Visualization of 3D booklet CT data based on FCN

Zhongjiang han, Akio doi, Koji koyamada

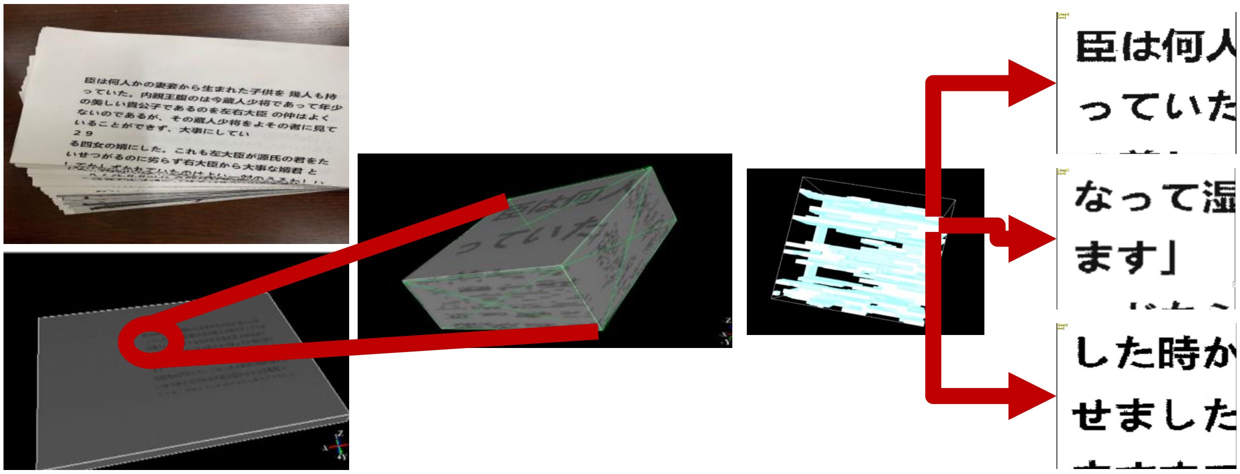

Fig. 1. Page extraction based on FCN framework for deep learning with Japanese text.

Abstract—CT technology has been widely used in medical, geographic, architectural, and other industries. In recent years, significant contributions have also been made to archaeological and ancient cultural relics research. In addition to the medical CT, there have been many studies using CT technology to observe the internal structure without destroying the surface of the object. For example, discover the ore in the rock layer, the mummy in the large Buddha statue, and the digital unwrapping of an ancient parchment. We have found that in the face of fragile paper ancient literature, how to digitize its content on the basis of protecting it from being destroyed, and provide valuable information for scholars becomes a problem. Therefore, we have digitally reconstructed the paper text in which metal-containing inks are applied for the printing process in CT machines, and used FCN-based deep learning to solve the problem of page bending. We experimentally demonstrated the feasibility of obtaining the internal information of the booklet through CT scanning, using deep learning to clear the printed text, and demonstrating the effectiveness of deep learning technology in restoring CT data of affine transformation books.

Index Terms—CT， booklet， page extraction， deep learning， 3DFCN

1 Introduction

Since the birth of CT technology, especially x-ray based CT scan imaging technology has been widely used in various fields. X-ray is an electromagnetic wave with a very short wavelength and a large energy. The wavelength of X-ray is shorter than the wavelength of visible light (about 0.001 to 10 nm, and the wavelength of X-ray applied in medicine is about 0.001 to 0.1 nm). Its photon energy is tens of thousands to hundreds of thousands of times greater than the photon energy of visible light [1], [2]. It was discovered by the German physicist WK Roentgen in 1895, so it is also called the roentgen ray. X-rays have a high penetration ability and can pass through many materials that are opaque to visible light, such as ink paper, wood, and the like [3], [4], [5], [6].

X-rays were originally used for medical imaging diagnostics and X-ray crystallography. X-rays are used in medical diagnosis, mainly based on X-ray penetration, differential absorption, photosensitivity and fluorescence. Since X-rays are absorbed by different degrees when they pass through the human body, for example, the amount of X-rays absorbed by the bones is larger than that of the muscles, the amount of X-rays passing through the human body is different, so that the portable belt has a density distribution of various parts of the human body [7], [8], [9]. The information, the intensity of fluorescence or photosensitivity caused on the screen or on the photographic film, is quite different, so that the shadows of different densities will be displayed on the screen or on the photographic film (developed, fixed). According to the comparison of shadow shading, combined with clinical manifestations, laboratory results and pathological diagnosis, it can be judged whether a certain part of the human body is normal. Thus, X-ray diagnostic technology became the world's first non-scratch visceral inspection technique.

We are inspired by the use of CT devices for visual research because of this characteristic of X-rays. We believe that the CT device can deal with the research problems of not visualizing the internal text on the basis of the ancient literature. In fact, there has been a research on the virtual expansion of the ancient sheepskin roll using the CT machine, which shows that as long as the use is guaranteed The ink contains metallic elements that can be recognized by X-rays to give an image that is different from paper or sheepskin [10].

In fact, the ancient literature represented by ancient China and Japan is different from the West. It is similar to a booklet rather than a roll. The material is different from the animal skin but very thin rice paper, which makes our research necessary from the foundation. At the beginning, whether the huge energy of X-rays is harmful to the paper, the content of metal elements in the ink affects the imaging results, and the CT machine sets the influence of voltage and current (x-ray energy) on the imaging results. In the first phase, we will briefly introduce the experimental design and results in this regard. Next we will introduce the active grid. Method we used before to deal with the problem of page extraction. We will explain the advantages and disadvantages of this approach and the reasons for moving to deep learning. Secondly, we will discuss the simple concept of FCN and the reasons for using FCN as the main means [11], [20].

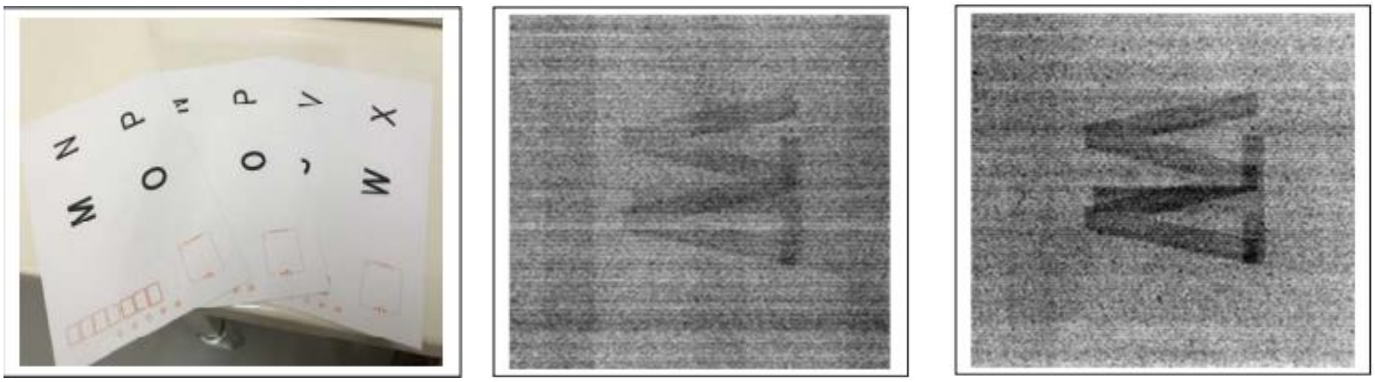

Fig.2. Specially printed cards and their imaging under the X-ray

2 CONVERTING BOOKLET DATA BY X-RAY

In this section, we will explain the experimental preparation and specific processes related to CT machines and X-rays, as well as the related research of active grid and its experimental results in the early stage of this study.

2.1 Experiment of CT machine parameters

We all know that CT technology uses tomography to acquire images of various layers and finally perform three-dimensional reconstruction to restore objects. In order to avoid damage to the real ancient literature before the test, it was decided to use our own booklet for scanning analysis. The experiment used a CT machine of the Shimadzu manufacturer's INSPEXIO SMX-90CT PULS model [17], [19]. The booklet uses a 0.2 mm thick card and uses a high-content ink to print a wide variety of English letters. Experimented on the following two points,

1) the influence factor of x-ray on the transmittance of white paper portion and printed portion

2) the effect of x-ray energy on the resulting image

Experiments were carried out on the voltage, current, number of card stacks, and three parameters to obtain a result map. The conversion range of the voltage parameters is 20kv to 90kv, the range of current parameters is 100μA to 250μA, and the number of cards stacked is from 1 to 30. The experimental result image has a pixel value of 1024×1024, and we extract the brightness value of the pixel therein as a result for evaluation.

Simply by the human eye recognition, the larger the difference in brightness value between the printed portion and the white paper portion, the easier it is to be recognized. The greater the value of the voltage and current, the greater the energy contained in the x-ray, the better the visualization. However, after increasing to a certain extent, the result will not change significantly.

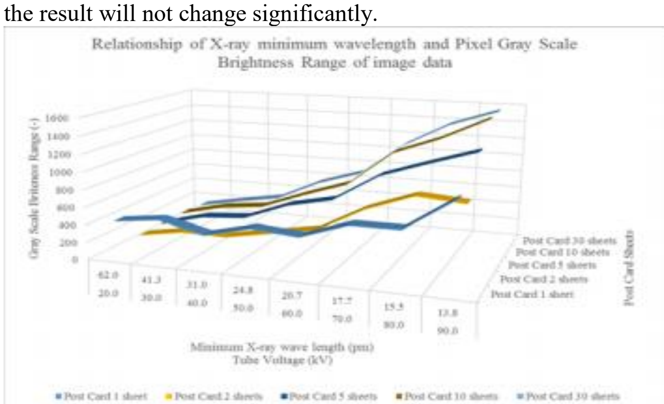

Fig.3. Pixel brightness value changes with CT machine voltage and current

X-ray Tube Voltage (kV)	minimum X-ray wave length λmin (pm)	Gray Scale Brightness Range (-) *	minimum Gray Scale Brightness	max Gray Scale Brightness
20.0	62.0	428	231	659
30.0	41.3	489	1730	2219
40.0	31.0	365	4541	4906
50.0	24.8	487	9849	10336
60.0	20.7	446	17030	17476
70.0	17.7	624	25501	26125
80.0	15.5	624	35576	36200
90.0	13.8	980	44582	45562

Fig.4. Brightness value and x-ray energy relationship (single page)

2.2 Active grid based model extraction

Three-dimensional image data by CT equipment are applied in various places such as surgical simulation and medical diagnosis performed in the medical field. Specifically, extraction of tissues such as bones, internal organs, and epidermis from three-dimensional image data as three-dimensional geometric models is performed by various methods. This experiment aimed at the extraction inside the volume, and was performed by the method called Active Grid [12], [13], [14].

The 3D Active Grid consists of a 3D grid model with internal nodes, and each grid is deformed to match the 3D region inside the volume data. The transformation of the grid is performed by solving the energy minimization problem for the internal energy of the grid which reduces the grid and the image energy inside the 3D region. As the experiment, 557 CT images of 1024×1024 pixels were reconstructed as three-dimensional models as volume data. In order to analyse the constructed three-dimensional model, we conducted an experiment in the following three steps.

Step 1: Refer to the visualized model and extract the booklet model from Active Grid

Step 2: Extract pages from Active Grid from the extracted model

Step 3: Check the accuracy of the extracted page

Step 1 is the extraction of the booklet model. Since the model constructed from 557 CT images is large, and only a part of it (booklet model) is to be studied, clipping of unnecessary parts are performed. The work involved in the next step can be reduced, the size of the model can be reduced, and it can be useful to prove the effect of Active Grid. In Step 2, extracting pages from the extracted booklet model is the most important procedure. The purpose is to adjust the initial calculation range from Active Grid one by one and extract a page completely. Evaluate the extracted results in Step 3. If not enough results, return to Step 2 and start over. This is the stage with the largest amount of work. Step 3 evaluates the extracted page. At this stage, it is not a scientific method, but to evaluate with several features (page bending angle, page printing). From the evaluation at this stage, we will also evaluate the extraction from Active Grid, and evaluate the possibility of the research of data conversion into booklets with this method [15], [16], [18].

The experiment was conducted by the above three procedures, and the evaluation of this study was made from the results. We will examine shortages in research and areas to be improved based on evaluation, and consider future research as useful.

Figure 5 shows that the grid converges to the model from the number of convergence of step1. The white side is the upper side, representing the six sides of the rectangular solid with their respective colors. Change from 500 times to 2500 times, and from 3000 to almost fit the model body. This is not shown here because the number of times is more than 3000 times and the convergence model is more likely to collapse than the external force. Figure 5 compares the final result with the model body. As a result, the number of convergences is 3000 and the following experiment is performed.

Figure 6 shows the results of extracting pages from the booklet model. As we mentioned earlier, we did not show the other results here, as we got the result by adjusting one pixel at a time. Even if only one pixel in the convergence range is different at the beginning, different results will be produced. It is thought that it is useful for this research to produce the best result from the work of repeating Step 2 and Step 3.

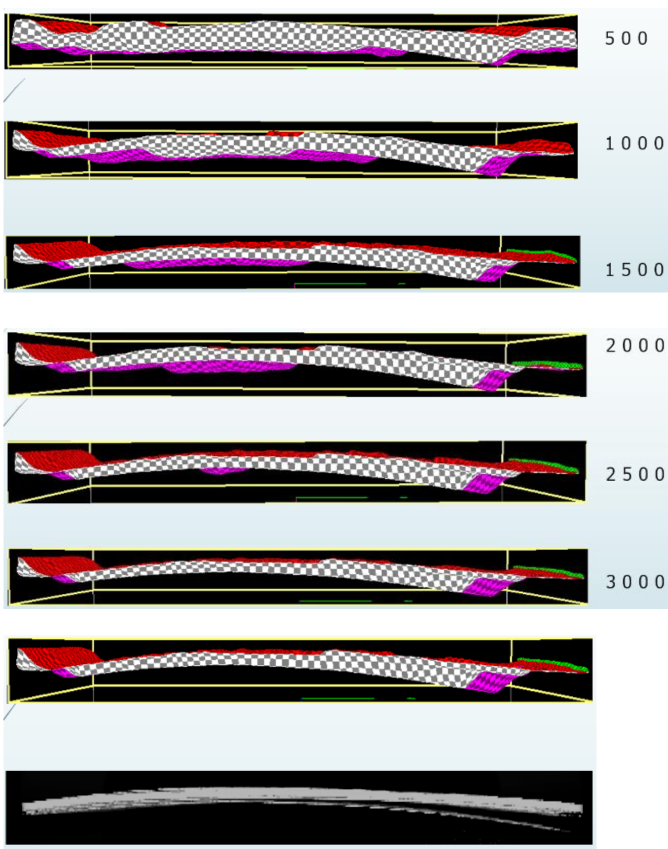

Fig.5. The result of the number of active grids and the final result compared with the model

2.3 Result

In this research, we conducted research on the conversion of booklet data by X-ray using a CT system. We evaluated the influence of the number of pages and the data conversion result by X-ray energy. I would like to consider how far we can go with the current technology. Then, we conducted an experiment on the extraction of the booklet model from Active Grid, the extraction of the page from the model, and the identification of the printing. The conclusion is as follows:

a). It is possible to extract pages from the CT data of the booklet by Active Grid.

b). If the extracted page is reflected on the plane, the contents of the page can be understood.

c). If the number of grids is further increased, the accuracy also increases.

From the experimental process and results, I think there is something to consider. The technical shortage in CT equipment is not considered here because it is not an expert. However, extraction of models and pages is more useful than software technology, i.e., Active Grid, but it is considered too much work. In addition, it is considered that noise correspondence of image data is insufficient, and processing such as noise removal of data is necessary before the experiment.

There are many studies in China on recovering distorted pages. The usual method is to create a large page database by combining a large number of page grayscale data and curved curves. It is then restored by scanning the curved page against the model in the database. Although some studies have shown higher accuracy in the experiment, in fact the database does not contain all the page bending.

We tried to find a general way to deal with all the page bending problems. This is why we decided to use the deep learning to perform 3D model extraction in the next experiment. We will introduce experiments on image processing using 3DFCN in the next section.

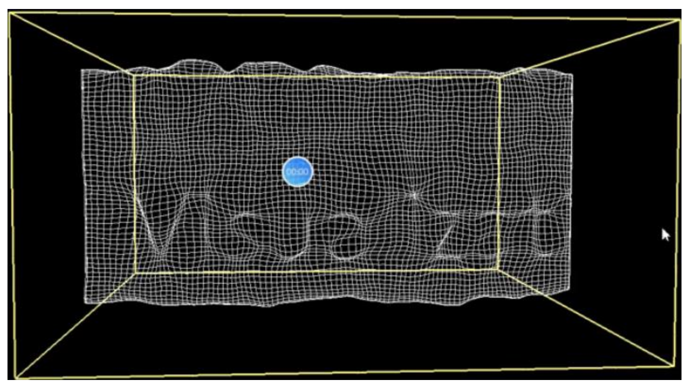

Fig.6. Single page extraction results based on active grid

3 DEEP LEARNING BASED ON 3DFCN

In this section we will explain in detail the principles of 3dfcn and the reasons for using 3dfcn deep learning. Finally, we will give an experimental design and a discussion of the results.

3.1 3DFCN

Deep learning is one of the most popular AI technologies in the world. Especially in the continuous upgrading of computer processing systems, deep learning evolves on the basis of machine learning to solve more complex problems [28], [29]. Especially in the field of image processing, the CNN-based method perfectly solves many computer vision problems such as image segmentation. FCN is improved on the basis of CNN to deal with the recognition and segmentation of 2D images.

Usually, the CNN network is connected to a number of fully connected layers after convolution, and the feature map generated by the convolution layer is mapped into a fixed length feature vector. The general CNN structure is suitable for image level classification and regression tasks because they ultimately expect a probability of classification of the input image.

The FCN classifies the image at the pixel level, thus solving the problem of image segmentation at the semantic level. Different from the classic CNN in the convolutional layer using the fully connected layer to obtain fixed-length feature vectors for classification, the FCN can accept input images of any size, and use the deconvolution layer to up-sampled the feature map of the last volume base layer, making it Reverting to the same size of the input image, one prediction can be made for each pixel while preserving the spatial information in the original input image, and finally the parity is sorted by the up-sampled feature map [30].

The Full Convolutional Network (FCN) recovers the categories to which each pixel belongs from abstract features. That is, the classification from the image level is further extended to the classification at the pixel level [26].

We focus on 3DFCN's ability in the field of image segmentation, using deep learning to automatically solve the page segmentation problem of booklets, which is also a new attempt in the field of visualization. In fact, there aren't many examples of image segmentation with a large number of segmentation targets, especially when we are faced with different pages of booklets, it is our ultimate goal to find a common processing method. We are exploring the possibility of deep learning in solving such problems, and this study will be used as a prior study for later research.

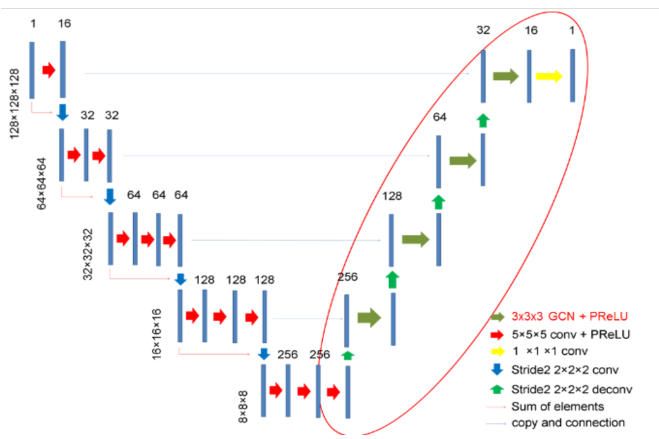

Fig.7. FCN framework used for experiments (convolution and up-sampling)

3.2 Experiment

In this section, we introduce the FCN-based deep learning framework and introduce two experiments: character restoration and restoration of the affine transformation model. The experimental results are illustrated to demonstrate the effectiveness of deep learning in restoring the affine transformation booklet model.

3.2.1 Deep learning framework

In the previous section we explained the basic principles and framework of the FCN. In order to use the FCN to deal with the booklet issues we faced, the FCN framework must be rebuilt. The basic framework of the FCN corresponds to a two-dimensional image and is used to solve the segmentation problem on a two-dimensional image. This does not solve the dilemma. We decided to use the 3DF model segmentation framework 2DFCN which is now widely used in the medical field. In simple terms, 2DFCN refers to the fact that, on the basis of FCN, the input and output are three-dimensional models, but in the computer operation process, the slice is taken for convolution, and the convolution kernel is still a two-dimensional matrix. The advantage of this is that the amount of computation of the computer is greatly reduced, and deep learning operations can still be performed even in the CPU environment [21], [22], [23].

In the actual experiment, we encountered a lot of problems in using 2DFCN, we will explain in the next part, which makes us pay attention to whether 3DFCN can solve these problems. 3DFCN is based on 2DFCN, the computer directly operates on the 3D model, that is, the convolution kernel becomes a three-dimensional matrix, which will cause the CPU to bear a huge amount of computation. Even if we use the GPU as the main operation, we must still input smaller. Model to ensure that memory does not overflow [24], [25].

As shown in Fig. 7, the 3DFCN deep learning framework used in this experiment is finally outputted by convolution and deconvolution with volume data of 128x128x128 size as input.

3.2.2 character restoration

In the previous section we introduced the 3DFCN framework used, which limits the size of our input model. In fact, the smaller the word of the booklet, the more it is necessary to improve the accuracy of the CT machine. Otherwise, under the influence of noise, the resulting image will be blurred and difficult to distinguish. At the same time, the use of high-precision CT machines will increase the image pixels, which will greatly increase the time-consuming of 3D reconstruction, and the reconstructed 3d model data will reach several GB. It is obviously unrealistic to enter the entire model at once [27].

The prepared experimental booklet is made up of 100 sheets of A4 Paper printing has a small word. We experimented with dicing it into small blocks of 128x128x128 to ensure that deep learning can proceed normally. Fortunately, the printed words that are themselves small under such conditions will become clear and discernible.

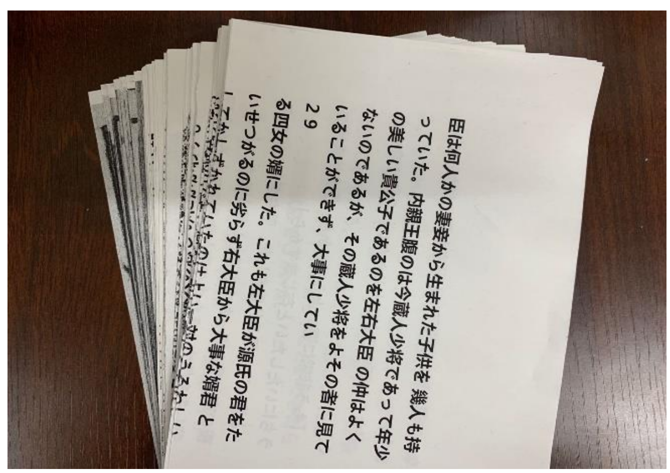

Fig.8. Experimental booklet (A4 100 pages double-sided printing)

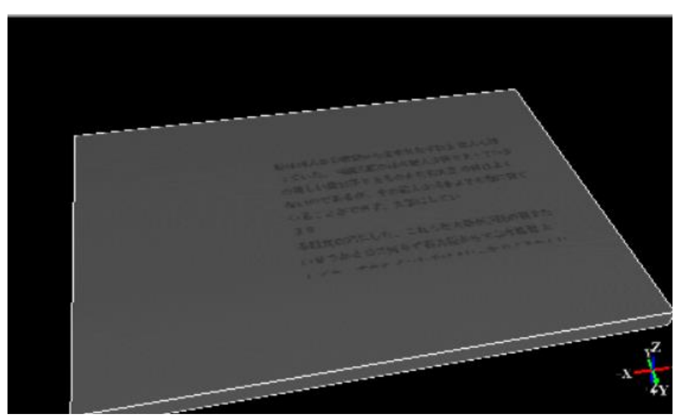

Fig.9. Booklet CT data three-dimensional model

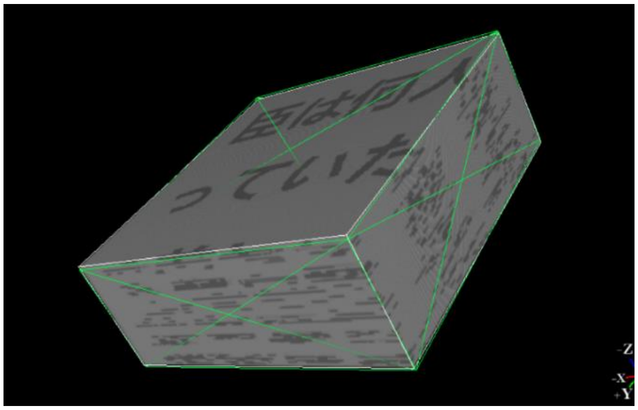

Fig.10 The dicing of the experimental model (size 128×128×128)

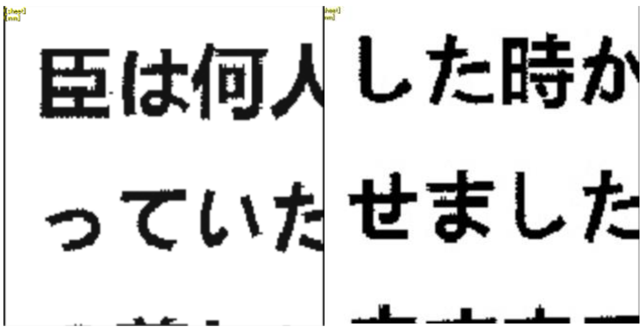

Fig.11 Text recovery and clarity results (two pages)

3.2.3 affine transformation model restoration

During the CT scan process, pages overlap, noise, jitter, etc. may cause the CT image to be distorted, part of the content to be lost, and the brightness of the characters to be inconsistent, etc., which needs to be sharpened.

Using 3DFCN as the basis, we used the clear image of the booklet as the correct data to train the reconstructed booklet 3D model and get a clear slice. Due to the large number of pages, the page is flat when scanning, and the resulting 3D model slice can directly recognize the text, which proves that deep learning can solve the problem of simple book page extraction.

It's not our research purpose to use simple learning to process simple models. We are committed to developing more versatile methods to deal with a wide variety of problems. In the visual study of booklets, the irregular curvature of the page is one of the most common problems. In the previous section, we tried using the active grid to process the curvature of the page, extract the curved page and project it on the plane to get the complete identifiable page. We also discussed the limitations of using active grids, and we also tried using deep learning to solve this dilemma.

As a semi-supervised learning method, FCN needs to provide a considerable number of training data sets and correct data sets. In solving the problem of curved pages, we set up the following experimental schemes for deep learning experiments:

1) Select a small dicing block as the correct data

2) Use the affine transform to coordinate the data, such as 5° in the x direction and 10° in the y direction.

3) Train multiple post-rotation model data to point to the same unrotated correct data, so that the learning process is converged

4) Test the input of other data after the training model

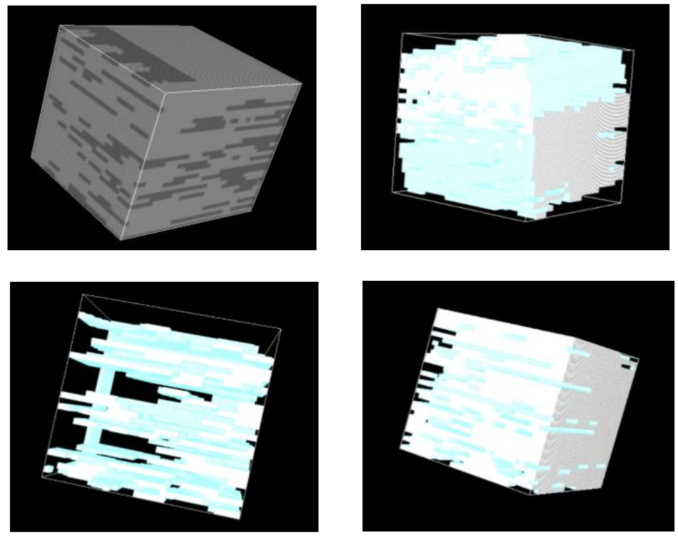

Fig.12. Deep learning data size 32×32×32 (top left)
Affine transformed data (unrecognizable) (top right)
FCN-based page extraction results (identifiable) (bottom left)
Raw data comparison (bottom right)

3.3 result

We have trained more than 100,000 times on 6 sets of data. According to Figure 14, the loss function does not directly receive the beam, but the overall trend of convergence is large, and there is a large fluctuation in a small range. This is because the 6 sets of data are not trained at the same time, but are trained in order, which means that we can get 6 sets of results in order (these results tend to be the same data), which is why the overall learning will be closed. Figure 12 shows the final training results. According to the results, we can determine the basic recovery smoothness of the data, and extract the text part of the original data, which indicates that it is effective and feasible to use the deep learning to restore the book data after the affine transformation. At the same time, Figure 13 shows that the result of the reduction is highly similar to the original data, and the text is in a discernible state. We will evaluate its accuracy by scientific methods.

Fig.13. Comparison of the original page (left) and the post-recovery restored page (right)

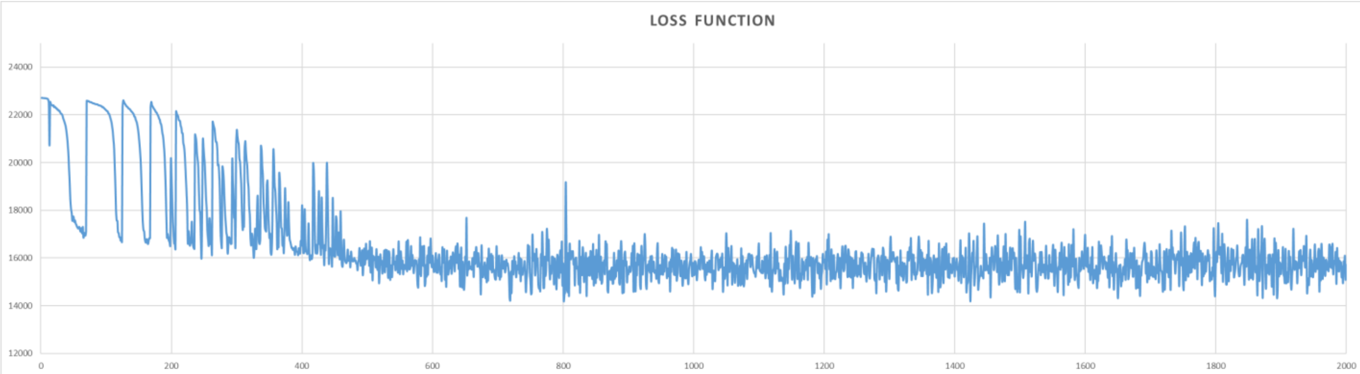

Fig.14. Loss function in the deep learning process of FCN framework.

4 CONCLUSION

Our method has successfully demonstrated that it is feasible to scan the booklet through the CT machine for data processing. It also shows that the 3DFCN-based deep learning framework can solve simple page extraction problems and curved page problems. Although there is no visual research on real ancient literature, we try proposing a general solution to provide a basis for future research by using deep learning experiments.

In the future work, we will try solving some problems existing at the present stage, such as more perfect visualization of each page of data and try reconstructing a new model, so that the relevant literature workers can conduct research. Create more new data sets to refine the existing models to handle more and more complex booklet data. Add text recognition to the visualized results page and use a more scientific approach to evaluate visualization results. Look for a deep learning framework other than 3DFCN to solve more related complex problems, and further research on ancient Chinese and ancient Japanese literature, and propose corresponding solutions.

REFERENCES

[1] E. A. Hoffman, G. McLennan, "Assessment of the pulmonary structure-function relationship and clinical outcomes measures: Quantitative volumetric CT of the lung", Academic Radiol., vol. 4, no. 11, pp. 758-776, 1997.

[2] L. W. Hedlund, R. F. Anderson, P. L. Goulding, J. W. Beck, E. L. Effmann, C. E. Putman, "Two methods for isolating the lung area of a CT scan for density information", *Radiology*, vol. 144, pp. 353-357, 1982.

[3] R. Uppaluri, T. Mitsa, M. Sonka, E. A. Hoffman, G. McLennan, "Quantification of pulmonary emphysema from lung CT images using texture analysis", *Amer. J. Resp. Crit. Care Med.*, vol. 156, no. 1, pp. 248-254, 1997.

[4] J.B. Ludlow, L.E. Davies-Ludlow, S.L. "BrooksDosimetry of two extraoral direct digital imaging devices: NewTom cone beam CT and Orthophos Plus DS panoramic unit", Dentomaxillofac Radiol, 32 (2003), pp. 229-234

[5] Ohnesorge B, Flohr T, Schwarz K, Heiken JP, Bae KT (2000) Efficient correction for CT image artifacts caused by objects extending outside of the scan field of view. Med Phys 27(1):39–46

[6] K. Sourbelle and W. Kalender, "Generalization of Feldkamp reconstruction for clinical spiral cone beam CT," in Proceeding of the International Conference on Fully 3D Reconstruction in Radiology and Nuclear Medicine, Saint Malo, France, 29 June–4 July 2003.

[7] X. Tang, J. Hsieh, R. Nilsen, S. Dutta, D. Samsonov, and H. Hagiwara, "A three-dimensional weighted cone beam filtered backprojection (CB-FBP) algorithm for image reconstruction in volumetric CT - helical scanning," Phys. Med. Biol. 10.1088/0031-9155/51/4/007 51, 855–874 (2006).

[8] M. Silver and K. Taguchi, "Windmill artifacts in multislice helical CT," in Proceedings of SPIE Conference on Medical Imaging, Vol. 5032, May 2003, pp. 1918–1927.

[9] H. Hu, "Multi-slice helical CT: Scan and reconstruction," Med. Phys. 10.1118/1.598473 26, 1–18 (1999).

[10] D. Heuscher, "Helical cone beam scans using oblique 2d surface reconstructions," in Proceedings of the International Conference on Fully 3D Reconstruction in Radiology and Nuclear Medicine, Egmond aan Zee, The Netherlands, 23–26 June 1999, pp. 204–207.

[11] Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 1090–1098 (2012).

[12] Tompson, J., Jain, A., LeCun, Y. & Bregler, C. Joint training of a convolutional network and a graphical model for human pose estimation. In Proc. Advances in Neural Information Processing Systems 27 1799–1807 (2014).

[13] Mikolov, T., Deoras, A., Povey, D., Burget, L. & Cernocky, J. Strategies for training large scale neural network language models. In Proc. Automatic Speech Recognition and Understanding 196–201 (2011).

[14] LeCun, Y. Une procédure d'apprentissage pour Réseau à seuil assymétrique in Cognitiva 85: a la Frontière de l'Intelligence Artificielle, des Sciences de la Connaissance et des Neurosciences [in French] 599–604 (1985).

[15] Dauphin, Y. et al. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In Proc. Advances in Neural Information Processing Systems 27 2933–2941 (2014).

[16] Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph. 2007; 31:198–211.

[17] Giger ML, Karssemeijer N, Schnabel JA. Breast image analysis for risk assessment, detection diagnosis, and treatment of cancer. Annu Rev Biomed Eng. 2013; 15:327–357.

[18] Heimann T, Meinzer HP. Statistical shape models for 3D medical image segmentation: a review. Med Image Anal. 2009; 13:543–563.

[19] Xu Y, Xu C, Kuang X, et al. 3D-SIFT-Flow for atlas-based CT liver image segmentation. Med Phys. 2016; 43:2229–2241.

[20] Shimizu A, Ohno R, Ikegami T, Kobatake H, Nawano S, Smutek D. Segmentation of multiple organs in non-contrast 3D abdominal CT images. Int J Comput Assist Radiol Surg. 2007; 2:135–142.

[21] A. Stijnman, in: M. Clarke, J.H. Townsend, A. Stijnman (Eds.), Art of the Past, Sources and Reconstructions. Proceedings of First Symposium of the Art Technological Source Research Group, Archetype, London, 2005, p. 125.

[22] E. Bulska, B. Wagner, B. Stahl, M. Heck, H.M. Ortner R. Van Grieken, K. Janssens, L. Van't dack, G. Meersman (Eds.), Seventh International Conference on Non-destructive Testing and Microanalysis for the Diagnostics and Conservation of the Cultural and Environmental Heritage Proceedings, University of Antwerp, Belgium (2002)

[23] P. Vandenabeele, L. Moens K. Janssens, R. Van Grieken (Eds.), Non-Destructive Microanalysis of Cultural Heritage Materials (Comprehensive Analytical Chemistry XLII), Elsevier, Amsterdam (2004), p. 635

[24] R. Bade, J. Haase, B. Preim, Comparison of fundamental mesh smoothing algorithms for medical surface models, in: Proceedings of the Simulation and Visualization, 2006, pp. 289–304.

[25] Y. Boykov, M.P. Jolly, Interactive graph cuts for optimal boundary and region segmentation, in: Proceedings of the ICCV, vol. 1, 2001, pp. 105–112.
[26] P. Campadelli, E. Casiraghi, G. Lombardi, Automatic liver segmentation from abdominal CT scans, in: Proceedings of the 14th International Conference on Image Analysis and Processing, ICIAP '07, 2007, pp. 731–736.
[27] Matsuyama, T. Miura, S.I., and Nagao, M., Structural Analysis of Natural Textures by Fourier Transformation, Comput. Vision, Graph. Image Process. 24, 347-362 (1983).
[28] Millán, M.S, and Escofet, J. Fourier-domain-based Angular Correlation for Quasiperiodic Pattern Recognition. Applications to Web Inspection, J. Appl. Opt., 35, 6253-6260 (1996).
[29] Shinohara, T, Takayama, J, Ohyama, S. et al., Analysis of Textile Fabric Structure with the CT Images, "Proceedings of SICE Annual Conference", Fukui, August 2003, pp. 234-238.
[30] Shafarenko, L, Petrou, M, and Kittler, J., Histogram-based segmentation in a perceptually uniform color space, IEEE Trans. on Image Processing, IP-7(9), pp. 1345-1358, Sep. 1998.