

High-precision page information extraction from 3D scanned booklets using physics-informed neural network

Zhongjiang Han, Jiarui Ou and Koji Koyamada

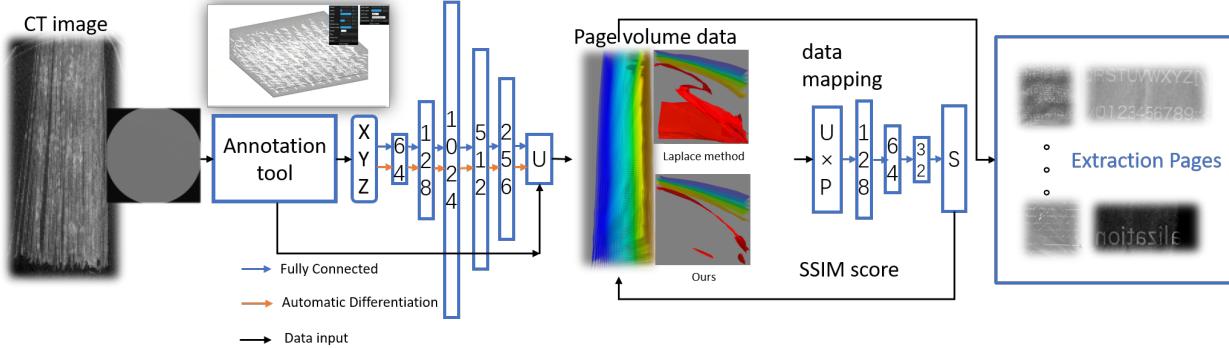


Fig. 1. The process of extracting page information using the physics-informed neural network model

Abstract—X-ray-based computerized tomography scans are used to analyze page information in closed booklets noninvasively. An important task is to extract the page information. Previously, the Laplace equation was used to calculate the page number field and extract the page information as an iso-surface. However, this technique cannot extract the page information properly. To solve this problem and improve the accuracy of the extracted page information, we propose a page information extraction method using a physics-informed neural network. The proposed method employs a structural similarity measure—often used in image processing research—to numerically evaluate the appropriateness of the page extraction. New history booklet are used to verify the effectiveness of this method in addition to the conventional booklet data.

Index Terms—3D CT, Page segmentation and extraction, deep learning, PINN

1 INTRODUCTION

Historical documents contain valuable descriptions of past events. They are fundamental to historical research and have important implications for historians. However, owing to aging and external erosion, the text in some of these documents cannot be easily deciphered. Additionally, some documents must not be invasively examined because they are too important to risk damage. Moreover, some of these documents are too charred to decipher. For example, in 1750, Karl Weber discovered over 1,800 papyri, which were carbonized by the eruption of Vesuvius in 79 AD, near Herakleaneum in Campania, Italy [8]. These papyri were wound into tubes and are difficult to open without damage, making them impossible to decipher without a noninvasive technique. In 2004, fire and firefighting water severely damaged over 62,000 ancient documents at the Duchess Anna Amalia Library in Germany [20]. Therefore, it is necessary to acquire and decipher the information contained in these documents by using noninvasive techniques.

X-ray computerized tomography (CT) imaging, wherein X-rays irradiate an imaging target to reveal the contents, is the most widely used noninvasive literature browsing method. The spatial resolution of an X-ray CT system is inversely proportional to the focal size of the X-ray source [11] [13]. Research on analyzing ancient documents

through noninvasive CT scanning methods has been conducted. The components of ancient documents that are of interest to this study mainly comprise paper and ink. During the middle ages, parchments and papyri were widely used in place of paper [17]. Gallic ink has been widely used since the Roman Empire [21], and its main constituents are gallic acid and iron (II) sulfate, in which iron ions contribute to the color characteristics [43]. Differences in X-ray absorptivity between inks containing iron and paper containing cellulose make it possible to visualize the written text and compare it with other types of ink (Stromer et al. [45].) Bergmann et al. used X-ray fluorescence imaging to decipher an ancient manuscript by Archimedes [4].

We sought to decipher the overlaid text of a document once written on parchment. A technique similar to ours was used by Bergmann et al. to distinguish between the two types of ink considered [5]. However, X-ray fluorescence imaging requires direct X-ray irradiation after opening a document. In this study, we used a micro-X-ray CT system to visualize documents that cannot be opened, such as the carbonized document examined by Seales et al. [41] and the paper prepared by Baum et al. [2]. Baum et al. made markings on two-dimensional (2D) slices, formed contour lines, and then formed mesh data to achieve page extraction. In contrast, we used deep learning to generate three-dimensional (3D) page volume data. There have been many studies on pattern recognition. For example, Vito Mocella et al. successfully visualized the inside of a scroll of carbonized papyrus using X-ray phase imaging [26]. In this study, because there was no significant difference between the density of the papyrus paper and that of the ink, visualization using a normal micro X-ray CT system was impossible; X-ray phase imaging was used instead.

Extant studies on page information extraction from historical documents have involved processing of 2D slices. However, we recently developed a novel method for page extraction directly from 3D volume data. This method efficiently extracts page data from 3D booklet im-

• Zhongjiang Han is with Kyoto University. E-mail: zhongjianghan@gmail.com
• Jiarui Ou is with Kyoto University. E-mail: oujiarui@outlook.com
• Koji Koyamada is with Kyoto University. E-mail: koyamada.koji.3w@kyoto-u.ac.jp

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

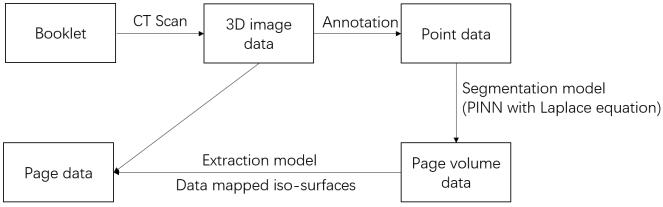


Fig. 2. Overview of our proposed approach. Data is obtained from CT scans, annotated for deep learning, page volume data is obtained by learning, and finally page data is obtained by mapping.

ages acquired by 3D CT devices by using the Laplace equation. The effectiveness and utility of this method were evaluated in Ou et al. [30]. This method cannot properly fit the page in the face of some data, such as Laplace method in Fig.1 (Represents using the previous method), the fitted page shape is wrong (red part). In addition to this, we also find that the Laplace equation does not fit the page correctly when faced with more complex page shapes, such as rolls, irregular folds, etc. The following is the result obtained by the method proposed in this paper.

Occasionally, a page surface may not be adequately represented by the Laplace equation. To overcome this limitation, we herein propose a method to improve the accuracy of page extraction while handling more complex data. Based on the universal approximation theorem, we believe that deep-learning models can be used instead of the Laplace equation to fit the spatial structure of booklets. We use a Laplace-equation-based physics-informed neural network (PINN) to approximate the page volume data and construct a structural similarity (SSIM)-based page-extraction model.

The approach of our proposed method is shown in Fig.2. The 3D image data were obtained by CT scanning, and the data were imported into the self-developed annotation tool to annotate the pages with page numbers. Thus, the 3D coordinates and the data corresponding to the page numbers were obtained. A deep-learning model trained these data to obtain the page volume data (where the scalar data value represents the page number), which can calculate all voxels where each page is located. Finally, the page data are obtained by mapping the corresponding iso-surfaces (i.e., the surface corresponding to the page).

A schematic of the proposed method is shown in Fig.1, where the input CT image data are annotated by a self-developed annotation tool fitted by a deep learning model. Finally, the text in the document is extracted.

Our main contributions are as follows:

- A page extraction method based on volume rendering and 2D slice annotation was developed.
- We built a page segmentation model based on a PINN using the Laplace equation (hereinafter referred to as the PINN model) and provide a complete set of automatic page-extraction methods.
- By fitting the spatial structure via deep learning, the accuracy of page extraction was improved. We confirm the effectiveness of the proposed method through several sets of experiments.

2 RELATED WORK

2.1 PINN and Universal approximation theory

Raissi et al. proposed PINN, an emerging deep-learning framework for solving forward and inverse (solution and derivation) problems of nonlinear partial differential equations [36] [34] [35]. Wang et al. used PINN to reconstruct dense fields from sparse image data for visualization and analysis [47]. Cheng et al. demonstrated that PINN outperformed traditional deep learning models in solving fluid dynamics problems [6]. SSIM is a quantitative metric often used in image processing studies to compare the similarity of images [12]. Hore et al. discuss the difference between the peak-signal-to-noise ratio (PSNR) and SSIM [12]. Sara et al. added a comparison on the feature

similarity indexing method (FSIM) and suggested that PSNR and mean-squared error (MSE) only output absolute errors, whereas SSIM and FSIM output perceptual errors, which are more easily understood by humans [39].

In their CT-based analysis of historical documents, Stromer et al. used SSIM as an evaluation criterion and analyzed the differences of various historical documents [44]. Su et al. used SSIM as a criterion for evaluating Mongolian historical documents [46].

Universal approximation theory forms the theoretical basis for replacing the Laplace equation using a deep-learning model in this study. The effectiveness of this theory on various neural networks has been demonstrated several times [16] [40]. Single-layer neural networks and deep multilayer networks are used in experiments to fit complex spatial structure equations [33]. We believe that a neural network can fit any continuous function; thus, we can improve the accuracy of page extraction by constructing a neural network to fit the structural equations of the page space. Multilayer perceptrons (MLPs) are universal function approximators, as shown by Cybenko's theorem [7]. They can be used to create mathematical models via regression analysis and are useful in research because of their ability to solve problems stochastically, thereby allowing approximate solutions for highly complex problems such as fitness approximations [29] [50] [10]. As classification is a particular case of regression in which the response variable is categorical, MLPs make good classifier algorithms [31] [48]. It has been established that overly complex neural networks are not conducive to solving the equation fitting problem [22], which is the reason for the uncomplicated structure of the proposed network.

2.2 historical documents page extraction

Previous studies have focused on page extraction at the 2D level, implementing slice-by-slice extraction tasks. For example, Samko et al. performed page segmentation on each slice to achieve the virtual unfolding of parchment documents [38]. Seales et al. used 3D scanning to virtually unfold charred historical documents by constructing a grid [42]. However, these studies did not directly fit the spatial structure of the data at the 3D level, leading to inaccuracies. Recently, some researchers have proposed processing 3D visualizations of CT images using deep learning. For example, Shreeraj et al. proposed an end-to-end visual diagnostic system to visualize 2D CT images at the 3D level while performing image segmentation and recognition by deep learning [18]. This is similar to our research concept.

Research on analyzing data by artificial labeling has been widely proposed. Web-based technology for image annotation tools at the 3D level is being developed extensively. Nguyen et al. created a manual annotation tool that can annotate 3d scene data from both 2D and 3D levels, facilitating further processing of 3D data [28]. Bowen et al. built a web-based visualization system that can provide interactive annotations, analyze genetic data of organisms, and unify multiple large regulatory network models [49]. In the field of image processing, besides the annotation of 3D scenes, CT medical images also face significant challenges. Moreira et al. developed a web-based plugin to perform 3D annotation on medical CT images [27] while Meng et al. discussed the performance of 2D annotation compared to 3D annotation based on experiments [25]. Rubin et al. discussed the web-based annotation in medical images and the application of the means [37].

Herein, we aim to process CT images on a 3D level. In previous research, spatial data collection was achieved by building a 3D annotation tool. To benefit from the promotion of the program, we constructed it on the Internet because HTML scripts do not require any extra plugin beyond the browser itself, while OpenGL is also a cross-platform standard. Using this technique, we built the working platform to visualize the 3D volume data containing the booklet information scanned by the CT machine.

Using annotation tools to generate page volume data to achieve segmentation has been proven feasible. Hu et al. used the annotation tool we built to visualize a 3D model of the magnetic lines of force by manually labeling the extent of magnetic lines of force for the nuclear fusion reaction, solving the boundary ambiguity problem [14, 15]. In

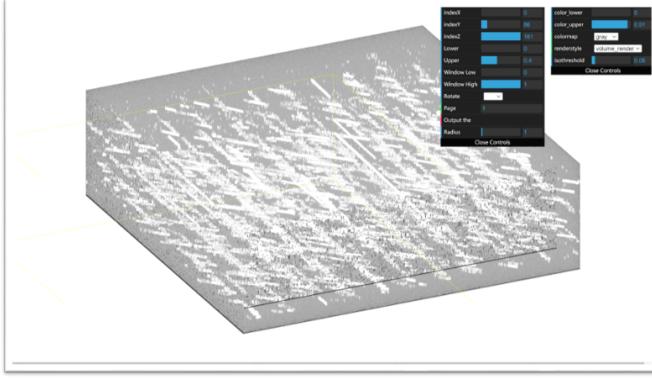


Fig. 3. The main interface of the annotation tool, including the visualization volume data interface and the control panel.

our previous research, we proposed a self-built annotation tool. Users can annotate data points in the 3D space in real-time by importing CT images, and the data obtained can be used for further experiments.

In the study by Mocella et al., the first x-rays were used to extract text from non-expandable historical documents. In this study, we developed volume renderings to demonstrate their results [26]. Based on the previously self-built annotation model, we added the volume rendering function, which can more intuitively determine the page distribution in the book CT data.

As shown in Fig.3, the main interface of the annotation tool includes the visualization volume data interface and panels for adjusting the visualization section and the voxel density, inputting the page data, and adding in-range annotation.

3 METHOD

As shown in Fig.4, the proposed model is divided into two main parts—the page-segmentation model (PINN model) and the page-extraction model. The PINN model is based on the Laplace equation, and the basic structure of the network is fully connected. The main task of the PINN model is to generate page volume data from the input annotation data. Based on the PINN theory, automatic differentiation is used in the fully connected network node. The automatic differentiation for inputs (X, Y, Z) constitutes the Laplace equation. This part constitutes the physical rules written into the loss function along with the product of the α factors.

Automatic differentiation techniques are often used in fluid dynamics calculations, and owing to the presence of gradients as derivative forms in deep learning, automatic differentiation becomes a general technique similar to backpropagation. The wide application of automatic differentiation in machine learning is discussed by Baydin et al. [3]. Many practical automatic differentiation tools have been developed for use in various computer languages [1] [32] [9]. We used the automatic differentiation toolkit developed by Lu et al. for our study, because it has a Python module for PINN for solving partial differential equations [23].

The second part of the proposed deep learning model is the page-extraction model, whose underlying structure is the fully connected network. The main task of the page-extraction model is to calculate the ground-truth SSIM scores from the input scalar values (page numbers). Previous studies have found that the best-extracted scalar values (page numbers) are often not integer values owing to annotation errors. For example, the iso-surface where 3.12 is located extracts a better page than 3.00. In previous studies, researchers manually found the best scalar value (page number); conversely, we use a deep-learning model based on the SSIM score as an evaluation criterion to find the best iso-surface. Notably, the page extraction model is required if ground truth test data are available. For real historical document data, only the results of the PINN model need to be texture mapped to extract the page information.

As shown in the Fig.4, in the page segmentation model, the input is annotated data (3D spatial coordinate points: x, y, z and their corresponding page values: U , where N represents the number of network nodes.) as training data. After the training, all the coordinate points in the whole 3D space are imported as test data to obtain the scalar field. By extracting the iso-surfaces in the scalar field, the corresponding pages are extracted. Thus, the segmentation of pages is achieved. The page extraction model is designed to correctly select the iso-surface corresponding to the page. In the presence of ground truth, P equivalence surfaces are randomly selected (P is the actual number of pages) and the average SSIM score is calculated. These data are imported into the page extraction model as training data, and P best U -values can be obtained by training. represents the best iso-surface. This is to solve the problem that U is not an integer value due to annotation error.

3.1 Segmentation model

Deep learning can be used to fit the spatial page structure; however, using only deep-learning networks leads to problems such as incoherence and significant errors in the fitted structure space. In previous studies, the Laplace equation, namely, the assumption that “two adjacent pages do not overlap,” was necessary to solve this problem. Therefore, we chose a PINN, a promising neural-network model for solving scientific computing problems [24]. The Laplace equation is a partial differential equation that we would like to utilize among deep-learning models. Further, one of the features of a PINN is that the model contains relevant physical laws (i.e., a priori information) and acts as a regularization term to improve the training process.

The neural network is based on traditional regression and can be constructed as a fully connected neural network with multiple hidden layers. This part is called the agent part. The physical information component uses the regression results of the neural network on the input data. Using automatic differentiation techniques, the partial differentiation terms can be computed once or several times, and the Laplace equation describes the laws of physics in the model. The details are as follows:

1. The fully connected network can output a loss function by solving the optimization problem; the loss function is MSE_u (mean square error).

2. The inputs and outputs together form the partial differential equation, in this case, the Laplace equation. Here another loss function can be obtained, i.e., MSE_f .

3. $MSE = (1 - \alpha) \times MSE_u + \alpha \times MSE_f$ is the loss function for the entire model, and the training process is run by minimizing. The equation for the loss function consisting of the physical law component is as follows:

$$MSE_f = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \quad (1)$$

The optimizer used in the steps is Adam [19]. In addition, one of the unique designs is to embed the automatic differentiation technique into the loss of the neural network [23]. We, in turn, used this technique to constitute the PINN network.

3.2 Extraction model

We used a $128 \times 64 \times 32$ MLP model to extract pages. In previous studies, owing to annotation errors, the user had to manually specify the iso-surface where the extracted pages were located (the result was not an integer value). Here, we extract the page corresponding to the random iso-surface and compare the SSIM with the actual page (ground truth) to calculate the score. The page-extraction model automatically obtains the best iso-surface by maximizing the score rather than manually specifying it by the user.

In particular, we perform page extraction with a scalar value of 0.03 in the page volume data. $U = 0.03$ is the smallest value of voxels across the CT booklet volume data. This value varies with the specific data and can be calculated from the following equation:

$$U = \frac{(U_{max} - U_{min})}{Resolution(z)} \quad (2)$$

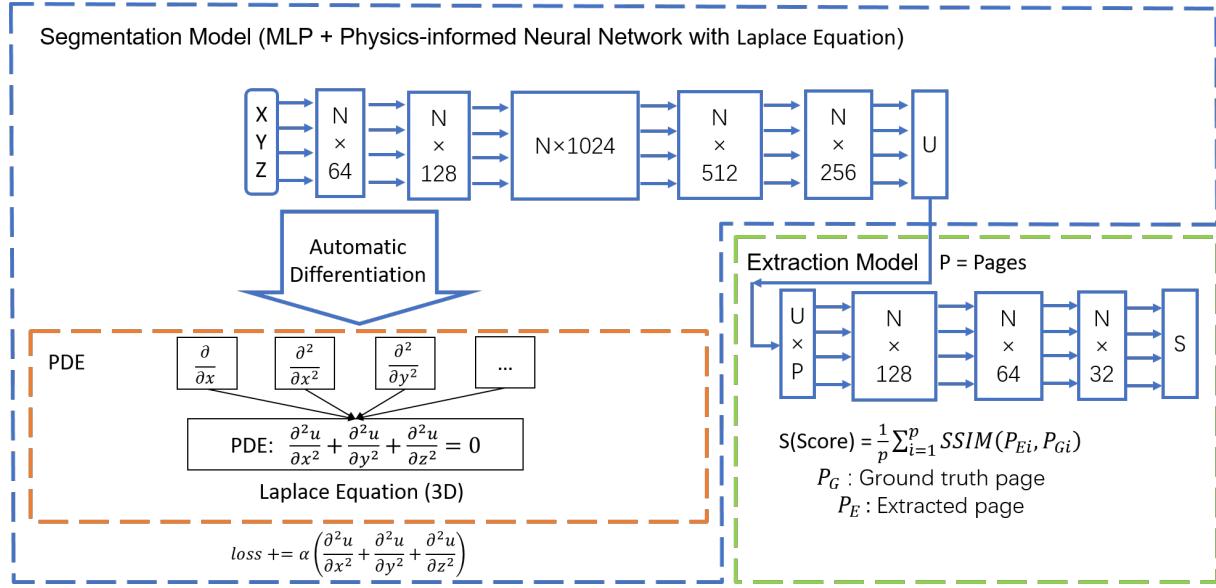


Fig. 4. Overview of the network. The model is divided into the page-segmentation model (PINN model) and the page-extraction model. The PINN model as the first part is based on Laplace equation, and the page-extraction part is an automatic extraction model based on the SSIM score. The input to the page extraction model is the number of pages corresponding to the output of the PINN model. The basic structure of both models is a fully connected network. The automatic differentiation function uses the PINN to apply the physical laws (Laplace equation here) to the deep-learning model.

Where U_{max} is the maximum value in the page volume data, U_{min} is the minimum value in the page volume data, and $Resolution(z)$ is the resolution in the z-axis direction of the CT-generated volume data (In general the z-axis in the direction of the change in the number of pages in the book, whichever is the actual case).

These extracted pages (which may also not be in the page area, at this point, the non page area) are generated by data mapping to 2D images. 2D images are scored against ground truth for SSIM calculation, which generates the training set for the page extraction model.

As shown in the Fig.4, the input to the model is the scalar value of the page volume data corresponding to the number of pages of the booklet, ($U_{p1}, U_{p2} \dots$). The output S of the model is the average score of all pages, with the following equation.

$$S(\text{Score}) = \frac{1}{P} \sum_{i=1}^P \text{SSIM}(P_{Ei}, P_{Gi}) \quad (3)$$

where P_{Ei} is the extracted page and P_{Gi} is the ground truth page. A network trained to extract the highest score, i.e., the best scalar value, can be selected (page number).

4 EXPERIMENT

4.1 Annotation

This section details the annotation tool, including the experimental and page extraction steps and the experimenters' opinion of the tool. In the first version [30], the annotation tool obtained the 3D coordinates of the mouse's current position by double-clicking the mouse and then assigning page numbers to the coordinate data. Later, to improve the efficiency of the annotation, we added the function of range annotation. In particular, with the mouse's position as the sphere's center, we displayed a sphere with an adjustable radius. With a double-click of the mouse, all points on pages in the sphere were annotated as being on the same page. Depending on the sphere's radius, a double click can annotate 7 to 100 data points, which is much more efficient than the previous version. However, more annotated points are not necessarily better. After many experiments, we deduced the following annotation protocols:

- Ensure that the annotation points are evenly distributed in each part of the volume data.

- Ensure that there are annotation points on the boundary of the volume data.

Fig.5 shows the distribution of points annotated in the bending page dataset in space. The points that need attention in the annotation rules are evenly distributed on the page.

The annotated points obtained during this process remarkably impact the results of subsequent experiments. In previous studies, obtaining the correct page-volume data from the Laplace equation model was difficult for poor-quality annotated data (Dataset 2). Moreover, this problem was solved by the PINN model, and we show this comparison in the results section.

4.2 Training

Under the deep-learning model, we still use the data obtained based on annotation tools for learning, such that the effects of two different methods can be compared for the same data. All experiments in this study use Geforce GTX 1060, 6G memory GPU for their calculations.

We used the same annotated data to perform experiments comparing the Laplace equation (previous method) with the PINN model. The experimental data are shown in Table1. We used two sets of annotated data in Ou et al. [30]: Dataset 1 and Dataset 2. Training from the two datasets revealed that the time to perform deep learning is much longer than the time to compute the Laplace equation. This time also becomes longer as the annotated dataset increases. However, the absolute error of deep learning is much smaller than the error of the Laplace equation (using the mean squared error). The reduced error helps to segment more dense pages.

4.3 Mapping

As indicated in the overview, page extraction through page volume data generated by the deep-learning model requires CT voxels to be mapped onto the iso-surface. Usually, we first extract the 3D coordinates of the corresponding iso-surface in the page volume data, which typically consists of several million points. Then, the gray values of the corresponding coordinates in the CT volume data are plotted onto the iso-surface to obtain a 3D visualization of the extracted page. Owing to the annotation error in the generated page volume data, the correct pages are not obtained when page extraction is performed through

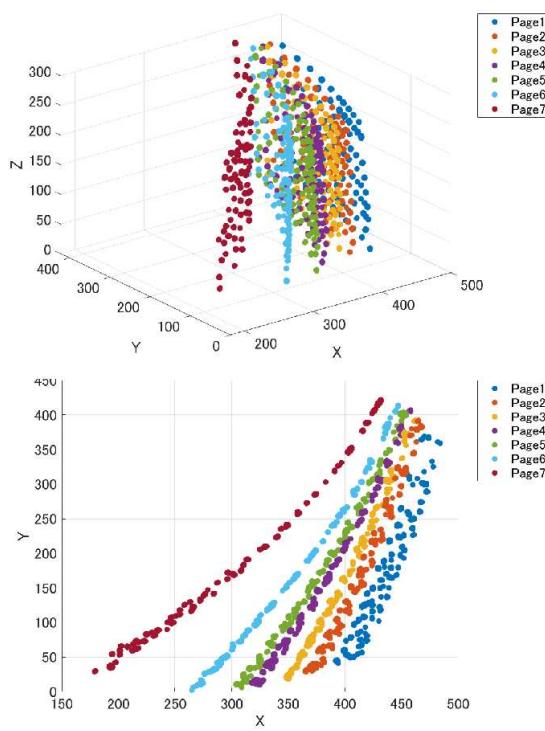


Fig. 5. In the experiment, the points (pages 1 to 7) obtained through annotations are evenly distributed in space (x , y , z are the number of slices of the image)

the iso-surface of integer values. Here, The annotation error is the mean squared error between the generated page volume data and the annotated data.

We proposed the range-extraction method in Ou et al. [30] to solve this problem. An extraction range (twice the calculation error) was specified based on the calculation error to extract the points in the volume data for the range. Now we build a page extraction model to assist the page-extraction process, making it possible to calculate the optimal extraction value U automatically.

In addition, we consider that a page has two sides: back and front, and that the front side is printed in ink. Therefore, when using CT machines, attention must be paid to the resolution to ensure the thickness of each page containing at least two voxels. Then, it is better to visualize the iso-surface by offsetting half the thickness of the page.

Table 1. Comparison of PINN model and Laplace equation

Dataset 1	The Laplace equation	PINN Model
Annotations	3511	3511
Time	5min	36min46sec
Train error	0.41	0.0568
SSIM (average)	0.408	0.894
Dataset 2		
Annotations	3458	3458
Time	5min	36min11sec
Train error	0.35	0.0487
SSIM (average)	0.259	0.876

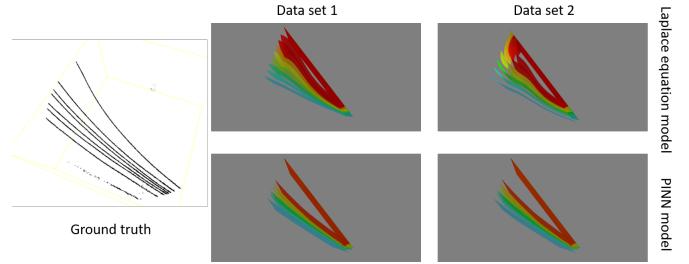


Fig. 6. visualization of iso-surfaces for page volume data generated on the basis of Laplace equation and PINN model. The color from blue to red indicates the value of the iso-surface from 1.0 to 7.0. (Left) The page boundaries shown by the CT images opened in the annotation tool. (Right) The four graph distributions are the page volume data generated by the Laplace equation model and the PINN model using Dataset 1 and Dataset 2

5 RESULT

5.1 Bending page dataset

Fig.6 presents the results of the two models in two sets of experiments. We generated two new page volume datasets for comparison with the page volume data generated by the Laplace equation [30]. The results show that all generated page volume data correspond to the original pages in their basic shapes, indicating that the method of replacing the Laplace equation with our model is feasible. By comparing the extraction results of separate pages, we can evaluate the Laplace equation and compare it with the PINN model approach. The results show that all generated page volume data correspond to the original pages in their basic shapes. In other words, the Laplace equation approach presents different curvatures from the original data at places where circles are drawn in the figure, whereas the PINN model obtains the same shape as the original data at all these places.

Here, we will show the 3D results of the pages extracted by the deep-learning method and compare them with the results of the Laplace equation. By comparing the extraction results of separate pages, we can compare the approaches of the Laplace equation and the PINN model. As shown in Fig.7, page 3 was extracted using the Laplace equation and the PINN model in Dataset 1. The results show that, in the case that the Laplace equation does not fit the page curve completely, the PINN model achieves a better fit; then, the page is complete, and the text is clear. In Dataset 2, page 6 was extracted using the Laplace equation and the PINN model. If the Laplace equation cannot fit the page correctly, the PINN model solves the problem and extracts a complete page illustrating the PINN model's ability in generalizability.

Additionally, in Fig.7, when using the same data (Dataset 2), the Laplace equation model has a significant error in fitting the space equation. The shape of the page has been deformed, and some areas do not match the shape of the original data. However, in the PINN model, this problem is solved very well. Moreover, the overall shape is smooth and conforms to the spatial structure of the original data. On page 6, the Laplace equation model has problems fitting the edge pages because the blank data on the periphery of the page have a greater impact on it. Simultaneously, the PINN model was not affected much and fitted the page structure correctly.

5.2 Booklet dataset

A4 data consisting of twenty single-sided printed pages were used for the experiment. Twenty tightly stacked A4 pages constituted a volume data of $432 \times 275 \times 438$. As shown in Fig8, the volume data were presented in a U shape to test whether the model could extract similarly shaped book data. The first and last pages were annotated during the annotation process as there were no gaps between the pages. To ensure the resolution of the data, we can only extract part of an A4 sheet, as shown on the right side in Fig.8. The distribution of pages in the dataset is simple, and the text in it is successfully extracted.

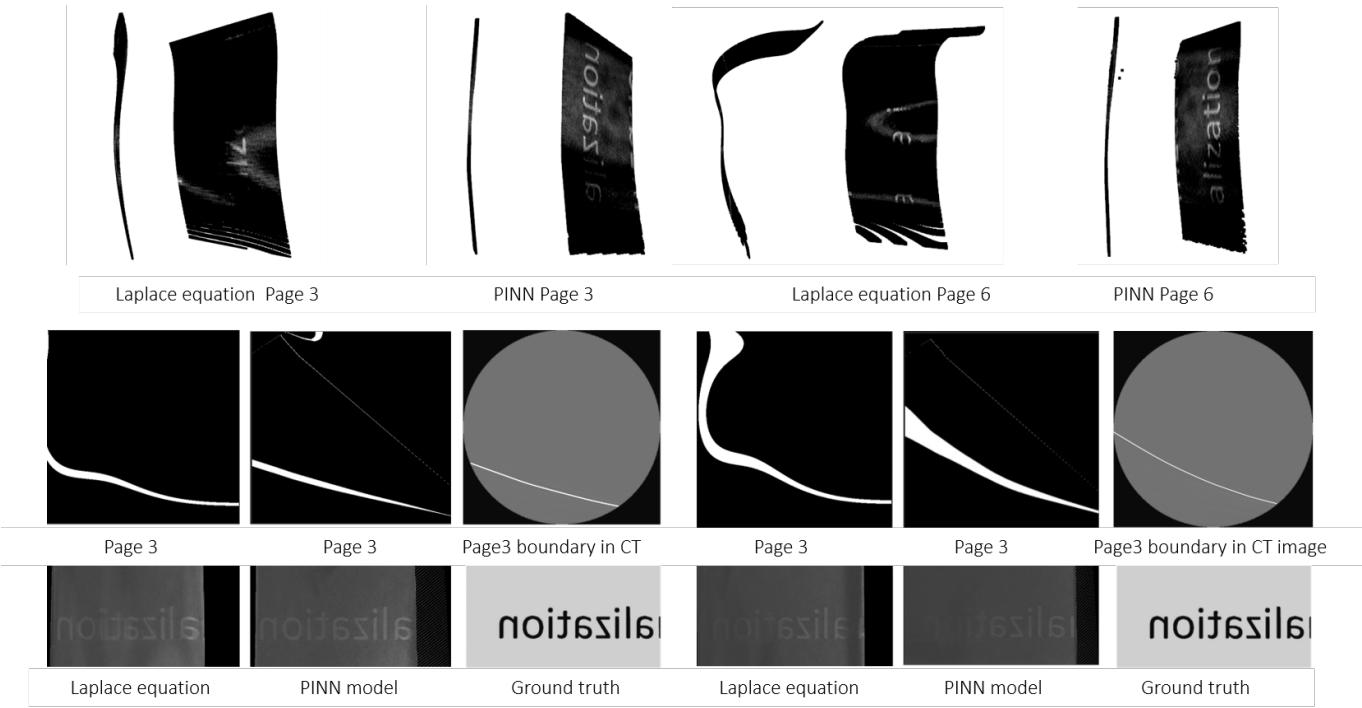


Fig. 7. (Top left) 3D visualization results from Dataset 1 using the Laplace equation to extract page 3, where only some words are recognizable owing to incomplete shape fitting. The 3D visualization result of page 3 is extracted using the PINN model in Dataset 1, and the complete page is extracted from the shape, and original data fit. (Top right) 3D visualization results from Dataset 1 using the Laplace equation to extract page 6, where only some words are recognizable owing to incomplete shape fitting. The 3D visualization result of page 6 is extracted using the PINN model in Dataset 1, and the complete page is extracted from the shape, and original data fit. (Bottom left) When using the Laplace equation model and PINN model to extract page 3, the PINN model achieves a better fit (clearer display of text in the page-extraction results of the PINN model.) (Bottom right) When using the Laplace equation model and PINN model to extract page 6, the PINN model can deal with deformation problems

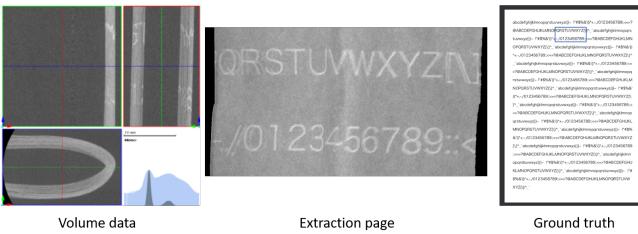


Fig. 8. The dataset consisted of 20 pages with a resolution of $432 \times 275 \times 438$ and a voxel size of 0.0538 mm . The extracted page is complete, and the text on it is clearly visible

We also used a dataset containing 50 pages for the experiment to test whether near-real modern books are available for page extraction. Of course, owing to the spatial resolution of the CT machine, data can only be obtained for some areas. The data were composed of 50 double-sided sheets of standard A4 paper. The content is a mixture of Chinese, Japanese, and drawings. CT data are shown in Fig.9.

We attempt to restore the most complex page extraction scenario, with many pages, no gaps between pages, and text on both sides of each page. The experimental results are unsatisfactory. Although we can fit the shape of the pages, it is not easy to extract each page completely because each page contains a few pixel points in thickness and the overall resolution is not high. The extracted pages, which contain both front and back information, are difficult to subdivide in terms of thickness.

In addition, this type of dataset faces a great challenge in manual annotation. It is difficult to distinguish each page with the naked eye, and we can only select recognizable pages for annotation. This also makes subsequent page extraction challenging.

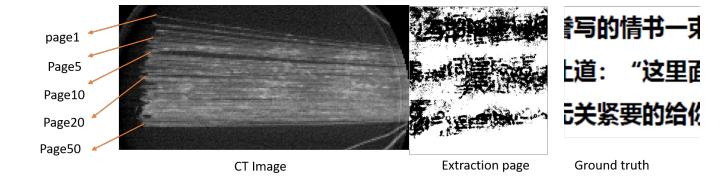


Fig. 9. The dataset consisted of 50 pages with a resolution of $970 \times 489 \times 976$ and voxel size of 0.0499 mm . Even if the thickness of the A4 paper reaches 0.1 mm , it is difficult to identify the position of each page from the figure. The results of page extraction (binarization-processed) contain superimposed text information that is difficult to identify

5.3 Real historical document dataset

In addition to the previous data for methodological comparison, we used new real historical data for validation experiments. The historical document is French from the 1800s and measures $29.5\text{cm} \times 21\text{cm}$. We extracted the pages of the book based on its closure through experiments. CT data are shown in Fig.10. By using our proposed model, the fitted page volume data are identical to the actual page structure in terms of shape. As shown in Fig.11, comparing two selected pages with the ground truth reveals that text was successfully extracted. Because the text is handwritten, we sent this result to an ancient document visualization expert for evaluation. We also used SSIM to calculate the similarity between the extracted pages and ground truth, and the results were used to optimize the page-extraction model.



Fig. 10. French instrument from the 1800s, containing six pages, which we extracted without opening. After opening, it was compared as ground truth with the results of page extraction

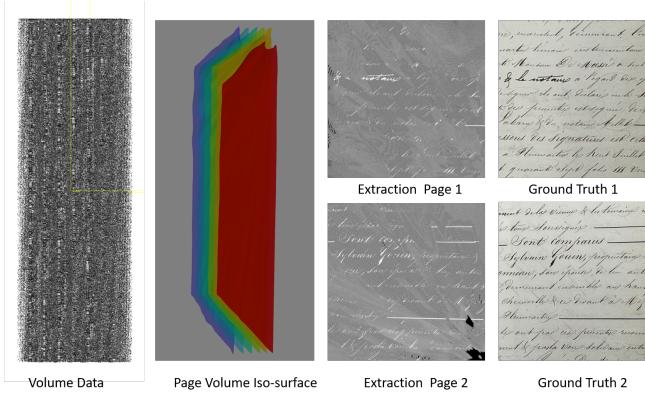


Fig. 11. The generated volume data, the page volume data, and two of the pages of results extracted were compared to the ground truth.

6 EVALUATION

6.1 Number of annotation data

To explore the effect of the number of annotated data points on the model, we conducted experiments using multiple sets of annotated data from the bending page dataset. We conducted seven experiments using different numbers of annotation points (from 500 to 3500, in increments of 500). As shown in Fig.12, the final training loss gradually decreases as the number of data points increases. Meanwhile, we found that the loss is no longer significantly reduced after 4000 times, indicating a low learning efficiency.

As shown in Fig.13, depending on the number of data points used in the experiment, the loss no longer changes significantly at approximately 3500. This gives the user an idea of the number of annotation points needed when using the annotation tool.

6.2 α factor

We verified the effect of the α factor on the model in the PINN through multiple controlled experiments. Finally, a suitable α factor was obtained to add to the loss function. The α factor represents the magnitude of the effect of the physical rules (in this case, the Laplace equation) on the loss function of the PINN model. The settings of the α factors differ in studies related to different PINN models. We conducted 11 experiments on a pamphlet dataset with the α factor set to (0, 0.1, 0.2 ... 1.0). The network was simple and fully connected at the moment when the α factor was 0. The results are shown in Fig.14. Finally, for the optimal α factor, the result is different on different data. The appropriate value must be selected when performing page extraction on different datasets. When the α is 1, the deep learning model at this point is similar to the method that uses only Laplace equation.

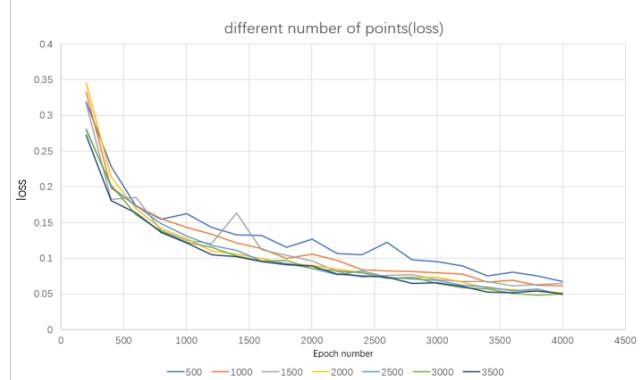


Fig. 12. The number of points in the annotated data ranges from 500 to 3500 for a total of seven sets of experiments, and there is a decline in the loss at the same epoch.

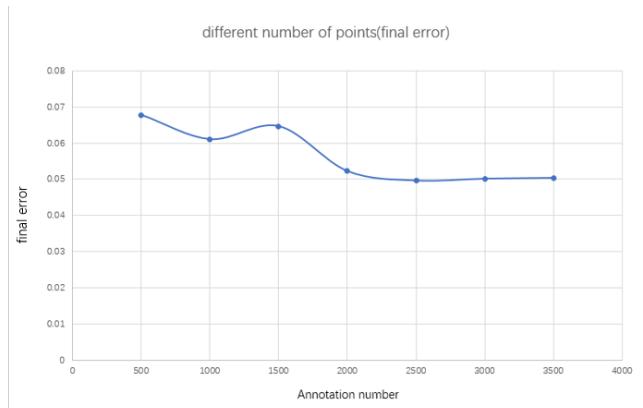


Fig. 13. The number of points for the annotated data ranges from 500 to 3500 for a total of seven sets of experiments, and the final error shows that the error no longer varies significantly when the number of points is around 3000.

6.3 Image

In the Ou et al. [30], Optical Character Recognition (OCR) was chosen as the evaluation criterion for extracting pages [30], but this was limited to the standard font. In this study, the fundamental historical data is in ancient handwriting, and the extracted pages are partially blurred, thereby difficult to recognize.

Therefore, in the page extraction model, we use SSIM as a quantitative evaluation criterion for judging the quality of page extraction. We perform page extraction on the iso-surface output by the model and obtain the SSIM score, as shown in the Table2.

Our current research focuses on ways to reduce the error of manual annotation and achieve the extraction of complete pages. It is clear from the SSIM results that the conventional image evaluation criterion is suboptimal. However, this is extracted from the CT images without processing. In historical literature-related studies, the process of image denoising, binarization, etc., can greatly improve the SSIM value and even reach the standard of OCR. For example, the study by Su et al. explores the effects of a subsequent processing [46].

Table 2. SSIM of page extraction result of each dataset

Dataset	bending page	20 pages	50 pages	real historical
SSIM (average)	0.876	0.6534	0.2218	0.5176

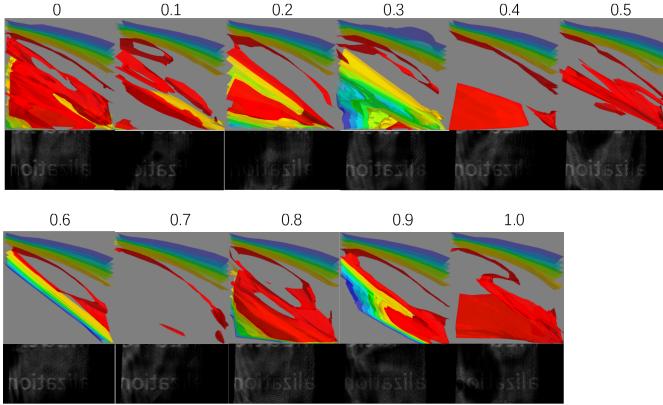


Fig. 14. Effect of different α factors on the PINN model (page volume data). When the PINN model is not used, a portion of the page volume data appears unfitted. As the α factor changes, gradually the spatial distribution of the pages can be fitted correctly. For the current experimental data, 0.7 can be considered as the optimal value. When the alpha value is equal to 1, according to the loss function formula, the model at this point can be seen as a fit using only the Laplace equation.

7 DISCUSSION

This study uses a deep-learning model to improve the Laplace equation model. The deep-learning-based page-extraction technique presented in this paper improves the accuracy of page extraction compared to previous methods. The proposed PINN and extraction model combines deep learning while maintaining the properties of the Laplace equation to achieve automatic page segmentation and extraction. We intend to use more data types for analysis and further research in the future. This study aims to provide more accurate visualization results using AI techniques. Further, we intend to provide a case study for future research on book-shaped page extraction.

The content and results of our research were evaluated by visualization and ancient literature research experts. The experts stated that it is pretty reasonable and necessary to hypothesize that “page surfaces of the booklet could be extracted by one scalar field describing them as iso-surfaces.” Nevertheless, because the scalar field generated by the Laplace equation model is the input of iso-surface extraction, it is crucial to ensure this model is the best one for our problem, allowing us to reduce or avoid the expensive manual annotation process. We believe this method will help improve the optimization of the final results.

“Regarding usability and effectiveness, I think your results are good. As we know, among the millions of ancient books in the world, many of them are in danger. Furthermore, in these books, most of them are fragile and owing to conservation aspects, we cannot unfold them because of their preciousness and uniqueness, or even if we can unfold them, in most cases, the contents will be destroyed during the unfolding process. You provide a general and nonintrusive method to obtain the contents of books; it may be the only feasible way for the digitalization practice of those fragile ancient books.”

As experts have commented, the annotation process significantly impacts page-extraction results. During the experiments, the page volume data obtained from training often could not fit the actual page space structure because of the inadequate annotated data. We had to improve the annotated data several times to achieve good results. This problem is severe in the case of many pages and complex page structures. We used a deep-learning approach to solve this problem, but owing to the small amount of historical book data and the large gap between each dataset, simple machine learning can quickly produce overfitting, thereby requiring further investigation.

8 CONCLUSION

We proposed a process of page extraction using a PINN model to replace the previous method and improve accuracy. A page-extraction model is also proposed to achieve quantitative evaluation and automatic

extraction. The principle is to approximate the page volume data from spatially annotated points utilizing a deep neural network. We have experimentally demonstrated that the proposed page-extraction method for generating page volume data through a deep-learning network is effective, and we have evaluated the annotation data and the α factor through multiple sets of experiments. We visualized the extracted pages and evaluated them quantitatively by SSIM. Experts have evaluated the study’s results as excellent, contributing to the subsequent research about page information extraction from historical documents in the direction of deep learning. In future research, we intend to conduct further experiments to reduce the annotation process and optimize deep-learning neural networks.

ACKNOWLEDGMENTS

The authors wish to thank A, B, and C. This work was partly supported by a grant from XYZ (# 12345-67890).

REFERENCES

- [1] J. Andersson, J. Åkesson, and M. Diehl. Casadi: A symbolic package for automatic differentiation and optimal control. In *Recent advances in algorithmic differentiation*, pp. 297–307. Springer, 2012.
- [2] D. Baum, N. Lindow, H.-C. Hege, V. Lepper, T. Siopi, F. Kutz, K. Mahlow, and H.-E. Mahnke. Revealing hidden text in rolled and folded papyri. *Applied Physics A*, 123(3):1–7, 2017.
- [3] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, 18:1–43, 2018.
- [4] U. Bergmann. Archimedes brought to light. *Physics World*, 20(11):39, 2007.
- [5] U. Bergmann, P. L. Manning, and R. A. Wogelius. Chemical mapping of paleontological and archeological artifacts with synchrotron x-rays. *Annual Review of Analytical Chemistry*, 5:361–389, 2012.
- [6] C. Cheng, H. Meng, Y.-Z. Li, and G.-T. Zhang. Deep learning based on pinn for solving 2 dof vortex induced vibration of cylinder. *Ocean Engineering*, 240:109932, 2021.
- [7] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- [8] D. Diringer. *The book before printing: ancient, medieval and oriental*. Courier Corporation, 2013.
- [9] L. Hascoet and V. Pascual. The tapenade automatic differentiation tool: principles, model, and specification. *ACM Transactions on Mathematical Software (TOMS)*, 39(3):1–43, 2013.
- [10] A. Heinecke, J. Ho, and W.-L. Hwang. Refinement and universal approximation via sparsely connected relu convolution nets. *IEEE Signal Processing Letters*, 27:1175–1179, 2020.
- [11] Y. Higo, F. Oka, T. Sato, Y. Matsushima, and S. Kimoto. Investigation of localized deformation in partially saturated sand under triaxial compression using microfocus x-ray ct with digital image correlation. *Soils and Foundations*, 53(2):181–198, 2013.
- [12] A. Hore and D. Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pp. 2366–2369. IEEE, 2010.
- [13] A. Horng, E. Brun, A. Mittone, S. Gasilov, L. Weber, T. Geith, S. Adam-Neumair, S. D. Auweter, A. Bravin, M. F. Reiser, et al. Cartilage and soft tissue imaging using x-rays: propagation-based phase-contrast computed tomography of the human knee in comparison with clinical imaging techniques and histology. *Investigative radiology*, 49(9):627–634, 2014.
- [14] K. Hu, K. Koyamada, H. Ohtani, T. Goto, and J. Miyazawa. Visualization of plasma shape in the lhd-type helical fusion reactor, ffhr, by a deep learning technique. *Journal of Visualization*, pp. 1–14, 2021.
- [15] K. Hu, Q. Wang, K. Koyamada, H. Ohtani, T. Goto, and J. Miyazawa. Visualization of the plasma shape in a force free helical reactor, ffhr. *Journal of Advanced Simulation in Science and Engineering*, 7(1):151–167, 2020.
- [16] G.-B. Huang, L. Chen, C. K. Siew, et al. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Networks*, 17(4):879–892, 2006.
- [17] M. A. Hubbe and C. Bowden. Handmade paper: A review of its history, craft, and science. *BioResources*, 4(4):1736–1792, 2009.
- [18] S. Jadhav, K. Dmitriev, J. Marino, M. Barish, and A. E. Kaufman. 3d virtual pancreatography. *IEEE transactions on visualization and computer graphics*, 2020.

- [19] I. K. M. Jais, A. R. Ismail, and S. Q. Nisa. Adam optimization algorithm for wide and deep neural network. *Knowledge Engineering and Data Science*, 2(1):41–46, 2019.
- [20] M. Knoche. The herzogin anna amalia library after the fire. *IFLA journal*, 31(1):90–92, 2005.
- [21] J. Kolar, A. Štolfa, M. Strlič, M. Pompe, B. Pihlar, M. Budnar, J. Simčič, and B. Reissland. Historical iron gall ink containing documents—properties affecting their condition. *Analytica chimica acta*, 555(1):167–174, 2006.
- [22] R. P. Lippmann. Pattern classification using neural networks. *IEEE communications magazine*, 27(11):47–50, 1989.
- [23] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis. Deepxde: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021.
- [24] S. Markidis. The old and the new: Can physics-informed deep-learning replace traditional linear solvers? *Frontiers in big Data*, p. 92, 2021.
- [25] L. Meng, D. Dong, X. Chen, M. Fang, R. Wang, J. Li, Z. Liu, and J. Tian. 2d and 3d ct radiomic features performance comparison in characterization of gastric cancer: a multi-center study. *IEEE journal of biomedical and health informatics*, 25(3):755–763, 2020.
- [26] V. Mocella, E. Brun, C. Ferrero, and D. Delattre. Revealing letters in rolled herculaneum papyri by x-ray phase-contrast imaging. *Nature communications*, 6(1):1–6, 2015.
- [27] D. A. Moreira, C. Hage, E. F. Luque, D. Willrett, and D. L. Rubin. 3d markup of radiological images in epad, a web-based image annotation tool. In *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, pp. 97–102. IEEE, 2015.
- [28] D. T. Nguyen, B.-S. Hua, L.-F. Yu, and S.-K. Yeung. A robust 3d-2d interactive tool for scene segmentation and annotation. *IEEE transactions on visualization and computer graphics*, 24(12):3005–3018, 2017.
- [29] M. A. Nielsen. *Neural networks and deep learning*, vol. 25. Determination press San Francisco, CA, 2015.
- [30] J. Ou, Z. Han, and K. Koyamada. Three-dimensional book data page segmentation and extraction method using laplace equation. *Journal of Advanced Simulation in Science and Engineering*, 8(2):223–236, 2021.
- [31] J. Park and I. W. Sandberg. Universal approximation using radial-basis-function networks. *Neural computation*, 3(2):246–257, 1991.
- [32] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. 2017.
- [33] D. Perekrestenko, P. Grohs, D. Elbrächter, and H. Bölcskei. The universal approximation power of finite-width deep relu networks. *arXiv preprint arXiv:1806.01528*, 2018.
- [34] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*, 2017.
- [35] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics informed deep learning (part ii): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10566*, 2017.
- [36] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.
- [37] D. L. Rubin, P. Mongkolwat, V. Kleper, K. Supekar, and D. S. Channin. Annotation and image markup: accessing and interoperating with the semantic content in medical imaging. *IEEE Intelligent Systems*, 24(1):57–65, 2009.
- [38] O. Samko, Y.-K. Lai, D. Marshall, and P. L. Rosin. Virtual unrolling and information recovery from scanned scrolled historical documents. *Pattern Recognition*, 47(1):248–259, 2014.
- [39] U. Sara, M. Akter, and M. S. Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 7(3):8–18, 2019.
- [40] O. Schabenberger and C. A. Gotway. *Statistical methods for spatial data analysis*. CRC press, 2017.
- [41] B. Seales and D. Delattre. *Virtual unrolling of carbonized Herculaneum scrolls: Research Status (2007-2012)*. Macchiaroli editore, 2013.
- [42] W. B. Seales, C. S. Parker, M. Segal, E. Tov, P. Shor, and Y. Porath. From damage to discovery via virtual unwrapping: Reading the scroll from en-gedi. *Science advances*, 2(9):e1601247, 2016.
- [43] A. Stijnman. Historical iron-gall ink recipes: art technological source research for inkcor. *Papierrestaurierung*, 5(3):14–17, 2004.
- [44] D. Stromer, V. Christlein, Y. Huang, P. Zippert, E. Helmecke, T. Hausotte, and A. Maier. Dose reduction for historical books digitization by 3-d x-ray ct. In *of Applied Sciences Upper Austria, U.(ed.) Proceedings of 8th Conference on Industrial Computed Tomography (iCT 2018)*, pp. 1–2, 2018.
- [45] D. Stromer, V. Christlein, C. Martindale, P. Zippert, E. Haltenberger, T. Hausotte, and A. Maier. Browsing through sealed historical manuscripts by using 3-d computed tomography with low-brilliance x-ray sources. *Scientific reports*, 8(1):1–10, 2018.
- [46] X. Su, H. Xu, Y. Zhang, Y. Kang, G. Gao, et al. An end-to-end preprocessor based on adversarial learning for mongolian historical document ocr. In *Pacific Rim International Conference on Artificial Intelligence*, pp. 266–272. Springer, 2019.
- [47] H. Wang, Y. Liu, and S. Wang. Dense velocity reconstruction from particle image velocimetry/particle tracking velocimetry using a physics-informed neural network. *Physics of Fluids*, 34(1):017116, 2022.
- [48] D. Yarotsky. Universal approximations of invariant maps by neural networks. *arXiv preprint arXiv:1804.10306*, 2018.
- [49] B. Yu, H. Doraiswamy, X. Chen, E. Miraldi, M. L. Arrieta-Ortiz, C. Hafemeister, A. Madar, R. Bonneau, and C. T. Silva. Genotet: An interactive web-based visual exploration framework to support validation of gene regulatory networks. *IEEE transactions on visualization and computer graphics*, 20(12):1903–1912, 2014.
- [50] D.-X. Zhou. Universality of deep convolutional neural networks. *Applied and computational harmonic analysis*, 48(2):787–794, 2020.