 % enhancement in the predictive precision compared to images generated by the U-Net architecture. In a specific instance, an improvement of 58.6 % in predictive accuracy was observed. Furthermore, the UNEMA GAN demonstrated substantially greater predictive accuracy for drilled depths less than 200 μm. Through experimental validations of the geometric depth of the drilled holes, it was concluded that the UNEMA GAN contributed to the refinement of geometric features, surpassing the performance of the original Pix2Pix model with U-Net architecture.

Following the validation of predictive accuracy and precision, a paired *t*-test was conducted to assess the statistical significance of the geometric metrics between U-Net and UNEMA, based on the results presented in Fig. 10 (Table 8).The paired *t*-test was conducted using the

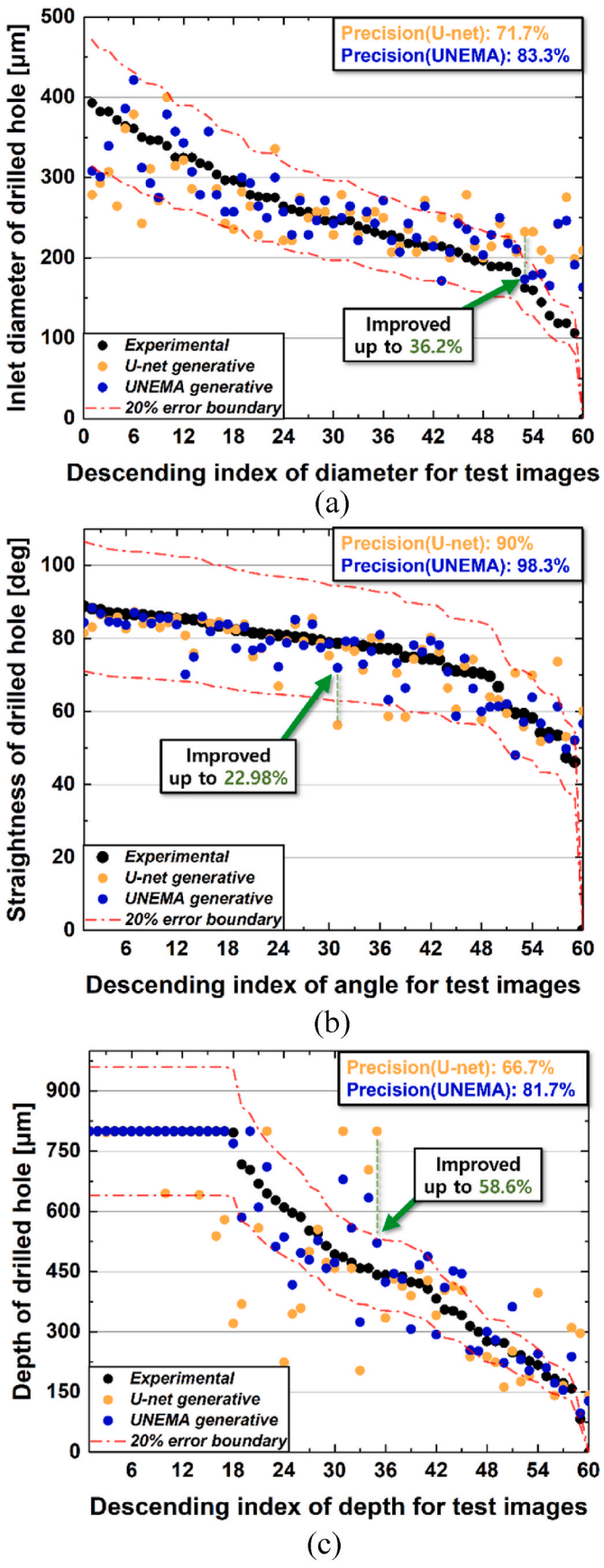

Fig. 10. Accuracy validation between experimental and generative images for (a) hole inlet diameter, (b) hole straightness, and (c) drilling depth.

MATLAB environment, with the number of test samples set to 60, resulting in 59 degrees of freedom. The analysis yielded a p-value of 0.4137 for inlet diameter, 0.0003 for drilled hole depth, and 0.0497 for hole straightness. Among these, the p-values for drilled depth and straightness were both below 0.05, indicating that the differences between the two models were statistically significant.

In the case of inlet diameter, no noticeable qualitative or quantitative differences were observed between the two models. However, in terms of drilled depth and straightness, clear and meaningful performance differences were evident. The extremely low p-value for drilled depth strongly supported a substantial performance gap between the models, while straightness also demonstrated a statistically significant difference. These results suggested that the performance improvements of the proposed model are not due to random variation or sampling error, but rather reflect statistically validated enhancements. Moreover, the statistical significance of model performance differences was most prominent in the order of drilled depth, straightness, and inlet diameter. In other words, the adoption of the UNEMA generator is supported not only by empirical results but also by statistical evidence.

5. Conclusion

To experimentally prepare image data for hole inlets and cross-sectional holes, a conditional generative adversarial network (GAN) model was developed to predict microholes in thin silicon wafers. Over 60 experiments were conducted using high-energy electron beam drilling on SiC wafers, varying parameters such as beam current, solenoid coil current, and irradiation time. A fully penetrated microhole with a minimum diameter of 223 μm and an aspect ratio of 3.58 was examined. The changes in inlet diameter and cross-sectional shape of the holes were analyzed under various experimental conditions. Using a limited dataset of images, a Pix2Pix model was successfully trained to accurately capture the geometrical dimensions of the drilled holes, rather than merely replicating the cross-sectional shape. The introduction of a U-Net with an Embedded Multi-head Attention module enhanced the Pix2Pix model's performance in predicting the geometrical dimensions of the drilled holes. This modified U-Net retained the core characteristics of the original architecture, while the multi-head attention module embedded in the bottleneck section improved training loss convergence and enhanced the representation of local features in the generative images. The UNEMA GAN demonstrated superior performance over conventional U-Net architectures in predicting penetration depth. The model achieved predictive precision rates of 83.3 % for inlet diameter, 98.3 % for hole straightness, and 81.7 % for drilled depth. Compared to the original Pix2Pix model with a U-Net architecture, the UNEMA architecture achieved maximum improvements of 36.2 % in inlet diameter prediction, 22.98 % in hole straightness, and 58.6 % in drilled depth. This research concludes that the UNEMA architecture effectively harnesses the strengths of both U-Net and the multi-head attention module, leading to significant enhancements in deep learning performance. The performance improvement of the images generated by the UNEMA generator was also verified, compared to those generated by the conventional U-Net, was statistically significant.

In developing the c-GAN model, it was assumed that the hole inlet contains deformation history, which provides insight into the causality between experimental conditions and outcomes. The successful training and validation of the proposed model confirmed this assumption. It was also demonstrated that images capturing deformation history offer valuable features for the deep learning architecture. The model only requires an RGB image of the hole inlet, which can be obtained using a simple camera tool. Given the ease of acquiring these images, integrating this approach into the manufacturing process for silicon wafers is expected to significantly reduce prediction and analysis time.

Moreover, with a larger dataset than the one used in this study, the quality of the generative images could potentially rival that of high-resolution microscopy equipment, such as scanning electron

Table 8
Paired *t*-test of geometrical evaluation between U-Net and UNEMA generative images.

Geometrical evaluation Criteria	Pairs	Statistical indicator		Results of paired *t*-test			
		Mean	SD	95 % C.I.		t-statistic	p-value
				Lower	Upper		
Inlet diameter	U-Net	257.61	45.67	−134.28	56.00	−0.8232	0.4137
	UNEMA	308.88	398.93				
Straightness	U-Net	485.63	235.25	0.0015	1.9020	2.0041	**0.0497**
	UNEMA	519.69	231.14				
Depth	U-Net	74.11	10.66	25.38	80.38	3.8480	**0.0003**
	UNEMA	74.00	10.96				

microscopes (SEM). When combined with analytical models, this approach also holds promise for digital transformation innovations, including metamodeling research for electron beam drilling on silicon wafers.

This study and the proposed model are expected to contribute in three significant ways. First, it experimentally demonstrates the feasibility of using high-energy electron beam drilling for TSV fabrication in bulk silicon wafers. Second, it introduces an AI-based non-destructive method for predicting microhole cross-sections, offering practical applicability in resource-constrained manufacturing environments. Third, it validates the effectiveness of deep learning, particularly the proposed UNEMA architecture under small-scale experimental conditions, emphasizing the importance of simplicity and domain-informed design in industrial AI applications. In the current landscape where novel deep learning architectures are constantly emerging, this work strongly suggests that the now-legacy U-Net architecture, when enhanced with add-on modules such as multi-head attention, can still serve as a viable and effective solution in a resource-limited manufacturing environment.

CRediT authorship contribution statement

Hyunmin Park: Validation, Formal analysis, Conceptualization, Writing – original draft. **Jun Goo Kang:** Validation, Data curation. **Jin Seok Kim:** Data curation, Validation. **Eun Goo Kang:** Validation, Data curation. **Seung-Kyum Choi:** Writing – review & editing. **Hyung Wook Park:** Writing – review & editing, Supervision, Funding acquisition, Data curation.

Consent to participate

The authors voluntarily agree to participate in this research study.

Consent to publish

Its publication is approved by all authors, and that, if accepted, it will not be published elsewhere in the same form in English or in any other language, without the written consent of the Publisher.

Ethical approval

The authors confirm that the content of the manuscript is original and has not been published previously, that it is not under consideration for publication elsewhere.

Availability of data and materials

The authors confirm that the data supporting the findings of this study are available within the article.

Funding

The authors declare any funding or research grants (and their source) received in the course of study, research or assembly of the manuscript.

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.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT of Korea (No. 2022R1A2C3007963 and No. RS-2025-02223634) and Development of smart manufacturing multiverse platform based on multisensory fusion avatar and interactive AI funded by the Ministry of Trade, Industry and Energy (20025702)

Data availability

Data will be made available on request.

References

Bakeev, I.Y., Klimov, A.S., Oks, E.M., Zenin, A.A., 2019. Double-coil magnetic focusing of the electron beam generated by a plasma-cathode electron source. Rev. Sci. Instrum. 90 (2). https://doi.org/10.1063/1.5078655.

Caya, M.V., Padilla, D., Ombay, G., Hernandez, A.J., 2019. Detection and counting of red blood cells in human urine using canny edge detection and circle hough transform algorithms. IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) 1–5. https://doi.org/10.1109/HNICEM48295.2019.9072708.

Cela, E., Shahidi, S., Parangi, P., Shrestha, R., Simpson, G., Widiez, J., et al., 2023. High sensitivity surface defect inspection of SiC and SmartSiCTM substrates using a DUV laser-based system. Defect Diffusion Forum 425, 57–61. https://doi.org/10.4028/p-4918s1.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., et al., 2021. Transunet: transformers make strong encoders for medical image segmentation. arXiv preprint arXiv: 2102.04306. https://doi.org/10.48550/arXiv.2102.04306.

Chen, P.C., Miao, W.C., Ahmed, T., Pan, Y.Y., Lin, C.L., Chen, S.C., et al., 2022. Defect inspection techniques in SiC. Nanoscale Res. Lett. 17 (1), 30. https://doi.org/10.1186/s11671-022-03672-w.

Chen, Q., Belyansky, M., Sulehria, Y., Horibe, A., Qin, L., Pujari, R., et al., 2023a. Next generation infrared (IR) laser debonding/silicon handle technology for precision chiplet technology applications. IEEE 73rd Electronic Components and Technology Conference, pp. 1175–1180. https://doi.org/10.1109/ECTC51909.2023.00201.

Chen, G., Zhang, G., Yang, Z., Liu, W., 2023b. Multi-scale patch-GAN with edge detection for image inpainting. Appl. Intell. 53 (4), 3917–3932. https://doi.org/10.1007/s10489-022-03577-2.

Cook, R.F., 2006. Strength and sharp contact fracture of silicon. J. Mater. Sci. 41, 841–872. https://doi.org/10.1007/s10853-006-6567-y.

Derby, N., Olbert, S., 2010. Cylindrical magnets and ideal solenoids. Am. J. Phys. 78 (3), 229–235. https://doi.org/10.1119/1.3256157.

Du, H., Shi, Z., 2020. Wafer SEM image generation with conditional generative adversarial network. J. Phys. Conf. 1486 (2), 022041. https://doi.org/10.1088/1742-6596/1486/2/022041.

Durall, R., Chatzimichailidis, A., Labus, P., Keuper, J., 2020. Combating mode collapse in gan training: an empirical analysis using hessian eigenvalues. arXiv preprint arXiv: 2012.09673. https://doi.org/10.48550/arXiv.2012.09673.

Egerton, R.F., 2005. Physical Principles of Electron Microscopy, vol. 56. Springer.

Goswami, A., 2020. Surface Micromachining—Advances and Advanced Characterization Techniques. Accuracy Enhancement Technologies for Micromachining Processes, pp. 165–198. https://doi.org/10.1007/978-981-15-2117-1_9.

Guan, J., Zhao, Y., 2022. Non-contact grinding/thinning of silicon carbide wafer by pure EDM using a rotary cup wheel electrode. Precis. Eng. 74, 209–223. https://doi.org/10.1016/j.precisioneng.2021.12.001.

Guo, H., Cao, S., Li, L., Zhang, X., 2022. A review on the mainstream through-silicon via etching methods. Mater. Sci. Semicond. Process. 137, 106182. https://doi.org/10.1016/j.mssp.2021.106182.

Heath, D.J., Grant-Jacob, J.A., Xie, Y., Mackay, B.S., Baker, J.A., Eason, R.W., Mills, B., 2018. Machine learning for 3D simulated visualization of laser machining. Opt. Express 26 (17), 21574–21584. https://doi.org/10.1364/OE.26.021574.

Henrie, J., Kellis, S., Schultz, S.M., Hawkins, A., 2004. Electronic color charts for dielectric films on silicon. Opt. Express 12 (7), 1464–1469. https://doi.org/10.1364/OPEX.12.001464.

Howitt, D.G., Chen, S.J., Gierhart, B.C., Smith, R.L., Collins, S.D., 2008. The electron beam hole drilling of silicon nitride thin films. J. Appl. Phys. 103 (2). https://doi.org/10.1063/1.2828157.

Hurtik, P., Tomasiello, S., Hula, J., Hynar, D., 2022. Binary cross-entropy with dynamical clipping. Neural Comput. Appl. 34 (14), 12029–12041. https://doi.org/10.1007/s00521-022-07091-x.

Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134.

Janocha, K., Czarnecki, W.M., 2017. On loss functions for deep neural networks in classification. arXiv preprint arXiv:1702.05659. https://doi.org/10.48550/arXiv.1702.05659.

Kang, J.G., Kim, J.S., Min, B.K., Kang, E.G., 2023. Inlet hole shape analysis depending on the focus conditions for electron beam micro-hole drilling. Int. J. Precis. Eng. Manuf. 24 (8), 1307–1317. https://doi.org/10.1007/s12541-023-00785-9.

Kar, T., Deshmukh, S.S., Goswami, A., 2024. Fabrication of micro holes using low power fiber laser: surface morphology, modeling and soft-computing based optimization. Advances in Manufacturing 1–22. https://doi.org/10.1007/s40436-024-00484-2.

Khuat, V., Chen, T., Gao, B., Si, J., Ma, Y., Hou, X., 2014. Uniform nano-ripples on the sidewall of silicon carbide micro-hole fabricated by femtosecond laser irradiation and acid etching. Appl. Phys. Lett. 104 (24). https://doi.org/10.1063/1.4883880.

Kim, J., Lee, W.J., Park, H.W., 2016. The state of the art in the electron beam manufacturing processes. Int. J. Precis. Eng. Manuf. 17, 1575–1585. https://doi.org/10.1007/s12541-016-0184-8.

Kim, J., Lee, W.J., Park, H.W., 2018. Temperature predictive model of the large pulsed electron beam (LPEB) irradiation on engineering alloys. Appl. Therm. Eng. 128, 151–158. https://doi.org/10.1016/j.applthermaleng.2017.08.142.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. *arXiv preprint arXiv*:1412.6980. https://doi.org/10.48550/arXiv.1412.6980.

Kizuka, T., Tanaka, N., 1997. Atomic desorption process in nanometre-scale electron-beam drilling of MgO in high-resolution transmission electron microscopy. Philos. Mag. Lett. 76 (4), 289–298. https://doi.org/10.1080/095008397179048.

Kumar, P., Dutta, I., Huang, Z., Conway, P., 2021. Materials and processing of TSV. 3D Microelectronic Packaging: from Architectures to Applications, pp. 47–70.

Kurniawan, R., Chen, S., Xu, M., Teng, H., Chen, J., Ali, S., et al., 2024. Understanding the mechanism of ultrasonic vibration-assisted drilling (UVAD) for micro-hole formation on silicon wafers using numerical and analytical techniques. Int. J. Adv. Des. Manuf. Technol. 1–31. https://doi.org/10.1007/s00170-024-13412-2.

Leitz, K.H., Koch, H., Otto, A., Maaz, A., Löwer, T., Schmidt, M., 2012. Numerical simulation of drilling with pulsed beams. Phys. Procedia 39, 881–892. https://doi.org/10.1016/j.phpro.2012.10.113.

Li, Z., Allegre, O., Li, Q., Guo, W., Li, L., 2021. Femtosecond laser single step, full depth cutting of thick silicon sheets with low surface roughness. Opt Laser. Technol. 138, 106899. https://doi.org/10.1016/j.optlastec.2020.106899.

Liu, Y., Wang, J., Li, W., Wang, C., Zhang, Q., Yang, X., Cheng, L., 2016. Effect of energy density and feeding speed on micro-holes drilling in SiC/SiC composites by picosecond laser. Int. J. Adv. Des. Manuf. Technol. 84, 1917–1925. https://doi.org/10.1007/s00170-015-7844-1.

Liu, S., Chen, J., Qu, C., Hou, R., Lv, H., Pan, T., 2021. LOSGAN: latent optimized stable GAN for intelligent fault diagnosis with limited data in rotating machinery. Meas. Sci. Technol. 32 (4), 045101. https://doi.org/10.1088/1361-6501/abd0c1.

Liu, M., Yu, S., He, L., Ni, Y., 2022. Recent progress on crack pattern formation in thin films. Soft Matter 18 (32), 5906–5927. https://doi.org/10.1039/D2SM00716A.

Martin, H., Hubert, R., Thomas, U., Bernhard, N., 2018. GANs trained by a two time-scale update rule converge to a local nash equilibrium. arXiv preprint arXiv:1706.08500. https://doi.org/10.48550/arXiv.1706.08500.

McDonnell, M.D.T., Grant-Jacob, J.A., Praeger, M., Eason, R.W., Mills, B., 2021. Identification of spatial intensity profiles from femtosecond laser machined depth profiles via neural networks. Opt. Express 29 (22), 36469–36486. https://doi.org/10.1364/OE.431441.

Mills, B., Heath, D.J., Grant-Jacob, J.A., Eason, R.W., 2018. Predictive capabilities for laser machining via a neural network. Opt. Express 26 (13), 17245–17253. https://doi.org/10.1364/OE.26.017245.

Moreno, M.V.R., Robles, G., Albarracín, R., Rey, J.A., Tarifa, J.M.M., 2017. Study on the self-integration of a rogowski coil used in the measurement of partial discharges pulses. Electr. Eng. 99, 817–826. https://doi.org/10.1007/s00202-016-0456-4. Electr Eng (2017) 99:817–826.

Mumbelli, J.D., Guarneri, G.A., Lopes, Y.K., Casanova, D., Teixeira, M., 2023. An application of generative adversarial networks to improve automatic inspection in automotive manufacturing. Appl. Soft Comput. 136, 110105. https://doi.org/10.1016/j.asoc.2023.110105.

Nasrollahi, V., Penchev, P., Batal, A., Le, H., Dimov, S., Kim, K., 2020. Laser drilling with a top-hat beam of micro-scale high aspect ratio holes in silicon nitride. J. Mater. Process. Technol. 281, 116636. https://doi.org/10.1016/j.jmatprotec.2020.116636.

Ozgur, M., Huff, M., 2017. Plasma etching of deep high-aspect ratio features into silicon carbide. J. Microelectromech. Syst. 26 (2), 456–463. https://doi.org/10.1109/JMEMS.2017.2661961.

Park, H., Kang, J.G., Kim, J.S., Kang, E.G., Choi, S.K., Kim, J., Park, H.W., 2022. Predictive modeling of microhole profile drilled using a focused electron beam with backing materials. Int. J. Therm. Sci. 177, 107584. https://doi.org/10.1016/j.ijthermalsci.2022.107584.

Popovich, V.A., Riemslag, A.C., Janssen, M., Bennett, I.J., Richardson, I.M., 2013. Characterization of multicrystalline silicon solar wafers fracture strength and influencing factors. Int. J. Mater. Sci. 3 (1), 9–17.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015 part III 18, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Sahli, H., Slama, A.B., Sayadi, M., 2023. Skin lesion segmentation based on modified U-NET architecture. 2023 IEEE International Conference on Advanced Systems and Emergent Technologies, pp. 1–5. https://doi.org/10.1109/IC_ASET58101.2023.10150888.

Sekhar, H., Fukuda, T., Tanahashi, K., Takato, H., Ono, H., Sampei, Y., Kobayashi, T., 2020. Mechanical strength problem of thin silicon wafers (120 and 140 μm) cut with thinner diamond wires (si kerf 120→ 100 μm) for photovoltaic use. Mater. Sci. Semicond. Process. 119, 105209. https://doi.org/10.1016/j.mssp.2020.105209.

Setiadi, D.R.I.M., 2021. PSNR vs SSIM: imperceptibility quality assessment for image steganography. Multimed. Tool. Appl. 80 (6), 8423–8444. https://doi.org/10.1007/s11042-020-10035-z.

Shen, X., Chen, L., Li, J., Zhao, J., 2016. Silicon microhole arrays architecture for stable and efficient photoelectrochemical cells using ionic liquids electrolytes. J. Power Sources 318, 146–153. https://doi.org/10.1016/j.jpowsour.2016.04.027.

Singh, S., Samuel, G.L., 2023. Ultrashort pulse laser micro drilling of silicon. Mater. Today Proc. https://doi.org/10.1016/j.matpr.2023.04.133.

Sinitca, A.M., Kayumov, A.R., Zelenikhin, P.V., Porfiriev, A.G., Kaplun, D.I., Bogachev, M.I., 2023. Segmentation of patchy areas in biomedical images based on local edge density estimation. Biomed. Signal Process Control 79, 104189. https://doi.org/10.1016/j.bspc.2022.104189.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., et al., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wang, Z., 2019. Microsystems using three-dimensional integration and TSV technologies: fundamentals and applications. Microelectron. Eng. 210, 35–64. https://doi.org/10.1016/j.mee.2019.03.009.

Wang, Z., Yu, L., Pu, L., 2021. Defect simulation in SEM images using generative adversarial networks. Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV 11611, 113–119. https://doi.org/10.1117/12.2581881.

Xu, M., Girish, Y.R., Rakesh, K.P., Wu, P., Manukumar, H.M., Byrappa, S.M., Byrappa, K., 2021. Recent advances and challenges in silicon carbide (SiC) ceramic nanoarchitectures and their applications. Mater. Today Commun. 28, 102533. https://doi.org/10.1016/j.mtcomm.2021.102533.

Yilbas, B.S., Aleem, A., 2004. Laser hole drilling quality and efficiency assessment. Proc. IME B J. Eng. Manufact. 218 (2), 225–233. https://doi.org/10.1243/095440504322886541.

Zhang, G., Pan, Y., Zhang, L., 2021. Semi-supervised learning with GAN for automatic defect detection from images. Autom. ConStruct. 128, 103764. https://doi.org/10.1016/j.autcon.2021.103764.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. UNet++: a nested U-Net architecture for medical image segmentation. In: Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis (DLMIA) and 8th International Workshop on Multimodal Learning for Clinical Decision Support (ML-CDS), Held in Conjunction with MICCAI 2018, vol 11045. Granada, Spain, pp. 3–11.