UMEDNERF: UNCERTAINTY-AWARE SINGLE VIEW VOLUMETRIC RENDERING FOR MEDICAL NEURAL RADIANCE FIELDS

Jing Hu[1], *Qinrui Fan*[1], *Shu Hu*[2], *Siwei Lyu*[3], *Xi Wu*[1], *Xin Wang*[4,†]

[1] Chengdu University of Information Technology, Chengdu, China
[2] Indiana University–Purdue University Indianapolis, IN, USA
[3] University at Buffalo, State University of New York, NY, USA
[4] University at Albany, State University of New York, NY, USA

ABSTRACT

In the field of clinical medicine, computed tomography (CT) is an effective medical imaging modality for the diagnosis of various pathologies. Compared with X-ray images, CT images can provide more information, including multi-planar slices and three-dimensional structures for clinical diagnosis. However, CT imaging requires patients to be exposed to large doses of ionizing radiation for a long time, which may cause irreversible physical harm. In this paper, we propose an Uncertainty-aware MedNeRF (UMedNeRF) network based on generated radiation fields. The network can learn a continuous representation of CT projections from 2D X-ray images by obtaining the internal structure and depth information and using multi-task adaptive loss weights to ensure the quality of the generated images. Our model is trained on publicly available knee and chest datasets, and we show the results of CT projection rendering with a single X-ray and compare our method with other methods based on generated radiation fields.

Index Terms— NeRF, Uncertainty, Medical Image, GAN, Deep Learning

1. INTRODUCTION

Digitally Reconstructed Radiographs (DRR) is a medical image processing technique employed for simulating and generating X-ray images. This method entails the creation of virtual X-ray projections using three-dimensional medical image data, such as CT or MRI, in conjunction with computer graphics algorithms. By generating DRRs, we can avoid the need for acquiring paired X-rays and corresponding CT reconstructions, which would otherwise subject patients to increased radiation exposure [1].

Deep learning models such as 3D convolutional neural networks, generative adversarial networks (GAN), and variational autoencoders can be applied to the reconstruction of

† Corresponding author (xwang56@albany.edu).

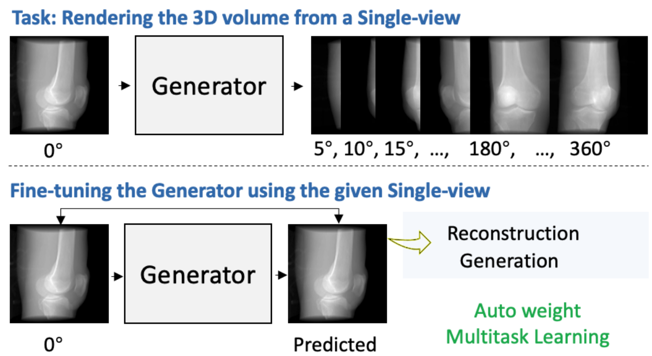

Fig. 1. Overview of the task of UMedNeRF. **Top:** The goal is to render the 3D volume from a single view. **Bottom:** Given the trained Generator, it is further fine-tuned for the final rendering process via a multitask learning method.

medical images like CT and MRI [2]. However, these methods often demand high-performance computing resources due to the computational complexity in three-dimensional space. Furthermore, they typically require extensive training data, and achieving satisfactory results can be challenging. Neural radiance fields (NeRF) [3] method has yielded surprising results in traditional 3D image reconstruction, but it cannot be directly applied to medical image reconstruction due to the unique internal details of medical images.

MedNeRF [4], a GAN-based Raditonal Filed, successfully untangles the surface shape, volumetric depth, and internal anatomical structures from 2D images by learning a function that assigns a radiance value to each pixel. However, due to the nature of GAN network architecture for image generation and using a single X-ray for CT projection reconstruction during image generation, the network may overfit to the single X-ray, leading to an imbalance between image clarity and accuracy. Thus, during the single-view volumetric rendering stage, MedNeRF fine-tuned the generator using the input view by jointly optimizing the reconstruction and the generator. The reconstruction loss is based on the disparity between the predicted and actual CT projections, while the

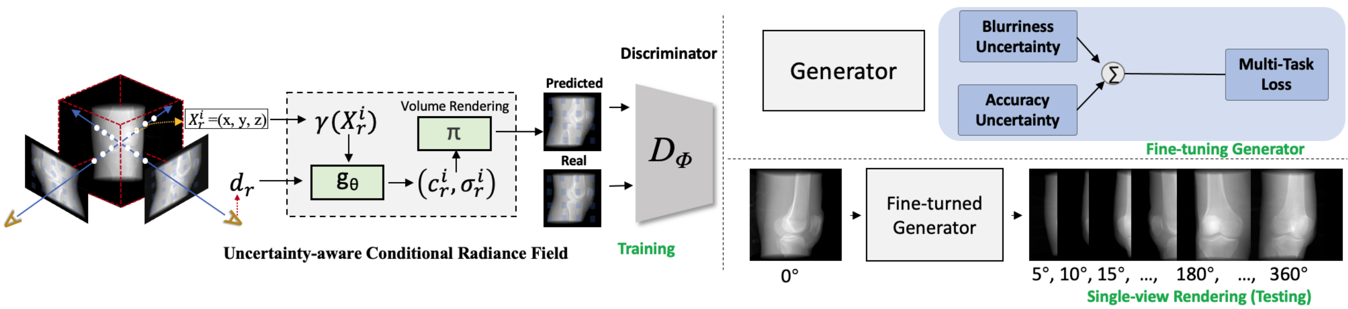

Fig. 2. Overview of our Uncertainty-aware MedNeRF. **Left:** Network structures. **Right:** (Top) Fine-tuning the generator for balancing the Blurrinenss and Accuracy. (Bottom) Rendering the whole 3D volume (each view taken at a 5-degree interval) using a single view slide.

generative loss is calculated using the variance between the predicted and actual radiance fields (See Fig. 1). The hyperparameters for balancing two losses are typically configured manually, which can pose challenges in achieving the optimal trade-off between image blurriness and accuracy.

To address this challenge, in this paper, we propose Uncertainty-aware MedNeRF (**UMedNeRF**) to learn weights for clarity and precision balance problems automatically. We propose to use an uncertainty-aware conditional radiance field that simultaneously weighs multiple loss functions automatically in the multi-task settings and generates the predicted patch for single-view volumetric rendering. The experiments illustrate that our method significantly improves the model's reconstruction abilities while also accelerating the rate of convergence during testing.

2. METHODOLOGY

Overview. The overview of the framework of our UMedNeRF is shown in Fig. 2 (Left). We use the same network structures in the MedNeRF [4]. The whole structure consists of a radiance field network and a multi-scale patch-based discriminator. The 2D image is encoded as 3D position $X = (x, y, z)$, and the viewing angle direction information d is fed into g as input. Network output volume density σ and pixel value c. At the same time, in order to learn the high-frequency characteristics of the input information, the input position information is mapped to a two-dimensional representation $\gamma(X)$, as Eq. (1). Finally, the output (c, σ) is rendered as a predicted patch by Volume Rendering. The multi-scale patch-based discriminator takes in a pair of images (either real or generated) and outputs a scalar value indicating whether the pair is real or generated.

$$\gamma(X) = \left(\ldots, \sin\left(2^i \pi X\right), \cos\left(2^i \pi X\right), \ldots\right) \tag{1}$$

2.1. Uncertainty-aware Generative Radiance Fields

Compared with natural images, medical images such as CT scans and MRI bring unique challenges to 3D reconstruction due to the problems of unclear boundaries and complex internal structure. It is difficult to establish a continuous representation of a CT scan directly using a neural radiation field, and it is impossible to effectively distinguish surface shape, volume depth, and internal structure from 2D images. And considering the well-known trade-off problem between distortion and perception in GAN methods, our method solves this problem by adding $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{\mathrm{MSE}}$.

$$\mathcal{L}_{\mathrm{r}} = \mathbb{E}_{\boldsymbol{f}\sim D(\boldsymbol{p}), \boldsymbol{p}\sim \boldsymbol{P}}\left[\frac{1}{whd}\left\|\phi_i(\mathcal{G}(\boldsymbol{f})) - \phi_i(\mathcal{T}(\boldsymbol{p}))\right\|_2\right] \tag{2}$$

Where $\phi_i(\cdot)$ represents the output of the ith layer of the pre-trained VGG16 network model, w, h and d is the width, height, and depth of the feature space, $\mathcal{G}$ represents the processing of $\mathcal{D}_\phi$ intermediate feature f, and $\mathcal{T}$ is the processing of the truth image block.

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

Where $\hat{y}$ is the predicted value, y is the true value, and N is the number of predicted and true values.

$$\mathcal{L}_{\mathrm{gen}} = \lambda_1 \mathcal{L}_r(VGG16) + \lambda_2 \mathcal{L}_{MSE}(G) \tag{4}$$

In most cases, hyperparameters λ_1 and λ_2 are traditionally configured manually, are subject to observation, and heavily depend on human expertise, which can pose challenges in achieving the optimal trade-off between image blurriness and accuracy.

Uncertainty-aware Multitask Learning. In this section, we proposed a novel approach to automatically learn λ_1 and λ_2 for these losses based on task-dependent uncertainty [5]. The learning objective can be represented as

$$\mathcal{L} = \frac{1}{2\sigma_1^2}\mathcal{L}_r(VGG16) + \frac{1}{2\sigma_2^2}\mathcal{L}_{MSE}(G) + \log \sigma_1\sigma_2, \tag{5}$$

where σ_1 and σ_2 are the VGG16 and G modules' observation noise parameters, which can capture how much noise we have in the outputs of each module. Minimizing $\mathcal{L}$ with respect to σ_1 and σ_2 can learn the relative weight of the losses adaptively. The last term in Eq. (5) is a regularizer, which can constrain the noise to be increased too much.

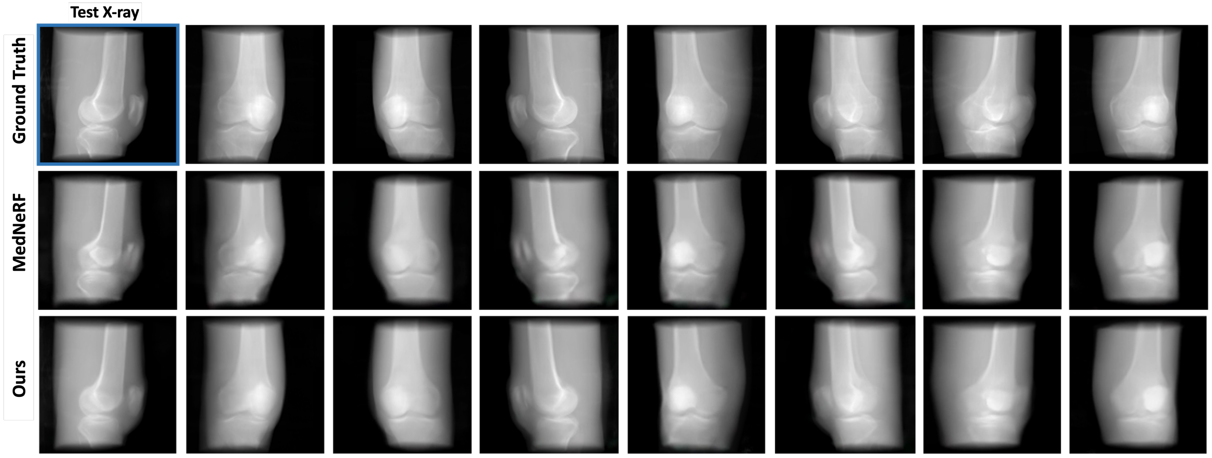

Fig. 3. The single knee X-ray given yields a complete CT projection by our uncertainty-aware fine-tuned generator.

Method	Chest dataset		Knee dataset	
	↑ PSNR (dB)	↑ SSIM	↑ PSNR (dB)	↑ SSIM
MedNeRF [4]	28.37	0.462	30.93	0.670
Ours	**28.41**	**0.484**	**31.62**	**0.723**

Table 1. Quantitative results based on PSNR and SSIM of rendered X-ray projections with single-view X-ray input.

Method	Chest dataset		Knee dataset	
	↓ FID ($\mu \pm \sigma$)	↓ KID ($\mu \pm \sigma$)	↓ FID ($\mu \pm \sigma$)	↓ KID ($\mu \pm \sigma$)
GRAF [10]	68.25 $\pm$ 0.954	0.053 $\pm$ 0.0008	76.70 $\pm$ 0.302	0.058 $\pm$ 0.0001
pixelNeRF [11]	112.96 $\pm$ 2.356	0.084 $\pm$ 0.0014	166.40 $\pm$ 2.153	0.158 $\pm$ 0.0010
GIRAFFE [12]	66.42 $\pm$ 0.706	0.064 $\pm$ 0.0012	184.34 + 0.281	0.064 $\pm$ 0.0012
Ours	**62.40** $\pm$ **0.352**	**0.049** $\pm$ **0.0008**	**76.41** $\pm$ **2.153**	**0.057** $\pm$ **0.0006**

Table 2. FID and KID analysis comparing other methods.

Task Weights		PSNR	SSIM
$\mathcal{L}_r$	$\mathcal{L}_{MSE}$		
0.6	0.4	30.93	0.67
0.5	0.5	30.67	0.61
0.4	0.6	31.26	0.70
Learned weights with task uncertainty		**31.62**	**0.72**

Table 3. Comparison of Loss Weights on Knee Dataset.

Fine-tuning Generator. In this phase, we reconstruct a complete CT projection across a full vertical rotation using a single X-ray image. In addition to adjusting the shape and appearance of latent vectors z_s and z_a, we also fine-tuned the generator parameters using the test image. Furthermore, we also normalize the output to enhance the quality of the CT projections using $y' = \frac{y - \min(y)}{\max(y) - \min(y)} \times 255$.

Testing. In the single-view rendering stage, we challenged our model by employing a single X-ray (referred to as a single view) for CT projection reconstruction over a full 360-degree range, each view taken at a 5-degree interval (See Fig. 2 (Right)). This test phase is instrumental in gauging the model's performance across the entire angular spectrum.

3. EXPERIMENTS

Dataset. To ensure comparability with existing methodologies, we leveraged the identical CT dataset employed in the MedNeRF paper to produce DRRs. The dataset for generating DRRs comprised 20 CT chest scans as referenced in [6] and [7], in addition to 5 CT knee scans documented in [8] and [9]. These scans represented a diverse array of patients, encompassing various contrast modalities and encapsulating both typical and atypical anatomical structures. The DRR generation process involved the emulation of the rotation of a radiation source and imaging panel around the vertical axis, resulting in the creation of DRRs with a resolution of 128 $\times$ 128 at five-degree intervals. Consequently, each object produced a total of 72 distinct DRRs.

Evaluation Metrics. We evaluate our model's performance in CT projection reconstruction using two key image quality metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR emphasizes luminance accuracy, measuring differences between original and model-generated images. SSIM provides a comprehensive assessment, considering luminance, contrast, and structural similarity. These metrics are crucial for accurately assessing our model's performance in CT projection reconstruction, ensuring reliability in our research findings.

Comparison with baseline models. We first compared it with the 3D baseline MedNeRF [4]. Table 1 displays our results, which were assessed using PSNR and SSIM. As shown in Fig. 3, our model can better find the balance between blurriness and accuracy in the quality of generated image results. Our findings highlight the effectiveness of Automatic Weighted Loss in improving the clarity of reconstructed images and revealing finer details within bone structures. Moreover, our model demonstrates the ability to consistently generate CT projections from multiple consecutive viewpoints.

We also evaluate our model on the task of 2D rendering and compare it to [10], [11], and [12] baseline, Table 2 presents a comparison of image quality using the Frechet Inception Distance (FID) and Kernel Inception Distance (KID) metrics, where lower values indicate better quality. When optimizing pixelNeRF and GIRAFFE on our datasets, it yields notably inferior results that do not match the performance of the GRAF baseline or our model. In contrast, our model surpasses the baseline models in terms of FID and KID metrics across all datasets.

Ablation Study. Due to the balance between image clarity and accuracy being a crucial concern in GAN network structures, we introduce Mean Squared Error (MSE) loss during the generation phase in an attempt to address this issue [13]. The model's performance is highly sensitive to the choice of weight parameters. A common approach for combining multiple objective losses is to simply assign weights to each objective and sum them up. However, this method heavily relies on manual expertise and can be quite costly in terms of experimentation, often taking many days to fine-tune. In our proposed UMedNeRF, we utilize Autoweightloss, which is capable of automatically learning the optimal weights for the task. As shown in Table 3, we employ various sets of loss weight parameters for ablation studies to assess the effectiveness of our approach. Experimental results demonstrate that UMedNeRF efficiently addresses the balance issue in generating images.

4. CONCLUSION

Our research proposes a novel deep learning architecture based on neural radiance fields for learning continuous representations of CT scans from 2D input X-rays. The proposed discriminator can provide comprehensive information to help generate the radiation field for 3D-aware CT projection reconstruction. In the test phase, our proposed UMedNeRF uses a single X-ray to achieve a complete CT projection reconstruction and solves the distortion and perception tradeoff common in GAN methods to generate images through Autoweightloss. The experimental results show that UMedNeRF has a significant improvement in the quality of CT projections compared with other neural radiance field methods. Although this model cannot completely replace CT at the current stage, this method can greatly reduce the radiation dose received by patients in clinical examinations and reduce the time and economic costs, which has great research potential in bone trauma and surgery.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by (Source information). Ethical approval was not required as confirmed by the license attached with the open access data.

6. ACKNOWLEDGMENTS

This work was supported by Sichuan province Key Technology Research and Development project under Grant 2023YFG0305. Xin Wang is supported by University at Albany Start-up Grant.

7. REFERENCES

[1] Pechin Lo, Bram Van Ginneken, Joseph M Reinhardt, et al., "Extraction of airways from ct (exact'09)," *IEEE Transactions on Medical Imaging*, vol. 31, no. 11, 2012.

[2] Payal Maken and Abhishek Gupta, "2d-to-3d: A review for computational 3d image reconstruction from x-ray images," *Archives of Computational Methods in Engineering*, 2023.

[3] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," 2021, vol. 65, ACM New York, NY, USA.

[4] Abril Corona-Figueroa, Jonathan Frawley, et al., "Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray," in *EMBC*. IEEE, 2022.

[5] Alex Kendall, Yarin Gal, and Roberto Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.

[6] E. B. Tsai, S. Lungren, M. P. Hershman, M. Roshkovan, L. Colak, et al., "Medical imaging data resource center (midrc) - rsna international covid open research database (ricord) release 1b - chest ct covid," 2021.

[7] K. Clark, B. Vendt, K. Smith, J. Freymann, et al., "The cancer imaging archive (tcia): Maintaining and operating a public information repository," 2013, vol. 26.

[8] Michael D. Harris, Adam J. Cyr, Azhar A. Ali, et al., "A combined experimental and computational approach to subject-specific analysis of knee joint laxity," 2016, vol. 138.

[9] Azhar A. Ali, Sami S. Shalhoub, Adam J. Cyr, et al., "Validation of predicted patellofemoral mechanics in a finite element model of the healthy and cruciate-deficient knee," 2016.

[10] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger, "Graf: Generative radiance fields for 3d-aware image synthesis," 2020, vol. 33.

[11] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa, "pixelnerf: Neural radiance fields from one or few images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[12] Michael Niemeyer and Andreas Geiger, "Giraffe: Representing scenes as compositional generative neural feature fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[13] Shu Hu, Xin Wang, and Siwei Lyu, "Rank-based decomposable losses in machine learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.