Recovering medical images from CT film photos

Quan Quan\textsuperscript{1}, Qiyuan Wang\textsuperscript{2}, Yuanqi Du\textsuperscript{3}, Liu Li\textsuperscript{4}, and S. Kevin Zhou\textsuperscript{1,2}

\textsuperscript{1} Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
\textsuperscript{2} Medical Imaging, Robotics, and Analytic Computing Laboratory and Engineering (MIRACLE) School of Biomedical Engineering & Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China
\textsuperscript{3} George Mason University
\textsuperscript{4} Imperial College London

Abstract. While medical images such as computed tomography (CT) are stored in DICOM format in hospital PACS, it is still quite routine in many countries to print a film as a transferable medium for the purposes of self-storage and secondary consultation. Also, with the ubiquitousness of mobile phone cameras, it is quite common to take pictures of CT films, which unfortunately suffer from geometric deformation and illumination variation. In this work, we study the problem of recovering a CT film, which marks the first attempt in the literature, to the best of our knowledge. We start with building a large-scale head CT film database CTFilm20K, consisting of approximately 20,000 pictures, using the widely used computer graphics software Blender. We also record all accompanying information related to the geometric deformation (such as 3D coordinate, depth, normal, and UV maps) and illumination variation (such as albedo map). Then we propose a deep framework called Film Image Recovery Network (FIReNet) to tackle geometric deformation and illumination variation using the multiple maps extracted from the CT films to collaboratively guide the recovery process. Finally, we convert the dewarped images to DICOM files with our cascade model for further analysis such as radiomics feature extraction. Extensive experiments demonstrate the superiority of our approach over the previous approaches. We plan to open source the simulated images and deep models for promoting the research on CT film image analysis.

Keywords: CT Film Image · Medical Image Restoration

1 Introduction

Recently, more and more attention has been drawn to medical image analysis. While medical images such as CT are stored in DICOM format in hospital PACS, it is still quite routine in many countries to print a film as a transferable medium for the purposes of self-storage and secondary consultation. For example, the market size of global medical radiography film, which is commonly used to print
a stack of CT slices, is estimated to reach 986.1 million USD by 2026, at a CAGR of 0.4% since 2016\(^5\).

With the ubiquitousness of mobile phone cameras, it becomes quite common to take pictures of the CT films to avoid physically transferring CT films or support remote access. While convenient, the pictures most time do not meet the standard for the doctors to process and analyze because sometimes the CT films are warped in the pictures. In addition to the geometric deformation, there is illumination variation caused by different the lighting condition and background scene. Therefore, it is a challenge to both dewarp the pictures and recover the content from rectified CT films in a framework. In this paper, we make the first attempt in the literature, to the best of our knowledge, to propose a CT film recovery solution. Its crucial part is to remove the background in the image and rectify it as in Fig. 1a.

Although there are no existing works about recovering warped CT films previously, a more related problem, the recovery and de-distortion of document images, has been studied. The previous methods usually rely on the geometric properties of papers [20,34], 3D shape [1,25,13,21,26,31,22,32] or visual cues [5,5,34,8,27,16,21,11,15,2,14,24,30,18,9,7]. However, these methods are resource- and time-intensive. Later, with the advancement of deep learning on computer vision, researchers start to tackle this problem utilizing the convolutional neural network [23]. Recently, DocUNet [17] firstly tackle this problem by constructing and training large 2D document datasets. DewarpNet [6] follows

---

5 https://www.marketwatch.com/press-release/medical-x-ray-film-market-size-share-global-trends-market-demand-industry-analysis-growth-opportunities-and-forecast-2026-2020-11-03
Recovering medical images from CT film photos

[17] to construct a 3D document dataset and uses the 3D coordinate map, which contains more information about the spatial structure of the documents, to generate a backward map, which is used to restore the warped image in the final step. However, in our CT film recovery task, as in Fig.1b, the backward map will cause many locally-warped areas, leading to poor visual performance. These locally-warped regions can negatively affect our further medical image processing and clinical analysis for CT films, hence we find another way to obtain the dewarped image—using a UV map. Though a UV map is generally applied to map textures to models, it can be also used to obtain textures from "CT film models ", and is verified to generate less locally-warped regions. Another problem is that the ambient light mixed with various colors reduce the quality of CT film images, and is harmful to the accurate diagnosis. Besides, RBG images have smaller value range and less details than DICOM files. Therefore, we need to remove the disturbance of lights and restore the value range and detail of the radiographs.

We propose a Film Image Recovery Network (FIReNet) that recovers CT film by the UV map. The FIReNet consists of two parts: 1) Dewarping module includes two modules: a) Multi-map module is designed to improve the quality of UV maps by outputting and combining multiple annotation maps. b) Transformation module is designed to generate UV maps, and also deformation maps for enhancing the UV map. In addition, to focus on disentangling local warps, both maps are trained with additional customized loss. 2) Quality Restoration Module includes two modules: a) De-illumination module is designed to solve the illumination problem by outputting albedo maps which retain only the content of the film without the light information. b) Medical Image restoration module is designed to convert the value range from RGB to CT, and predicts more details ignored in photos. We conduct our experiment based on the below CTFilm20K dataset.

2 CTFilm20K

As three are no public data for CT film recovery, we construct our CTFilm20K Dataset, which contains a large number of CT films with various real-world warping scenarios and different contents. After collecting high-quality background images and CT data, we utilize photorealistic rendering to generate images. Each sample contains several labeled maps including 3D coordinate, normal, UV, and albedo maps (Fig.2a). In total, the dataset contains about 20,000 richly annotated high-quality film images.

As in Fig.2b, we first extract slices from the public medical CT DICOMs [3] and convert them in a grid format to generate the film texture. In comparison with Doc3D [6], which captures the 3D shape of naturally deformed real objects, we rely on the embedded physics computing engines in Blender [4], a rendering software, which simulates deformation for different situations. After that, we render the images with synthetic film textures laid in various HDR scenes from [10]
and HDR Haven\(^6\) using path tracing. We randomly set camera positions in a certain range and vary illumination conditions in rendering.

3 Method

As in Fig. 3, our network consists of two modules: dewarping module and content recovery module. In general, FIReNet takes a warped CT film \( I \in \mathbb{R}^{h \times w \times 3} \) (\( h \) and \( w \) are image height and width, respectively) as input and predicts a UV map \( M_{uv} \in \mathbb{R}^{h \times w \times 2} \), where each point indicates its position on a texture map. To achieve the purpose of obtaining the texture map, we transform UV map \( M_{uv} \), which maps textures to objects, to the backward map \( B \in \mathbb{R}^{h \times w \times 2} \), which maps objects to textures by linear interpolation. In addition, to satisfy the need of radiologists, we convert the dewarped images to DICOM file, which is used to compute radiomics features.

3.1 Dewarping module

**Multi-map module.** We believe that the original image and its various annotations such as 3D coordinate map \( M_{3D} \in \mathbb{R}^{h \times w \times 3} \), normal map \( M_{nor} \in \mathbb{R}^{h \times w \times 3} \), and depth map \( M_{dp} \in \mathbb{R}^{h \times w \times 1} \) are closely related, hence unlike DewarpNet [6] using only 3D map, we use the information of three maps mentioned above by a shared encoder which extracts the features of the input image and 4 different decoders which generate \( M_{3D}, M_{nor}, M_{dp} \) and background map \( M_{bg} \in \mathbb{R}^{h \times w \times 1} \) respectively, with U-Net [23] as the backbone network. When these maps are

\(^6\) HDRHAVEN: https://hdrihaven.com/
Fig. 3: Overview of the proposed network: FIReNet.

generated, they are combined and filtered by $M_{bg}$ and sent to the subsequent modules for more processing.

**Transformation module.** Instead of generating a backward map $B$ directly from a model [6], we design the transformation module to generate the UV map $M_{uv}$, which is pixel-wise aligned with the original image and is stable and robust for restoration, but some regions especially around the edges are often missed because the number of valid points in $M_{uv}$ is less than that in the backward map $B$ and most of the valid points in $M_{uv}$ are concentrated in the middle part. Therefore, we also generate a deformation map $M_{df} \in \mathbb{R}^{h \times w \times 2}$, which is the relative movement position of each point from the input image to the unwarped image, and we utilize $M_{df}$ to enhance $M_{uv}$. Finally, we will use interpolation to make the final image. We also invoke the U-Net as the basic framework and take the integrated features from the previous module as inputs, and then use two different decoders to generate the corresponding UV map $M_{uv}$ and deformation map $M_{df}$, both of which have the same size as the original image. Afterward, we combine the UV map $M_{uv}$ and deformation map $M_{df}$ by converting the deformation map to UV map $M'_{uv}$, which is finally converted to backward map $B$ for CT film recovery. We follow [6] to recover the warped images with $B$.

### 3.2 Quality restoration module

For de-illumination, inspired by DMPHN [33,19] designed for denoising, we adopt the cascade network as our base model to meet our requirement of large scale, content consistency and clarity. In detail, the dewarped image will be sent to the de-illumination module to get the de-illuminated image $I_d$. We train this module with noisy images/ground-truth pairs following settings in [19]. In addition, we apply another cascade module to better restore the medical images $I_r$ with CT values.

Comparison between the predicted image and ground-truth in visual level is often insufficient. To this, besides the common evaluation metrics like SSIM and
PSNR, we perform a comparison based on radiomics features. We leverage the statistical tool, Chi-square Analysis, to assess the similarity of two distributions of multiple radiomics features including first-order features like energy, entropy, standard variation, etc. and shape features, and radiomics features like GLCM, GLDM, etc.

3.3 Loss function

In our training of dewarping module, we perform the end-to-end training with multiple objectives. In the multi-map module, we train the four maps with \( L_1 \) loss. For example, \( L_{3D} = ||M_{3D} - \hat{M}_{3D}||_1 \), where \( \hat{M}_{3D} \) represents the predicted 3D coordinate map, \( M_{3D} \) is the ground truth, and \( L_{3D} \) represents the loss value on \( M_{3D} \). Similarly, We define other losses of \( L_{bg}, L_{nor}, \) and \( L_{dp} \). Therefore, the total loss for shape module, \( L_{shape} \), is formulated as follows:

\[
L_{shape} = L_{3D} + L_{nor} + L_{dp} + L_{bg}
\]

In transformation module, we first perform the same operations as before on the maps of \( M_{uv} \), and \( M_{df} \) to get the corresponding losses, which are denoted as \( L_{uv}, \) and \( L_{df} \), respectively. In addition, we also add a shifting loss \( L_{shift} \) to reduce shifting, a disturbance loss \( L_{disturb} \) to reduce total variance, and a de-shifted difference loss \( L_{diff} \) to reduce variance of each point of \( M_{df} \). Denoting the predicted deformation map by \( \hat{M}_{df} \) and the ground truth by \( M_{df} \), we represent the entire process as follows:

\[
L_{shift} = ||\mu||_1, \quad L_{disturb} = ||\sigma||_1,
\]

\[
L_{diff}^c = \min(||\Delta M_{df}||_1, ||\Delta M_{df} - \mu||_1),
\]

\[
L_{diff} = \begin{cases} L_{diff}^c & \Delta M_{df} \odot (\Delta M_{df} - \mu) > 0, \\
0 & \Delta M_{df} \odot (\Delta M_{df} - \mu) \leq 0. \end{cases}
\]

where \( \Delta M_{df} = \hat{M}_{df} - M_{df}; \ \mu = \mu(\Delta M_{df}); \ \sigma = \sigma(\Delta M_{df}). \)

For deformation map, we hope our model to focus on relative deformation, hence in (2) we first calculate the difference \( \Delta M_{df} \) between the predicted deformation map \( \hat{M}_{df} \) and ground truth \( M_{df} \), and then calculate the mean \( \mu \) and the variance \( \sigma \) of \( \Delta M_{df} \). To focus on the relative image warping, we separate the shift loss \( L_{shift} \) from \( \Delta M_{df} \), and add standard variation as disturbing loss \( L_{disturb} \). In particular, we hope the prediction \( \hat{M}_{df} \) to approximate the relative location if far away from both relative and the absolute location, but we should not affect points already close to the absolute location, hence we calculate \( L_{diff} \) as mentioned above. Therefore, the total loss of the deformation map \( L_{df} = L_{shift} + L_{disturb} + L_{diff} \). Similarly, we can get \( L_{uv} \) for UV map according to \( L_{df} \) according to \( L_{3D} \). The loss for transformation module can be represented as follows: \( L_{trans} = L_{df} + L_{uv} \). Finally, the total loss for training the dewarping module is given as : \( L_{dewarp} = L_{shape} + L_{trans} \).
For quality restoration module, we follow [19] to perform a end-to-end training respectively. The objective is given as: \( \mathcal{L}_{\text{recover}} = \frac{1}{2} \sum_{i=1}^{N} ||S_i - G||^2 \), where \( N \) is the number of sub-models used in the cascade network, \( S_i \) is the output of sub-model \( i \), \( G \) is the ground-truth image.

### 4 Experiments

**Evaluation metrics.** To evaluate the similarity between the original and the dewarped CT film image, we choose PSNR, SSIM [28], and MS-SSIM [29] as our evaluation metrics. In these experiments, to avoid the effects of image shifting, we estimate the de-shifted results on these metrics. In medical image restoration part, we add chi-square tests on radiomics features to check if it is practical.

**Training Details.** For training, firstly, input images are resized to 256 \( \times \) 256 and we set Adam [12] as the optimizer, learning rate to 0.001 and batch size to 16. All maps and images are linearly scaled to the range \([-1, 1]\).

#### 4.1 The performance of FIReNet

We test DewarpNet [6] and our method on our CTFilm20k dataset. DewarpNet directly generates a backward map \( B \), while we generate a UV map, which is more robust for film restoration. Fig.1b which demonstrate a difficult example and two real photos and results in Table 1 show our advantage over other methods. The improvement on the performance of evaluation in Table 1 confirms the benefit of using the UV map and our network architecture.

| Methods              | PSNR ↑ | SSIM ↑ | MS-SSIM ↑ |
|----------------------|--------|--------|-----------|
| DewarpNet [6]        | 16.98  | 0.4501 | 0.6879    |
| FIReNet              | 25.30  | 0.8621 | 0.9523    |
| FIReNet(de-shifted)  | 25.60  | 0.8715 | 0.9564    |

Table 1: **Quantitative comparison between our FIReNet and DewarpNet** [6]. Methods are evaluated by PSNR, SSIM and MS-SSIM. For all of these metrics, higher value means better. “FIReNet” and “FIReNet(de-shifted)” are the evaluated with normal backward map, and “de-shifted” backward map (eq.2), respectively. Our method performs better than DewarpNet [6].

Besides, we test the medical image restoration module on our paired data including dewarped images from DewarpNet and the original DICOM files which is used for standard medical analysis. We directly input the dewarped images and predict the DICOM slice. The table in Fig. 4 shows the huge gap between photos and DICOM files. Fig 4 shows that the recovered DICOM files from our medical image restoration module passed the Chi-square test and is practical in the extent of reliable accuracy. We first suppose \( H_0 \): the distribution of prediction
is identical to the ground-truth distribution, and we sample 101 predictions so the DOF (degrees of freedom) is 100. The significance level $\alpha$ is set to 0.05, 0.01, 0.001, and $t_{100} = 1.660, 2.364, 3.390$. If $t > t^2_{\alpha}, n - 1$, the assumption is rejected and the sample mean is different with population mean, vice versa. As shown in Fig. 4, t-scores of almost features are smaller than the threshold. 67 in 101 are smaller than $t_{100}^{0.05}$, 83 are smaller than $t_{100}^{0.01}$, 98 are smaller than $t_{100}^{0.001}$.

### Metrics on DICOM

| Output | PSNR | SSIM |
|--------|------|------|
| $I_d$  | 5.23 | 0.593|
| $I_1$  | 27.1 | 0.942|

Fig. 4: Metrics (Table) and t-score (Barplot) between our predictions and medical images (CT) as ground-truth.

#### 4.2 Ablation Study

To further analyze our method, we conduct an ablation study to characterize the roles of each module. Results are shown in Table 2. **UV map:** UV map is the key part of our method to restore the image. We compare our model which generates UV maps for restoration, with the baseline model which generates backward maps directly for restoration. According to Table 2, we can find that UV map outperforms the backward map (25.16dB vs 23.76dB in PSNR). **Multi-map:** We study the comparison between using only a 3D coordinate map and using multiple maps that include 3D coordinate, normal, and depth maps. From Table 2, we can find an improvement on these metrics (from 25.16dB to 25.51dB in PSNR). **Deformation module:** We compare the performance of networks with and without the deformation module. The performance is improved from 25.51dB to 25.60dB in PSNR.

| UV MM Deform | PSNR $\uparrow$ | SSIM $\uparrow$ | MS-SSIM $\uparrow$ |
|--------------|----------------|----------------|-------------------|
| ✓            | 23.76          | 0.8450         | 0.9446            |
| ✓            | 25.16          | 0.8594         | 0.9515            |
| ✓ ✓          | 25.51          | 0.8691         | 0.9553            |
| ✓ ✓ ✓        | 25.60          | 0.8715         | 0.9564            |

Table 2: **Ablation study.** UV: using UV maps. MM: generating multiple maps. Deform: adding deformation module.
5 Conclusions and Future Work

In this work, we proposed our FIReNet, a novel dewarping network for CT film recovery. Our approach is robust to film content, reflection, and background and is insensitive to slight image shifting. Through extensive experiments, we validated the advantages of our approach, which outperforms the previous approaches. Moreover, we contribute the CTFilm20K dataset, which is the first large dataset for CT film dewarping, and full of various types of maps and annotations. However, there still exists some limitations like limited types of CT film. We will collect more data and validate our method to real scenarios.

References

1. Brown, M.S., Seales, W.B.: Document restoration using 3d shape: a general deskewing algorithm for arbitrarily warped documents. In: ICCV. IEEE (2001)
2. Cao, H., Ding, X., Liu, C.: A cylindrical surface model to rectify the bound document image. In: ICCV. IEEE (2003)
3. Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N.G., Venugopal, V.K., Mahajan, V., Rao, P., Warier, P.: Development and validation of deep learning algorithms for detection of critical findings in head ct scans. arXiv preprint arXiv:1803.05854 (2018)
4. Community, B.O.: Blender - a 3D modelling and rendering package. Blender Foundation, Stichting Blender Foundation, Amsterdam (2018), http://www.blender.org
5. Courteille, F., Crouzil, A., Durou, J.D., Gurdjos, P.: Shape from shading for the digitization of curved documents. Machine Vision and Applications 18(5), 301–316 (2007)
6. Das, S., Ma, K., Shu, Z., Samaras, D., Shilkrot, R.: Dewarpnet: Single-image document unwarping with stacked 3d and 2d regression networks. In: CVPR. IEEE (2019)
7. Das, S., Mishra, G., Sudharshana, A., Shilkrot, R.: The common fold: utilizing the four-fold to dewarp printed documents from a single image. In: Proceedings of the 2017 ACM Symposium on Document Engineering. pp. 125–128 (2017)
8. Ezaki, H., Uchida, S., Asano, A., Sakoe, H.: Dewarping of document image by global optimization. In: Eighth International Conference on Document Analysis and Recognition (ICDAR’05). pp. 302–306. IEEE (2005)
9. Forsyth, D.A.: Shape from texture and integrability. In: ICCV. IEEE (2001)
10. Gardner, M., Sunkavalli, K., Yumer, E., Shen, X., Gambaretto, E., Gagné, C., Lalonde, J.: Learning to predict indoor illumination from a single image. ACM Trans. Graph. 36(6), 176:1–176:14 (2017). https://doi.org/10.1145/3130800.3130891, https://doi.org/10.1145/3130800.3130891
11. Kim, B.S., Koo, H.I., Cho, N.I.: Document dewarping via text-line based optimization. Pattern Recognition 48(11), 3600–3614 (2015)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015), http://arxiv.org/abs/1412.6980
13. Koo, H.I., Kim, J., Cho, N.I.: Composition of a dewarped and enhanced document image from two view images. IEEE Transactions on Image Processing 18(7), 1551–1562 (2009)
14. Liang, J., DeMenthon, D., Doermann, D.: Geometric rectification of camera-captured document images. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(4), 591–605 (2008)
15. Liu, C., Zhang, Y., Wang, B., Ding, X.: Restoring camera-captured distorted document images. International Journal on Document Analysis and Recognition (IJDAR) 18(2), 111–124 (2015)
16. Lu, S., Tan, C.L.: Document flattening through grid modeling and regularization. In: 18th International Conference on Pattern Recognition (ICPR’06). IEEE (2006)
17. Ma, K., Shu, Z., Bai, X., Wang, J., Samaras, D.: Docunet: document image unwarping via a stacked u-net. In: CVPR. IEEE (2018)
18. Malik, J., Rosenholtz, R.: Computing local surface orientation and shape from texture for curved surfaces. International journal of computer vision 23(2), 149–168 (1997)
19. Mehri, A., Ardakani, P.B., Sappa, A.D.: Mrpnet: Multi-path residual network for lightweight image super resolution. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2704–2713 (2021)
20. Meng, G., Su, Y., Wu, Y., Xiang, S., Pan, C.: Exploiting vector fields for geometric rectification of distorted document images. In: ECCV (2018)
21. Meng, G., Wang, Y., Qu, S., Xiang, S., Pan, C.: Active flattening of curved document images via two structured beams. In: CVPR (2014)
22. Östlund, J., Varol, A., Ngo, D.T., Fua, P.: Laplacian meshes for monocular 3d shape recovery. In: ECCV. Springer (2012)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. Springer (2015)
24. Tian, Y., Narasimhan, S.G.: Rectification and 3d reconstruction of curved document images. In: CVPR. IEEE (2011)
25. Tsoi, Y.C., Brown, M.S.: Multi-view document rectification using boundary. In: CVPR. IEEE (2007)
26. Ulges, A., Lampert, C.H., Breuel, T.: Document capture using stereo vision. In: Proceedings of the 2004 ACM symposium on Document engineering. pp. 198–200 (2004)
27. Ulges, A., Lampert, C.H., Breuel, T.M.: Document image dewarping using robust estimation of curled text lines. In: Eighth International Conference on Document Analysis and Recognition (ICDAR’05). pp. 1001–1005. IEEE (2005)
28. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4), 600–612 (2004)
29. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. vol. 2, pp. 1398–1402. Ieee (2003)
30. Witkin, A.P.: Recovering surface shape and orientation from texture. Artificial intelligence 17(1-3), 17–45 (1981)
31. Yamashita, A., Kawarago, A., Kaneko, T., Miura, K.T.: Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. Icpr 2004. vol. 1, pp. 482–485. IEEE (2004)
32. You, S., Matsushita, Y., Sinha, S., Bou, Y., Ikeuchi, K.: Multiview rectification of folded documents. IEEE transactions on pattern analysis and machine intelligence 40(2), 505–511 (2017)
33. Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5978–5986 (2019)

34. Zhang, L., Yip, A.M., Brown, M.S., Tan, C.L.: A unified framework for document restoration using inpainting and shape-from-shading. vol. 42, pp. 2961–2978. Elsevier (2009)