Ground Truth base Deep Learning Model Initialization for Medical Image Segmentation

Zijin Yang[*1] MICHAEL.YANG@FAU.DE
[1] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany*
Nazmus Sakib[†1] NAZMUS.SAKIB@FAU.DE
[1] *Friedrich-Alexander-Universität Erlangen Nürnberg, Germany*
Chang Liu[1] CHANG.CH.LIU@FAU.DE
[1] *Friedrich-Alexander-Universität Erlangen Nürnberg, Germany*
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. Quantitative assessment was performed using the Dice similarity coefficient as the primary evaluation metric.

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: List of keywords, comma separated.

[*] Contributed equally
[†] Contributed equally

1. Introduction

The develop 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. 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.

This work seeks to enhance PCA-based kernel initialization by integrating ground truth (GT) images and introducing a unique initialization approach, referred to as Patient Reference (PR) - 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

The methodology employed in this study leverages advanced deep learning architectures, specifically focusing on the U-Net and Residual U-Net frameworks. U-Net, designed by Ronnenberg et al.(Ronneberger et al., 2015), addresses challenges in image segmentation

by incorporating skip connections between its encoder and decoder parts. This architectural choice enables the preservation of crucial information during the decoding process. The 2D U-Net model utilized in our study comprises encoder and decoder blocks with a maximum depth of 5 blocks, featuring convolutional operations, batch normalization, ReLU activation, and max-pooling for downsampling.

Expanding our methodology to three-dimensional data, we incorporate the 3D U-Net architecture, essential for medical applications involving voxel-based image processing. Additionally, we integrate the Residual U-Net, a variant developed to counteract performance degradation in complex U-Nets over time. Residual units are strategically inserted into every encoder block, while upsampling units, employing transpose convolution layers, facilitate effective decoding in the corresponding decoder layer.

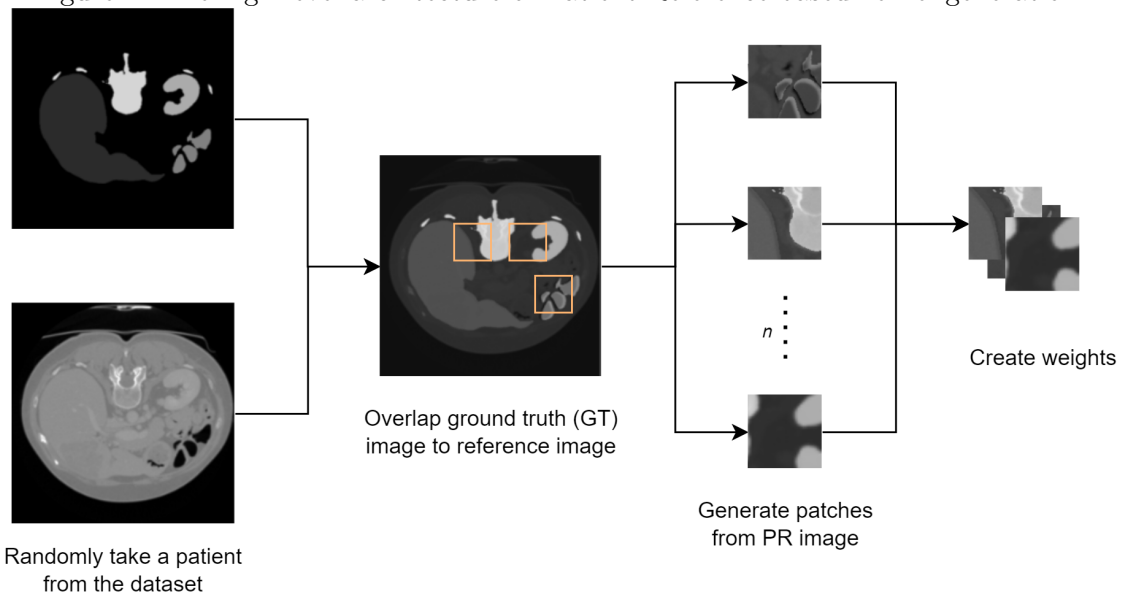
This methodological approach ensures the robust exploration of image-related tasks, capitalizing on the strengths of U-Net and Residual U-Net architectures to address segmentation challenges and enhance overall performance.

2.3. Initialization using GT-PCA

2.3.1. Ground Truth (GT) Initialization

The Ground Truth initialization process commences by selecting a random patient from the dataset, with data structured as 3D voxels having dimensions $(W \times H \times T)$. The network's initial state is represented as a 5-dimensional tensor $(B \times C \times W \times H \times T)$, where B is the batch size and C denotes the initial channel dimension.

Figure 1: The high-level architecture of Patient Reference based kernel generation.



Data Traversal: Beginning from the top-left corner, the algorithm traverses the data horizontally and vertically using strides proportional to the kernel's dimensions. This pro-

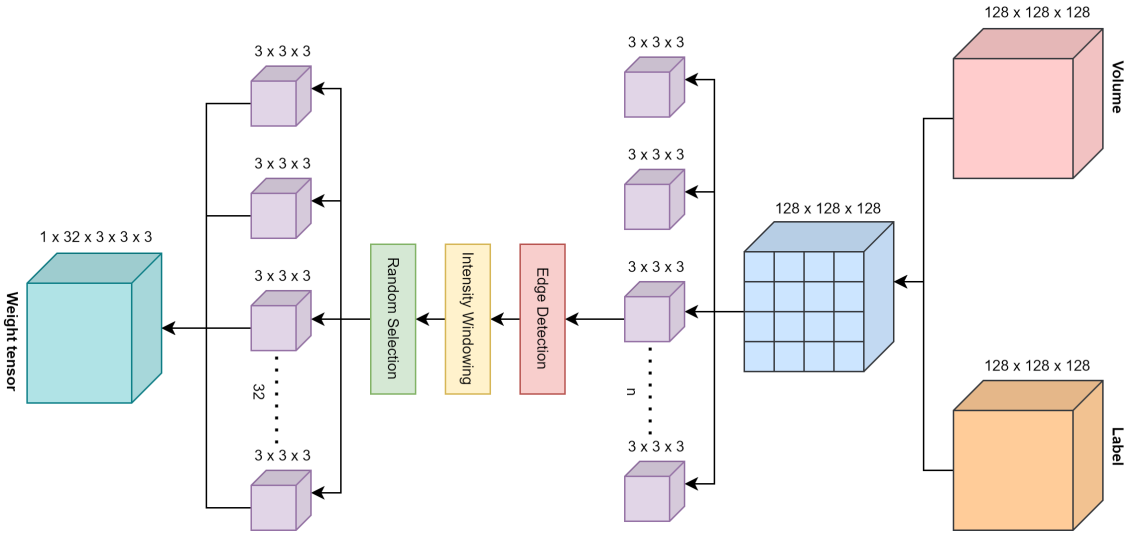


Figure 2: The block-level architecture of Patient Reference based kernel generation.

cess generates "potential kernels" and the actual kernels chosen align with the network's kernel size, for example, $7 \times 7 \times 7$.

Choosing 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, a random selection process yields a number equal to the channel dimension, C. These chosen PR weights are saved and reused for network initialization.

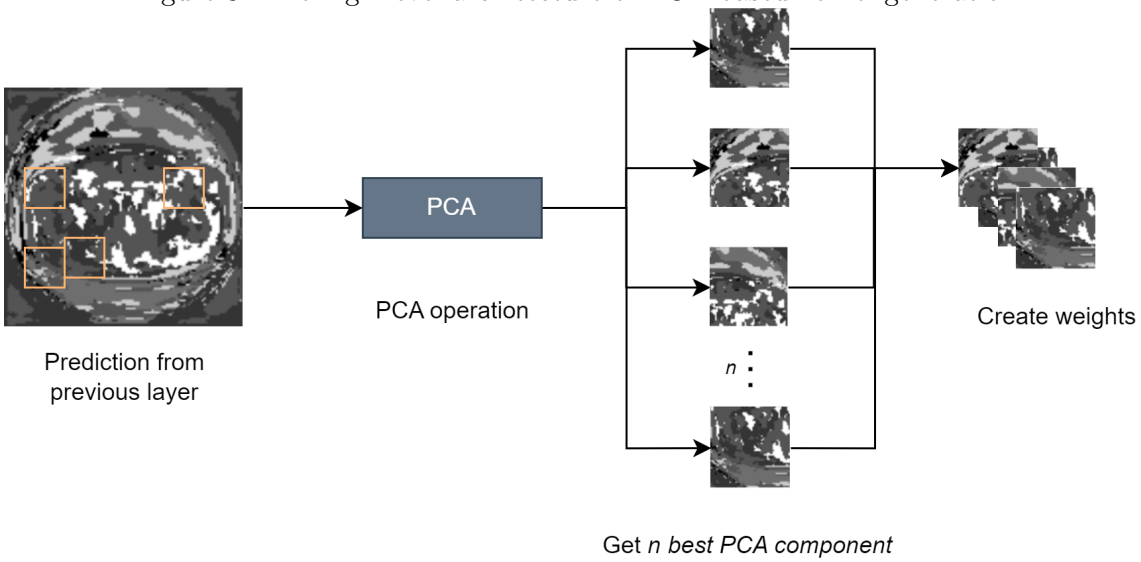
Figure 2 illustrates the weight generation process. A $128 \times 128 \times 128$ volume and label size are combined, yielding $3 \times 3 \times 3$ patches as potential kernels. After edge detection and intensity windowing, a random subset is selected and saved as $b \times n \times 3 \times 3 \times 3$ PyTorch tensors, where b is the batch size, and n is the kernel size matching the network architecture. This ensures congruence between the chosen kernel size and the network configuration.

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: 1. Gathering Output, 2. Generate PCA Weights, and 3. Generate Kernel Weights, as illustrated in Figure 3.

Gathering Output: Beginning from the second layer, the algorithm collects outputs from the preceding layer, generating small patches, e.g., $3 \times 3 \times 3$ tensors, dependent on the network's kernel size. The spatial shape of these tensors is determined by the equation $size = (\frac{s}{k})^k$, where s represents the spatial shape of the output, and k is the kernel shape.

Figure 3: The high-level architecture of PCA based kernel generation.



Generate PCA Weights: The individual four-dimensional outputs undergo reshaping, with the size dimension retained while shapes are flattened. Principle components are then derived through dimensional reduction, yielding n components required for the subsequent layer.

Generate Kernel Weights: The final step involves combining and resizing tensors to generate PyTorch tensors, which are saved for future model use.

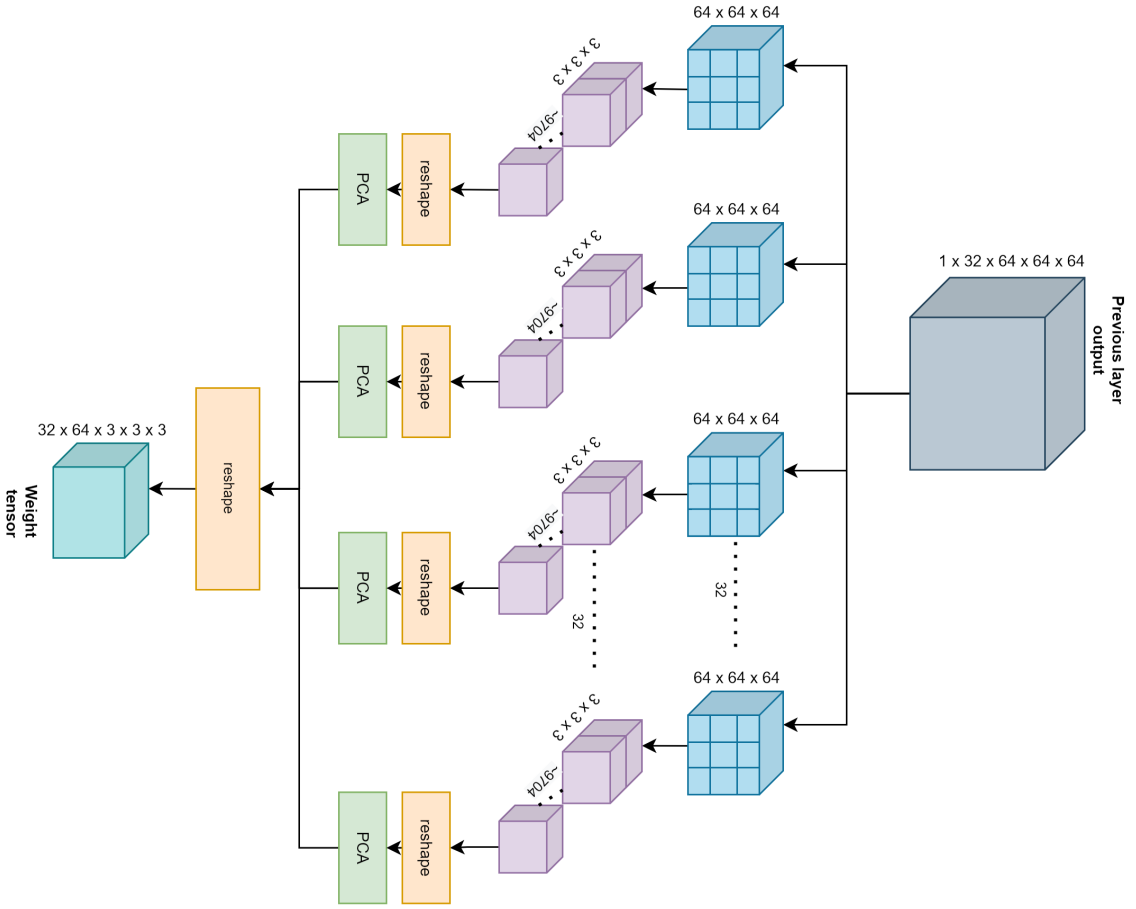
For instance, considering a previous layer output tensor of $1 \times 32 \times 64 \times 64 \times 64$, the tensor is decomposed into 32, $64 \times 64 \times 64$ tensors. Patches of $3 \times 3 \times 3$ tensors are generated, resulting in approximately 9704 patches per individual $64 \times 64 \times 64$ tensor. After reshaping and PCA operations, the tensors are combined into a weight tensor of shape $32 \times 64 \times 3 \times 64 \times 64$.

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. 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 compare 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. Remarkably, GT-PCA and Kaiming exhibit convergence, while

Figure 4: The high-level architecture of PCA based kernel generation.



Xavier initialization fails to converge to a reasonable minima with the smaller kernel during experiments.

	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.

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

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

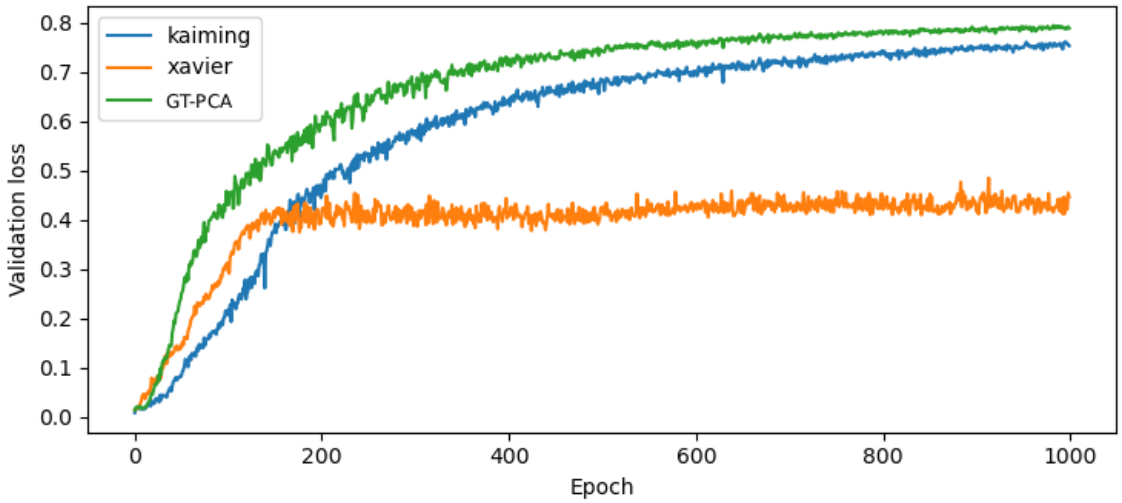
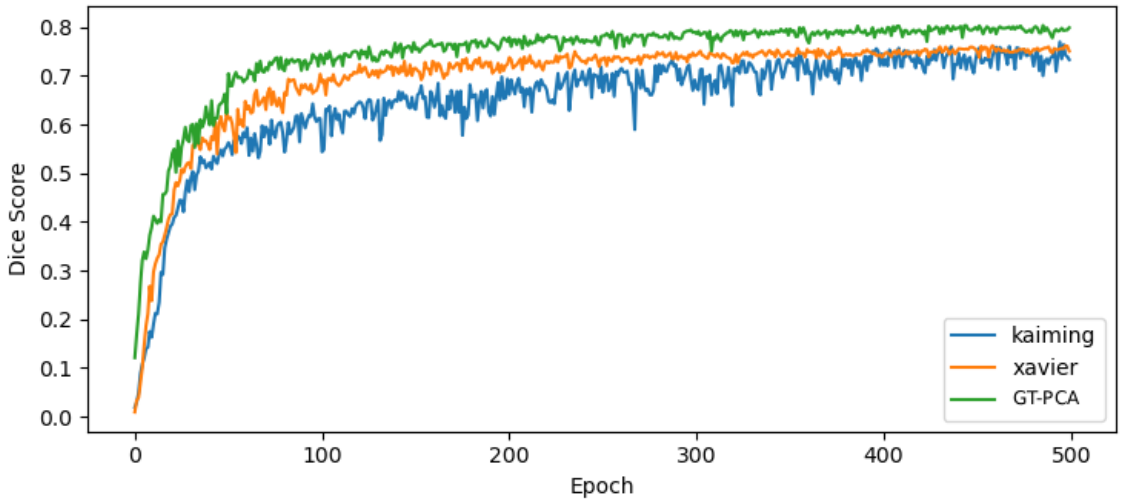


Figure 6: Validation DICE score of AMOS dataset for kernel size $7 \times 7 \times 7$

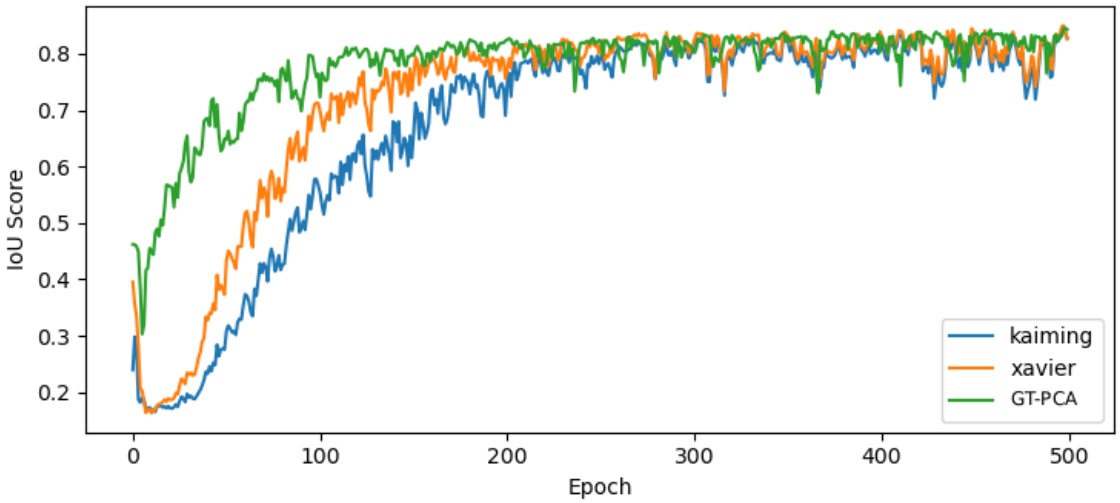


scores for GT-PCA, Kaiming, and Xavier are 0.79, 0.76, and 0.76, respectively. Together with the performance on CT-ORG dataset, GT-PCA consistently outperforms Xavier and Kaiming in terms of overall convergence speed, providing further evidence of its effectiveness in enhancing performance in segmentation 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 layer with patient reference-based weights, the model gains meaningful kernels tailored to

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



	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.

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 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.

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.