Ground Truth base Deep Learning Model Initialization for Medical Image Segmentation

Zijin Yang[*1] MICHAEL.YANG@FAU.DE
Nazmus Sakib[†1] NAZMUS.SAKIB@FAU.DE
Chang Liu[1] CHANG.CH.LIU@FAU.DE
Andreas Maier[1] ANDREAS.MAIER@FAU.DE
[1] *Friedrich-Alexander-Universität Erlangen Nürnberg, Germany*

Editors: Under Review for MIDL 2026

Abstract

Deep learning-based approaches for medical image segmentation have demonstrated considerable promise. However, their clinical translation remains challenging due to limited annotated datasets and the suboptimal performance of standard weight initialization strategies when applied to volumetric medical imaging data. To address these limitations, we propose a novel data-driven initialization framework, termed Principal Component Analysis-based Reference Patient initialization (GT-PCA), which leverages prior anatomical knowledge extracted from a reference patient to customize convolutional kernel weights. Our method employs Principal Component Analysis (PCA) to decompose structural information from the reference patient, which is then systematically propagated to initialize kernels across network depths, thereby encoding domain-specific anatomical priors into the network architecture.

We systematically evaluated the proposed initialization strategy against conventional Xavier and Kaiming initialization methods using U-Net and Residual U-Net architectures on two distinct medical imaging datasets. Experiments were conducted with varying kernel configurations ($3 \times 3 \times 3$ and $7 \times 7 \times 7$) across multiple training iterations. Our results demonstrate that the GT-PCA initialization consistently outperforms baseline methods across both kernel configurations. For the $3 \times 3 \times 3$ kernel architecture trained over 1000 epochs, GT-PCA achieved a Dice score of 0.79, representing improvements of 3.9% and 31% over Kaiming (0.76) and Xavier (0.48) initialization, respectively. Similarly, for the $7 \times 7 \times 7$ kernel configuration trained over 500 epochs, GT-PCA achieved a Dice score of 0.79, compared to 0.76 for both Xavier and Kaiming methods. These findings validate that incorporating anatomical priors through PCA-based initialization significantly enhances segmentation performance, particularly in scenarios with limited training data. This approach provides a principled framework for integrating domain knowledge into deep learning models for medical image analysis.

Keywords: Neural Network, Model Initialization, Data-Driven Method, Medical Image Segmentation

[*] Contributed equally
[†] Contributed equally

1. Introduction

The development of U-Net and its derivatives has led to significant advancements in deep learning-based medical image segmentation, enabling automated diagnosis and detection of pathological conditions. Yet, when using an architecture that has not been pre-trained with a large-scale dataset or trained from scratch, the outcome is not satisfactory. Especially, training a neural network with dataset e.g. a medical dataset can be extremely difficult due to the limited size of training data.

The purpose of training a neural network is to adjust parameters i.e. kernels in the layers to obtain a satisfactory result. A critical aspect of optimizing deep learning neural networks is the proper initialization of convolution kernels (Xu and Wang, 2022). Common approaches for kernel initialization include random weight assignment (Krizhevsky et al., 2017), Xavier initialization (Kumar, 2017), and Kaiming initialization (He et al., 2015).

These kernel initialization techniques typically rely on independent weight initialization without considering the available training data. The weights are adjusted during training to match the local image patterns, a process that requires numerous iterations. In this case, valuable information on the training data itself e.g. patterns, edges, spatial information, has not been used. This can result in a slower network, longer training times, and extended convergence rates. To address this issue, alternative methods have been explored. OrthoNorm, for instance, employs orthogonal matrix initialization, suitable for non-linear networks, unlike random assignment (Saxe et al., 2013). Another method, "Layer sequence unit variance (LSUV)," incorporates orthogonal initialization into the iterative process by utilizing singular value decomposition (SVD) to replace weights initially set with Gaussian noise (Mishkin and Matas, 2015). Additionally, Chan et al. introduced a Principal Component Analysis (PCA) based approach for convolution kernel initialization (Chan et al., 2015). This method calculates the principal components of image patches from feature maps to initialize convolution kernels.

Instead of relying on random weight initialization, one could initialize the network with kernels that are already proximate to the optimum for a given task or domain. If feasible, this approach could transform network optimization into a form of targeted fine-tuning, substantially reducing dependency on large-scale training datasets and extensive computation from scratch. In this work, we seek to enhance PCA-based kernel initialization by integrating ground truth (GT) images and introducing a unique initialization approach, referred to as Ground Truth (GT) - PCA-based initialization. Leveraging the labeled GT images, dominant features can be identified and utilized as convolution kernel weights, establishing a relationship between training images and convolution kernels. This has the potential to reduce training time and enhance convergence rates (Xu and Wang, 2022). The proposed initialization method has undergone extensive benchmarking and is subject to various quantitative measures, all of which are within the scope of this work.

2. Methods

2.1. Data

This work compares different initialization methods for segmentation tasks using medical data. We used two public medical segmentation datasets for experiments: Abdominal Multi-

Organ Segmentation (AMOS) dataset (Ji et al., 2022) as well as computed tomography (CT) volumes with Multiple Organ Segmentation (CT-ORG) dataset (Rister et al., 2020).

2.2. Network architectures

This study employs U-Net-based deep learning architectures for image segmentation. The 2D U-Net (Ronneberger et al., 2015) uses a symmetric encoder–decoder structure with skip connections to retain fine spatial details during upsampling. Our implementation comprises five depth levels, each employing convolutional layers, batch normalization, ReLU activation, and max-pooling.

For volumetric data, we extend the approach using a 3D U-Net. To improve gradient flow and stability in deeper networks, we also implement a Residual U-Net variant, which integrates residual blocks within each encoder stage and uses transposed convolutions for decoder upsampling. These architectures provide a flexible and effective framework for segmentation tasks across two and three dimensions.

2.3. Model Initialization with GT-PCA

Model initialization with the proposed GT-PCA consists of two major steps, i.e. GT-based initialization for the initial layer and PCA-based initialization for the subsequent layers. In particular, the features from the selected reference patient together with the ground truth mask are extracted as the convolutional kernels of the initial convolutional layer. The initial layer propagates the masked reference CT image and the subsequent layer output is used to initialize the next layers, until all layers are initialized The GT initialization and PCA initialization are here elaborated.

2.3.1. GT Initialization

The GT initialization process starts with selecting a random patient from the dataset, with data structured as 3D voxels having dimensions ($W \times H \times T$). With the define kernel size as $3 \times 3 \times 3$, the network's initial state is represented as a 5-dimensional tensor ($1 \times C \times 3 \times 3 \times 3$), where C denotes the initial channel dimension. This initialization step includes the following three steps:

Data Traversal: The reference CT volume is first masked by the foreground of the mask for feature extraction. It is then traversed horizontally and vertically using strides proportional to the kernel's dimensions. This process generates "potential kernels" and the actual kernels chosen align with the network's kernel size, for example, $3 \times 3 \times 3$.

Filtering Potential Kernels: Filtering potential weights involves adjusting values in the range of 0 to 255 and applying the Canny Edge detection algorithm (Canny, 1986) with a hysteresis threshold of 100 to 200. This step emphasizes weights with discernible curves, crucial for capturing features like abdominal organs.

Generating Kernel Weights: Among the potential kernels, C kernels are randomly selected, where C equals the channel dimension. These chosen weights are saved and concatenated as the initialization for this layer.

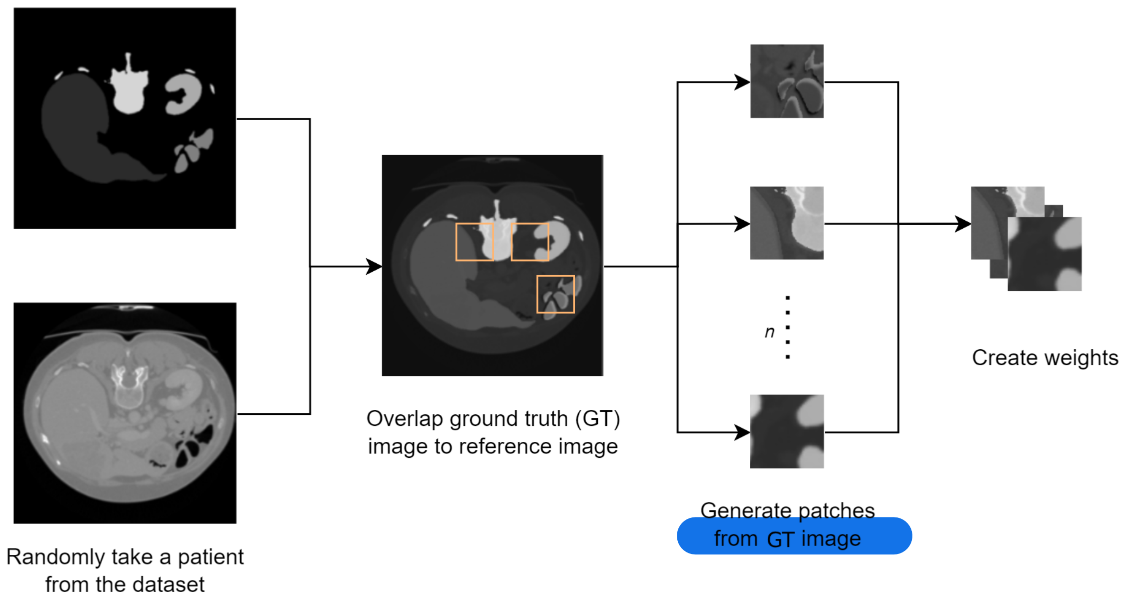

Figure 1: Illustration of the high-level architecture of GT-based kernel generation. Kernels (weights) are created by normalizing the selected patches from the extracted targets using the ground truth image and the original image

2.3.2. PCA Initialization

The subsequent layers of the encoder in U-Net or residual U-Net architectures are initialized using a Principle Component Analysis (PCA)-based method. This method is applied from the second layer onward, recursively traversing the network depth. The process involves three key steps, as illustrated in Figure 3.

Gathering Output: Beginning from the second layer, the algorithm collects outputs from the preceding layer and separates them into individual feature maps (see Figure 4). From each feature map, small patches, e.g., $3 \times 3 \times 3$ tensors are cropped directly by iterating over all directions through the feature map.

Generate PCA Weights: All cropped kernels from each feature map are then concatenated, making a 4-dimensional tensor. Principal components are then derived through these 4-dimensional tensors, n largest components required for the subsequent layer are then selected and dimensionality reduction is performed based on these largest components (see Figure 4).

Generate Kernel Weights: The final step involves combining and resizing the calculated tensors to generate PyTorch tensors to match the shape between the curent layer and the next layer.

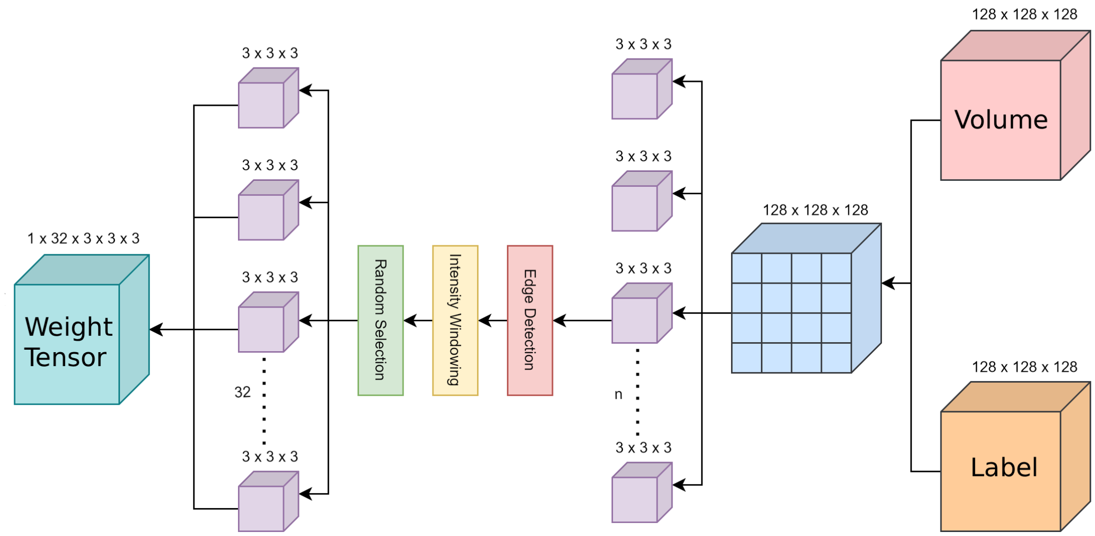

Figure 2: The block-level architecture of GT-based kernel generation. (N) Large number of patches are randomly extracted from the masked CT volume. Afterwards, the weight tensor is created by 32 randomly selected patches after performing edge detection and intensity windowing.

3. RESULTS

3.1. Overall performance

The study encompasses three experiments comparing DICE scores, two on the AMOS dataset with distinct kernel sizes and one on the CT-ORG dataset. Kernel sizes for the AMOS dataset include $3 \times 3 \times 3$ and $7 \times 7 \times 7$, while the CT-ORG dataset exclusively employs the $3 \times 3 \times 3$ kernel. All curves are averaged by 3 runs. The first table in this section (see Table 1) presents the DICE score comparisons at various experiment phases. The DICE score metric reveals the superiority of GT-PCA-based initialization in terms of model convergence across different kernel sizes. Notably, GT-PCA initialization achieves faster convergence and better performance compared to both Xavier and Kaiming initializations. In all cases, GT-PCA starts from a similar point but consistently outperforms both Kaiming and Xavier in mid-graph and subsequent epochs. Furthermore, GT-PCA initialization reaches the same validation dice (0.7, see Figure 5) approximately 40% faster than Kaiming initialization.

	Kaiming	Xavier	GT-PCA
Start score	0.014	0.008	0.013
End score	0.75	0.44	0.78
Best score	0.76	0.48	0.79

Table 1: Dice score comparisons for kernel size $3 \times 3 \times 3$ in AMOS dataset.

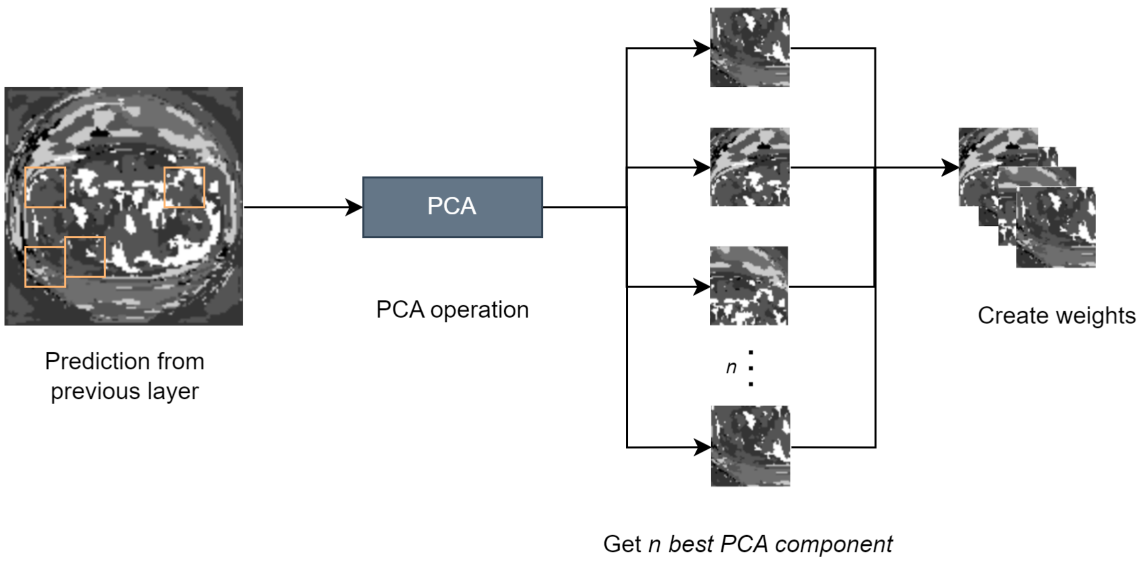

Figure 3: The high-level architecture of PCA based kernel generation. With the initial layer initialized by GT initialization, the next layer is initialized using the output from the first layer with PCA-based selection.

	Kaiming	Xavier	GT-PCA
Start score	0.004	0.008	0.007
End score	0.63	0.35	0.67
Best score	0.63	0.38	0.68

Table 2: IoU score comparisons for kernel size $3 \times 3 \times 3$ in AMOS dataset.

Figures 5 provide an overview of GT-PCA, Kaiming, and Xavier initialization performance on the AMOS dataset. In the $3 \times 3 \times 3$ kernel experiment, GT-PCA starts at 0.014, surpassing the initial scores of Kaiming (0.008) and Xavier (0.014). The concluding DICE scores are 0.78, 0.75, and 0.44, respectively. The peak DICE scores for GT-PCA, Kaiming, and Xavier are 0.76, 0.48, and 0.79. In the $7 \times 7 \times 7$ kernel experiment (Figure 6), GT-PCA initiates at 0.012, outperforming Kaiming (0.009) and Xavier (0.19). The final DICE scores are 0.73, 0.75, and 0.79 for Kaiming, Xavier, and GT-PCA. The best DICE scores for GT-PCA, Kaiming, and Xavier are 0.79, 0.76, and 0.76, respectively. Together with the performance on CT-ORG dataset, GT-PCA consistently outperforms Xavier and Kaiming in terms of overall convergence speed, providing further evidence of its effectiveness in enhancing performance in segmentation tasks.

4. DISCUSSION

This work introduced a novel initialization method, GT-PCA initialization, aiming to incorporate prior knowledge into the segmentation network. By initializing the first encoder

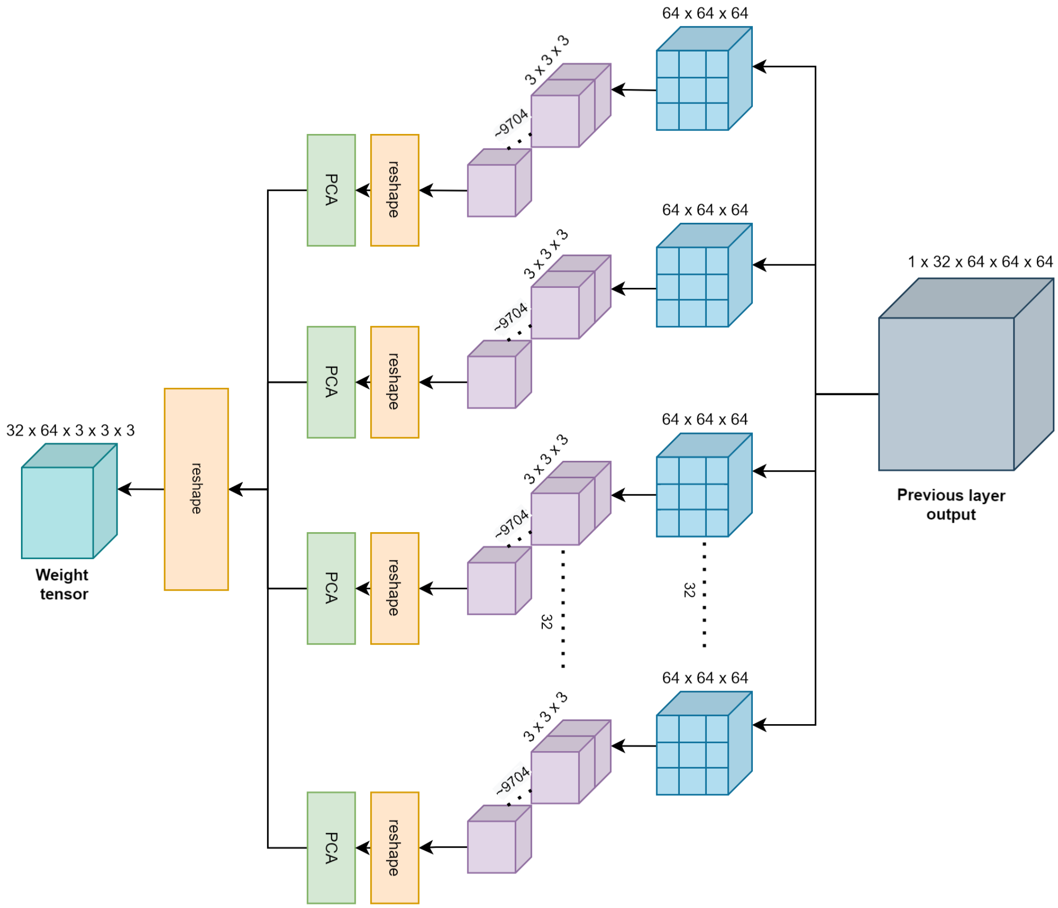

Figure 4: The block-level architecture of PCA based kernel generation. In this example, convolutional kernels of the current layer are of size $32 \times 64 \times 3 \times 3 \times 3$ and are extracted from the previous layer output. Considering a previous layer output tensor of $1 \times 32 \times 64 \times 64 \times 64$, the tensor is decomposed into 32, $64 \times 64 \times 64$ tensors. Patches of $3 \times 3 \times 3$ tensors are generated, resulting in approximately 9704 patches per individual $64 \times 64 \times 64$ tensor. After reshaping and PCA operations, the tensors are combined into a weight tensor of shape $32 \times 64 \times 3 \times 3 \times 3$.

layer with patient reference-based weights, the model gains meaningful kernels tailored to edge cases within a specific intensity range. However, it's crucial to note that this approach introduces some dependency on the ground truth of the image data. Subsequent layers of the encoder employ a specialized Principal Component Analysis (PCA) initialization, which calculates the best principal components from the previous layer's output.

Experimental results, conducted on different datasets (AMOS and CT-ORG), reveal certain advantages of GT-PCA over widely used Xavier and Kaiming initialization methods. GT-PCA exhibits superior convergence speed in certain scenarios, particularly when the

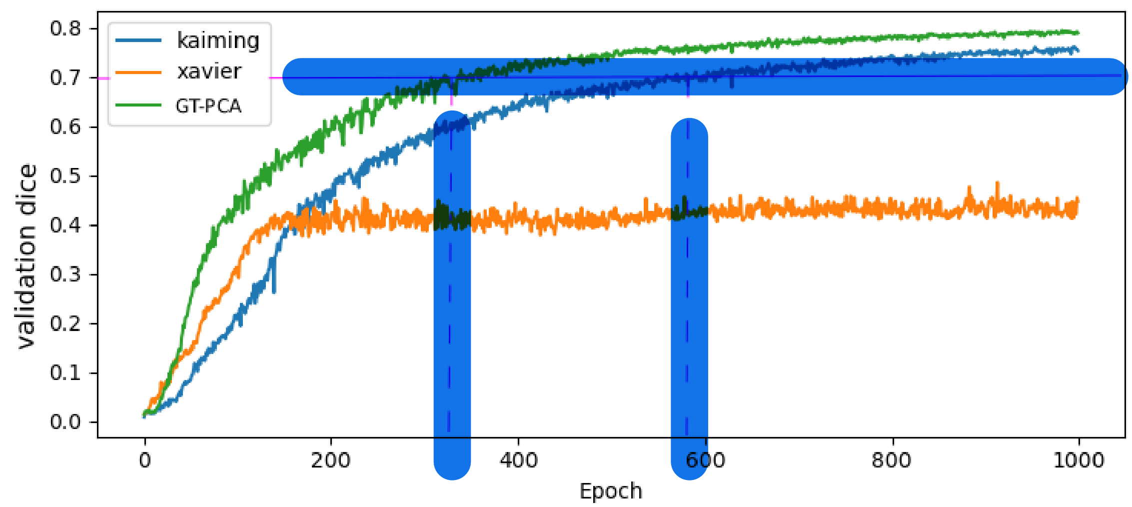

Figure 5: Validation DICE score of AMOS dataset for kernel size $3 \times 3 \times 3$

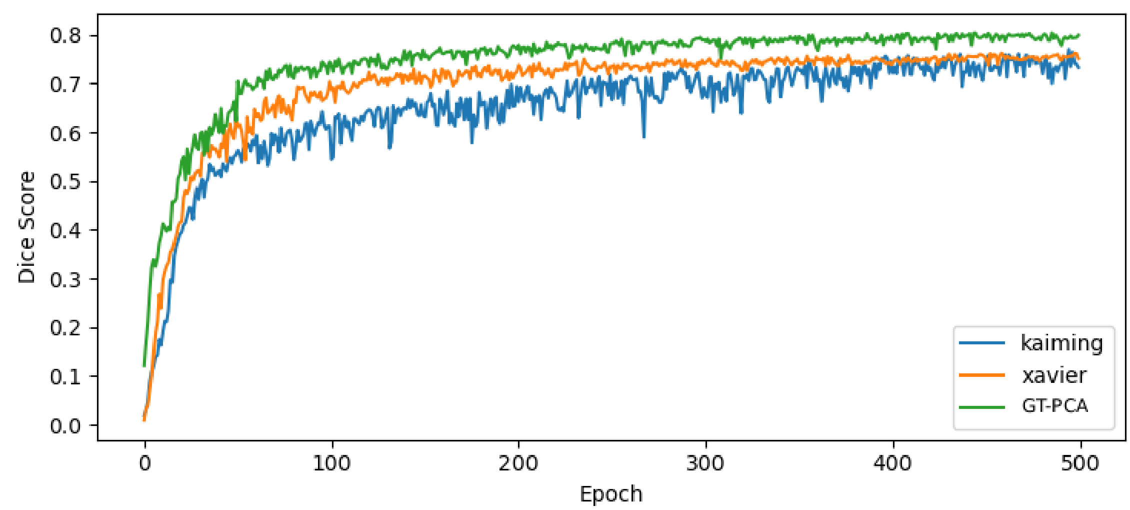

Figure 6: Validation DICE score of AMOS dataset for kernel size $7 \times 7 \times 7$

network is initialized with a larger kernel size (7 x 7 x 7). This is particularly pronounced in the AMOS dataset, where GT-PCA initialization leads to higher Dice scores at the outset.

Despite the observed benefits, the overhead associated with GT-PCA initialization must be considered. Unlike Xavier and Kaiming, which rely on simple Gaussian or normal distribution for weight initialization, GT-PCA involves a more complex process of weight generation based on specifications. Incorporating these weights into the network adds an extra layer of complexity, necessitating an additional epoch or a pre-training step for weight generation.

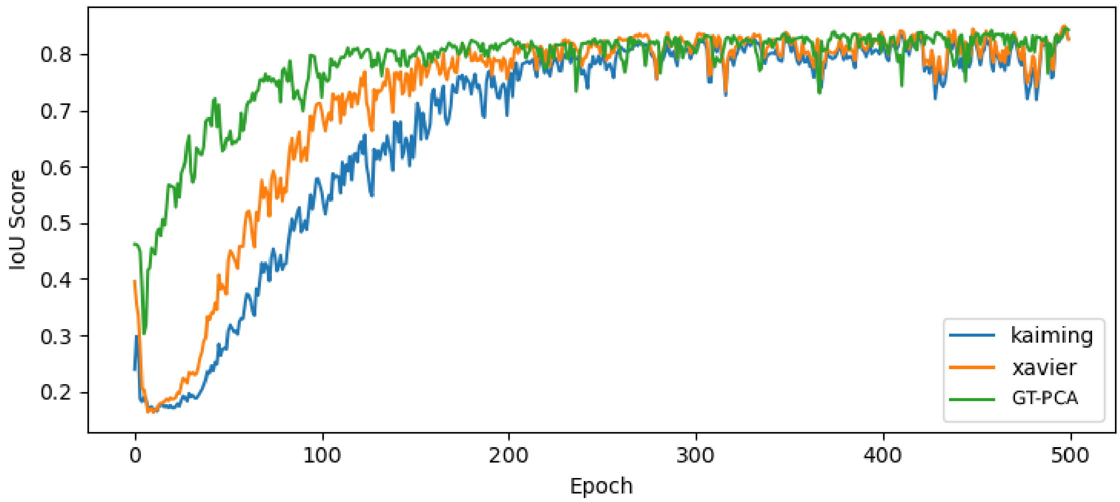

Figure 7: Validation IoU score of CT-ORG dataset for kernel size $3 \times 3 \times 3$.

5. CONCLUSION

The presented methodology of incorporating Ground Truth prior knowledge and PCA-based weight generation for segmentation models (GT-PCA initialization) in medical images yields positive results. The experimental findings indicate improvements in DICE and IoU scores over popular initialization methods e.g. Kaiming initialization and Xavier initialization, both in terms of performance and convergence rates. The introduced GT-PCA approach demonstrates promise in enhancing the performance of segmentation networks, especially useful for models that are not pre-trained with large-scale of dataset. This shows the potential of leveraging prior knowledge and specialized weight initialization methods for advancing segmentation models for tasks with limited training dataset e.g. medical image processing.

References

John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma. Pcanet: A simple deep learning baseline for image classification? *IEEE transactions on image processing*, 24(12):5017–5032, 2015.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, 2015.

Yuanfeng Ji, Haotian Bai, Jie Yang, Chongjian Ge, Ye Zhu, Ruimao Zhang, Zhen Li, Lingyan Zhang, Wanling Ma, Xiang Wan, and Ping Luo. Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation, 2022.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

Siddharth Krishna Kumar. On weight initialization in deep neural networks, 2017.

Dmytro Mishkin and Jiri Matas. All you need is a good init, 2015. URL https://arxiv.org/abs/1511.06422.

Blaine Rister, Darvin Yi, Kaushik Shivakumar, Tomomi Nobashi, and Daniel Rubin. Ct-org, a new dataset for multiple organ segmentation in computed tomography. *Scientific Data*, 7, 11 2020. doi: 10.1038/s41597-020-00715-8.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

Andrew M Saxe, James L McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv preprint arXiv:1312.6120*, 2013.

Chunyu Xu and Hong Wang. Research on a convolution kernel initialization method for speeding up the convergence of cnn. *Applied Sciences*, 12(2):633, 2022.