




## Article

# Deep Learning Framework with Multi-Head Dilated Encoders for Enhanced Segmentation of Cervical Cancer on Multiparametric Magnetic Resonance Imaging

Reza Kalantar <sup>1,2</sup> , Sebastian Curcean <sup>3</sup>, Jessica M. Winfield <sup>1,2</sup>, Gigin Lin <sup>4</sup> , Christina Messiou <sup>1,2</sup>, Matthew D. Blackledge <sup>1,2,\*</sup>  and Dow-Mu Koh <sup>1,2</sup>

<sup>1</sup> Division of Radiotherapy and Imaging, The Institute of Cancer Research, London SW7 3RP, UK; reza.kalantar@icr.ac.uk (R.K.); jessica.winfield@icr.ac.uk (J.M.W.); christina.messiou@rmh.nhs.uk (C.M.); mu.koh@icr.ac.uk (D.-M.K.)

<sup>2</sup> Department of Radiology, The Royal Marsden Hospital, London SW3 6JJ, UK

<sup>3</sup> Department of Radiation Oncology, Iuliu Hatieganu University of Medicine and Pharmacy, 400347 Cluj-Napoca, Romania; sebastian.curcean@iocn.ro

<sup>4</sup> Department of Medical Imaging and Intervention, Chang Gung Memorial Hospital at Linkou, Chang Gung University, Guishan, Taoyuan 333, Taiwan; giginlin@cgmh.org.tw

\* Correspondence: matthew.blackledge@icr.ac.uk

**Abstract:** T<sub>2</sub>-weighted magnetic resonance imaging (MRI) and diffusion-weighted imaging (DWI) are essential components of cervical cancer diagnosis. However, combining these channels for the training of deep learning models is challenging due to image misalignment. Here, we propose a novel multi-head framework that uses dilated convolutions and shared residual connections for the separate encoding of multiparametric MRI images. We employ a residual U-Net model as a baseline, and perform a series of architectural experiments to evaluate the tumor segmentation performance based on multiparametric input channels and different feature encoding configurations. All experiments were performed on a cohort of 207 patients with locally advanced cervical cancer. Our proposed multi-head model using separate dilated encoding for T<sub>2</sub>W MRI and combined b1000 DWI and apparent diffusion coefficient (ADC) maps achieved the best median Dice similarity coefficient (DSC) score, 0.823 (confidence interval (CI), 0.595–0.797), outperforming the conventional multi-channel model, DSC 0.788 (95% CI, 0.568–0.776), although the difference was not statistically significant ( $p > 0.05$ ). We investigated channel sensitivity using 3D GRAD-CAM and channel dropout, and highlighted the critical importance of T<sub>2</sub>W and ADC channels for accurate tumor segmentation. However, our results showed that b1000 DWI had a minor impact on the overall segmentation performance. We demonstrated that the use of separate dilated feature extractors and independent contextual learning improved the model's ability to reduce the boundary effects and distortion of DWI, leading to improved segmentation performance. Our findings could have significant implications for the development of robust and generalizable models that can extend to other multi-modal segmentation applications.

**Keywords:** tumor segmentation; multiparametric MRI; cervical cancer; deep learning; dilated convolution; radiology; radiation oncology



**Citation:** Kalantar, R.; Curcean, S.; Winfield, J.M.; Lin, G.; Messiou, C.; Blackledge, M.D.; Koh, D.-M. Deep Learning Framework with Multi-Head Dilated Encoders for Enhanced Segmentation of Cervical Cancer on Multiparametric Magnetic Resonance Imaging. *Diagnostics* **2023**, *13*, 3381. <https://doi.org/10.3390/diagnostics13213381>

Academic Editors: Francis Bui, Mohammad Ali Moni and Kawsar Ahmed

Received: 26 September 2023

Revised: 29 October 2023

Accepted: 1 November 2023

Published: 3 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cervical cancer is the fourth most common cancer in women worldwide [1]. In the epidemiological data of 2020, a total of 604,127 cervical cancer diagnoses were reported globally [2]. Despite the evolution in diagnostic and therapeutic modalities, projections for 2023 indicate a persistent challenge, with an anticipated 13,960 cases of invasive cervical cancer in the United States alone. This forecast also predicts a consequent 4310 mortality events [3]. Such statistics emphasize the imperative of advancements in precise diagnostic measures, including state-of-the-art imaging and segmentation methodologies.

Multiparametric magnetic resonance imaging (mpMRI) is a crucial tool in the diagnosis and management of gynecological malignancies, including cervical cancer. It provides detailed anatomical and functional information, which is applied for disease staging, treatment planning, response monitoring, and surveillance for disease recurrence [4–6]. An important aspect of many mpMRI protocols is diffusion-weighted imaging (DWI), which enhances the contrast and visualization of cellular tumors. DWI is sensitive to the rate of diffusion of water molecules *in vivo*, and offers insights into the intricate tissue microenvironment. Specifically, it has been shown to relate to cellularity, *i.e.*, the density of cells within a tumor, and is also sensitive to microstructural alterations, such as necrosis and changes in the extracellular matrix, which can influence water diffusion patterns [7]. The rate of diffusion can be quantified at each spatial location through the estimation of maps of the apparent diffusion coefficient (ADC). ADC measurements offer the potential to be used as a surrogate biomarker of tumor grade [8,9], and have shown promising results in identifying early treatment responses, making them desirable for monitoring therapeutic outcomes in cervical cancer [10,11]. In conventional anatomical T<sub>2</sub>-weighted (T<sub>2</sub>W) MRI, primary and metastatic tumors exhibit intermediate to high signal intensities. This is used to identify cervical abnormalities, as well as for disease staging and directing MRI-guided interventions [12,13].

Image segmentation in medical imaging involves the demarcation of regions of interest (ROIs) into semantically and clinically meaningful areas. Automating disease detection and delineation in medical images is a critical task, primarily because it aids in extracting valuable biomarkers from images, which enhances clinical decision-making. This process is currently impeded by the requirement for extensive annotated datasets, leading to a high dependency on clinicians and potential inconsistencies due to human contouring variations. Automated tumor segmentation on mpMRI therefore holds great significance, not only for reducing the burden on clinicians but also for its potential in improving accuracy and consistency. Furthermore, in contexts such as treatment planning, where manual delineation may not be feasible due to time constraints, these automated methods become particularly indispensable. Therefore, the development of fully automatic tumor segmentation techniques is a crucial step towards achieving more efficient and reliable clinical processes. Beyond operational benefits, this could also translate to profound patient-centric advantages, including shorter and more comfortable treatments, as well as more accurate and consistent treatment delivery. Such advancements could potentially mitigate collateral toxicity and optimize treatment outcomes.

Among these segmentation techniques, U-Net [14] stands out as a prevalent architecture for image segmentation. Numerous derivatives of U-Net have been developed for various medical segmentation tasks, with a particular emphasis on neuroimaging. These include the multi-scale densely connected U-Net (MDU-Net) [15], U-Net with interconnected skip connection pathways (U-Net++) [16], U-Net with residual extended skip connection and wide context modules (BU-Net) [17], U-Net with a feature enhancer block (BrainSeg-Net) [18], multi-scale recurrent residual U-Net with dense skip connections (R2U++) [19], the redesigned U-Net with full-scale feature fusion and ghost modules (half-U-Net) [20], U-Net augmented with multi-scale feature fusion (SegR-Net) [21], U-shaped architecture with residual spatial pyramid pooling modules and attention gates (RAAGR2-Net) [22], and the pyramid dual-decoder attention U-Net (PDAtt-UNet) [23].

In parallel, architectures based on the vision transformer model [24] have shown promise for the accurate segmentation of medical images. For instance, Wang et al. [25] developed the mixed transformer U-Net for multi-organ segmentation in computed tomography (CT) images. Furthermore, Huang et al. [26] proposed a transformer-based generative adversarial network for multimodal brain tumor segmentation in MRI.

Despite these advancements, a significant challenge persists in accurately segmenting small-scale tumors in mpMRI images, particularly when there is spatial misalignment between the input channels of the models. This issue is especially prominent in pelvic images, where internal tissues are prone to spatial deformation during scanning sessions.

Such deformations could impede the segmentation performance of networks that rely on conventional multi-channel inputs.

While current segmentation architectures demonstrate promising versatility for complex medical segmentation applications, there remains a scarcity of research on cancer tumor segmentation utilizing multiparametric MRI, especially in the context of cervical cancer. The segmentation of cervical cancer presents several challenges. These include the complex anatomy of the pelvis with closely situated organs, the appearance of inflammatory processes that can mimic or obscure tumor regions, and the potential presence of metastases that might be intermingled with the primary tumor or situated nearby. Moreover, the inherent variability of tumor appearances, challenges arising from low image resolutions or noisy DWI images, and the intricate distinctions between the primary cervical mass and adjacent lymph nodes further complicate the segmentation process.

Several studies have explored the DL-based segmentation of cervical cancer in MRI [27–31]. Specifically, some studies have combined semantic knowledge between T<sub>2</sub>W MRI and DWI/ADC in this context. Among related studies, Lin et al. [32] developed a U-Net model for segmenting cervical cancer in DWI and ADC images, and found that multi-channel input (b<sub>0</sub>, b<sub>1000</sub>, and ADC) produced the best segmentation performance. However, this study focused on two-dimensional (2D) images and did not incorporate multimodal MRI (e.g., MRI images derived from different sequences). Kano et al. [33] combined 2D and three-dimensional (3D) U-Net models for cervical tumor segmentation on DWI images using an ensembling approach. Yoganathan et al. [34] segmented primary tumors along with organs-at-risk (OARs) on T<sub>2</sub>W MRI, reporting that integrating segmentations from 2.5D training in axial, coronal, and sagittal planes improved segmentation performance compared with previous 2D models. However, this study was limited to 39 patients and single-channel inputs. Wang et al. [29] proposed a 3D CNN model for cervical tumor segmentation on multimodal MRI, while Hodneland et al. [35] utilized a U-Net with residual connections, employing T<sub>2</sub>W MRI, b<sub>1000</sub> DWI, and ADC as input channels. However, none of these studies examined the impact of spatial mismatch between multimodal MRI inputs, resulting from distortion in echo-planar imaging (EPI) and soft-tissue deformations between scans [36,37], on cervical tumor segmentation outcome.

The aim of this study was to develop a novel 3D framework based on the U-Net architecture that included multi-head dilated residual encoding blocks for combined fine-grained and contextual feature aggregation and training on anisotropic sub-volumes of images, enhancing the segmentation of locally advanced cervical tumors on multiparametric MRI. To our best knowledge, no previous studies have incorporated and investigated this strategy for the automated segmentation of pelvic malignancies.

## 2. Materials and Methods

### 2.1. Patient Populations and Imaging Parameters

In this study, we utilized a retrospective cohort of 207 patients diagnosed with cervical cancer. The histopathology types of enrolled patients included squamous cell, adenosquamous carcinomas (either non-keratinizing or keratinizing), and all were HPV-associated, according to the World Health Organization (WHO) classification of female genital tumors introduced in the 4th edition (2014) [38]. The categorization of locally advanced cervical cancer was based on the 2018 guidelines set by the International Federation of Gynaecology and Obstetrics (FIGO) [39]. These patients underwent pelvic T<sub>2</sub>W MRI, and DWI on a 3T MAGNETOM TrioTim MRI scanner (Siemens Healthcare, Erlangen, Germany) was selected. The ground-truth tumor contours were defined by a clinician with 3 years of relevant experience on T<sub>2</sub>W MRI images, with the DWI data available for all patients. The acquisition parameters for this dataset are shown in Table 1. The ADC maps were calculated from DWI images with varying diffusion-weighting magnitude (b-value) (Equation (1)),

using a mono-exponential fit for two b-values (b<sub>0</sub>, b<sub>1000</sub>) (Equation (2)) and least-square exponential fit for multiple b-values (b<sub>200</sub>, b<sub>600</sub>, b<sub>1000</sub>) (Equation (3)).

$$S_{b_i} = S_{b_0} \cdot e^{(-b_i D)} \tag{1}$$

$$D = -\frac{1}{b_i} \ln\left(\frac{S_{b_i}}{S_{b_0}}\right) \tag{2}$$

$$D = -\frac{N \sum_{i=1}^N b_i \ln(S_{b_i}) - \sum_{i=1}^N b_i \sum_{i=1}^N \ln(S_{b_i})}{N \sum_{i=1}^N b_i^2 - (\sum_{i=1}^N b_i)^2} \tag{3}$$

where  $b_i$  is the b-value,  $S_{b_0}$  represents the signal intensity with no diffusion weighting ( $b = 0$ ),  $S_{b_i}$  is the signal intensity at  $b_i$ ,  $N$  is the number of b-values and  $D$  denotes the ADC value.

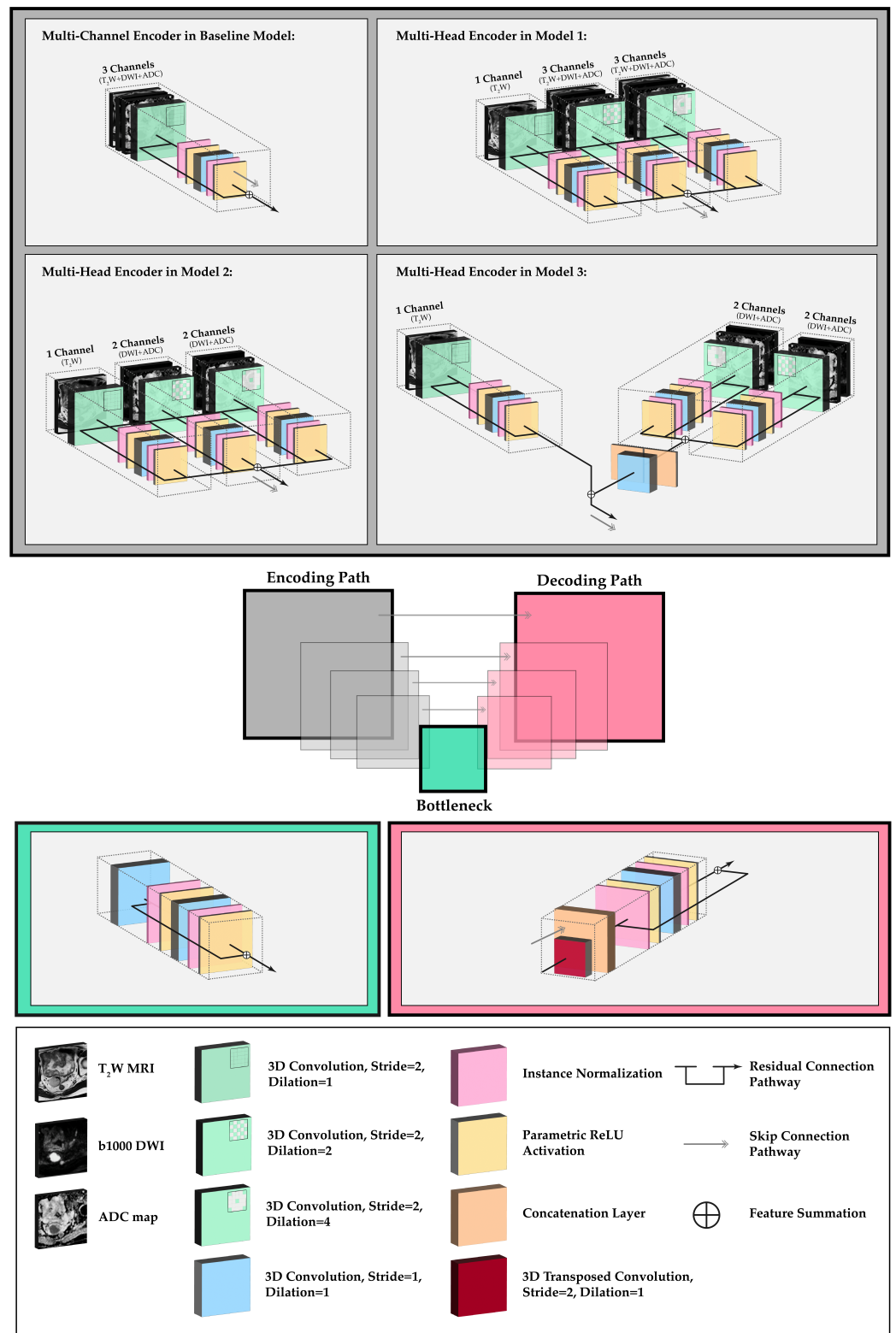
**Table 1.** Imaging parameters of the cohort used in this study ( $n = 207$ ). The values are presented as the range from minimum to maximum values, or are variables using different protocols within the dataset.

Parameter	T <sub>2</sub> W MRI	DWI
Manufacturer Name	Siemens Healthineers	Siemens Healthineers
Scanner Model	MAGNETOM Trio	MAGNETOM Trio
Magnetic Field Strength (T)	3	3
Sequence	Turbo Spin Echo (TSE)	Echo-Planar Imaging (EPI)
Slice Orientation	Axial	Axial
Echo Time (ms)	80–101	60–80
Repetition Time (ms)	3600–8060	3300–10,844
Acquired Matrix Size (read)	224–320	128–172
Reconstructed Matrix Size (read)	256–320	240–248
Reconstructed Pixel Size (mm <sup>2</sup> )	0.5 × 0.5–0.8 × 0.8	1.2 × 1.2–1.4 × 1.4
Slice Thickness (mm)	4.0–5.0	4.0–5.0
Flip Angle (°)	120–160	180
Phase Encoding Direction	Anterior–Posterior or Left–Right	Anterior–Posterior
Receiver Bandwidth (Hz/pixel)	190–200	1940–2441
b-values (s/mm <sup>2</sup> )	-	[0,1000] or [200,600,1000]

### 2.2. Network Topology and Architectural Experiments

In this study, we employed a residual U-Net model, a modified version of the conventional U-Net [14], as the benchmark for our segmentation framework. The model architecture was based on an encoder–decoder structure with symmetrical skip connections at each level (Figure 1). Each encoding level consisted of two residual blocks with 3D convolutional layers followed by instance normalization and a parametric ReLU activation layer (PReLU). Each upsampling block included two residual blocks with 3D strided transposed convolutions and skip connection concatenation layers. The architecture included four downsampling steps with kernel filters of 32, 64, 128, and 256 and a feature map depth of 512 in the bottleneck. Downsampling operation was performed in the first convolutional layer in the block (stride = 2). The model was trained on (i) T<sub>2</sub>W MRI-only; (ii) T<sub>2</sub>W and ADC; and (iii) T<sub>2</sub>W, ADC, and b1000 DWI training data, with input channels set to 1, 2, and 3, respectively.

To address the boundary effects of DWI in multi-channel training, as the result of distortions from EPI compared with morphological imaging, a series of architectural experiments were performed. In the first experiment, we replaced the first encoder block in the baseline multi-channel model with three encoding heads. The first head utilized a non-dilated 3D convolutional operation on T<sub>2</sub>W MRI inputs only. The other two heads included 3D convolution operations with dilations of 2 and 4 respectively, accepting three input channels: T<sub>2</sub>W, b1000 DWI, and ADC. These heads were connected by residual weight sharing and feature activation summation nodes (multi-head model 1).



**Figure 1.** The network topology and the multi-head segmentation training experiments. The operational blocks incorporate residual connections to facilitate the flow of information between different layers. The multi-head models use various encoding and weight-sharing configurations for T<sub>2</sub>W MRI, b1000 DWI and ADC maps using multiple heads with dilated convolution and connective residual operations.

In the second multi-head model, we employed the separate encoding of T<sub>2</sub>W MRI, and combined b1000 DWI and ADC images while still maintaining weight sharing. The dilated heads in this model only included the b1000 DWI and ADC channels. This was performed to investigate the impact of these channels on the overall contextual learning (multi-head model 2).

Finally, in the third experiment, we employed a multi-head encoding strategy with no weight sharing between the T<sub>2</sub>W and b1000 DWI/ADC heads. Similar to the approach used for the second model, we applied dilated convolutions exclusively to the b1000 DWI and ADC images to facilitate independent contextual learning and reduce boundary-based attention. However, this architecture included an additional concatenation layer followed by a convolutional layer for dimensionality reduction and feature summation with the T<sub>2</sub>W head (multi-head model 3).

### 2.3. Image Pre-Processing and Implementation Details

Prior to training, all mpMRI images were resampled to an in-plane resolution of  $0.6 \times 0.6 \text{ mm}^2$ , which was the most common T<sub>2</sub>W MRI resolution in the dataset (approximately 85%), and a slice thickness of 4 mm. Bilinear interpolation was applied for the resampling of T<sub>2</sub>W and b1000 DWI images, as it allows for smooth intensity transitions, preserving the details within the images. On the other hand, ADC images and contours were resampled using the nearest-neighbor interpolation method. This method was selected due to its ability to retain the original discrete values of the images, which is crucial for ADC maps that are quantitative in nature, and for contours as they represent categorical labels or boundaries in segmentation tasks. Each channel was independently normalized to a mean of zero and unit variance. Finally, the dataset was randomly split into 157, 25, and 25 patients for training, validation, and testing, respectively.

During training, sub-volumes of size  $256 \times 256 \times 16$  voxels were extracted stochastically from the training data, ensuring that each patch contained at least one annotated tumor slice. Random data augmentation operations, such as intensity shifting, scaling, and cropping, were applied to improve network generalizability. The models were trained for 100,000 iterations using the Dice loss function, which outperformed the combined Dice and cross-entropy and Tversky [40] losses during the initial experiments on input channels. The Adam optimizer with an initial learning rate of  $1 \times 10^{-4}$  and weight decay of  $1 \times 10^{-5}$  was used, and a cosine annealing learning rate scheduler was employed after each epoch. Validation was performed after each epoch, based on the Dice scores of whole image volumes, and the best-performing weights were saved. Volumetric segmentations were generated using a sliding window algorithm with a 75% overlap between adjacent patches. The models were evaluated using the Dice similarity coefficient (DSC), 95th percentile Hausdorff distance (HD), mean surface distance (MSD), and percentage relative volume similarity metrics [41]. PyTorch and Monai [42] DL libraries were used for all implementations.

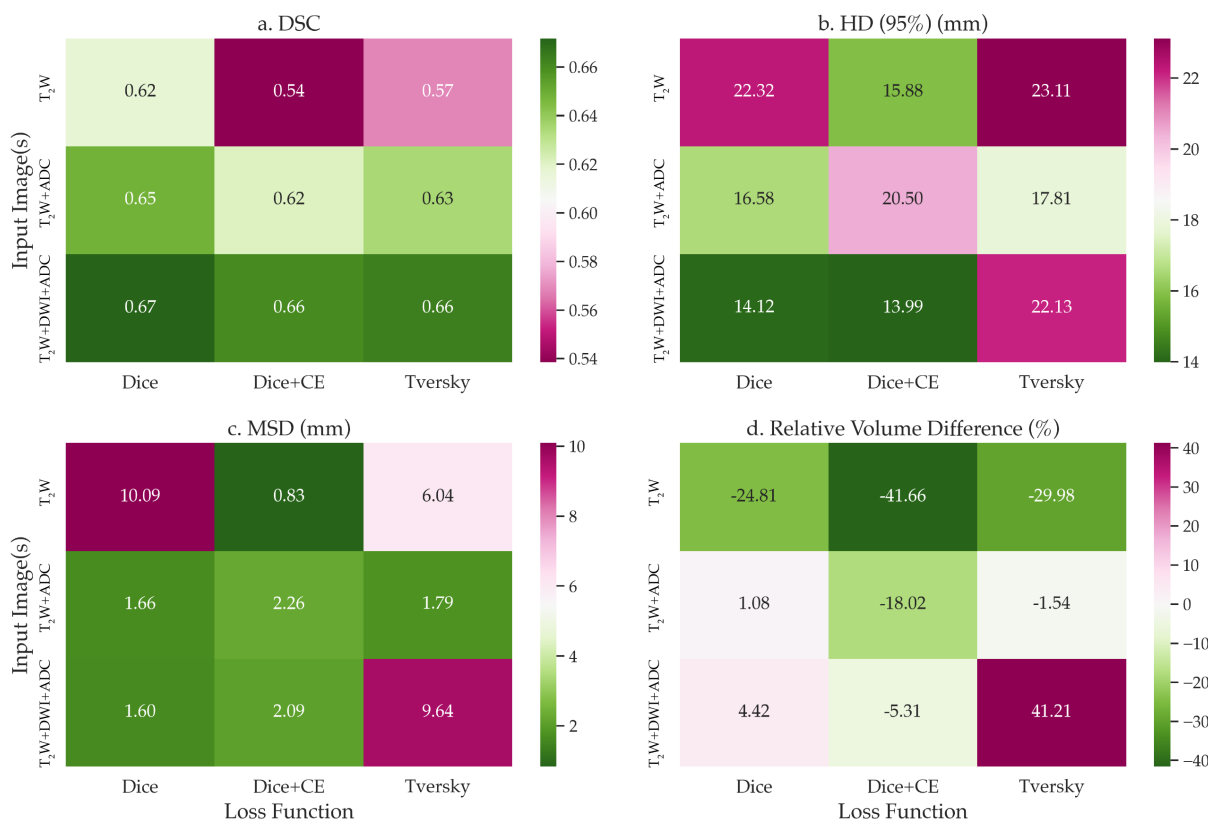
### 2.4. Channel Sensitivity Analysis and Visualization

To assess the significance of individual channels in our models, we conducted a channel sensitivity analysis using sequential channel dropout. By setting each channel to zero one at a time, we compared the segmentation results obtained from the baseline multi-channel and proposed multi-head models to those achieved with no channel dropout. We employed the same quantitative metrics utilized in our previous analyses to perform a comprehensive comparison between the proposed architectures. To delve deeper into the channel-wise significance within our models, we incorporated a 3D adaptation of Gradient-weighted Class Activation Mapping (GRAD-CAM) [43]. GRAD-CAM serves as a method to visually interpret the relevance of particular regions within an input image in relation to the final prediction of a DL model. It does so by leveraging the gradient information flowing into the last convolutional layer of the model to generate a coarse localization map, thereby highlighting regions of interest.

In the context of our research, we tailored this technique to cater to the 3D data modality. The produced saliency maps from the penultimate layers of the investigated models explicitly emphasized the areas of highest importance for the segmentation tasks at hand. Specifically, when this technique was applied to the center-cropped patches sourced from the test image volumes, the resulting visualizations not only detailed the salient features but also provided insight into the channel-wise contributions. Such intricate visual insights are paramount in decoding the rationale behind the model’s predictions and further fine-tuning its segmentation capabilities.

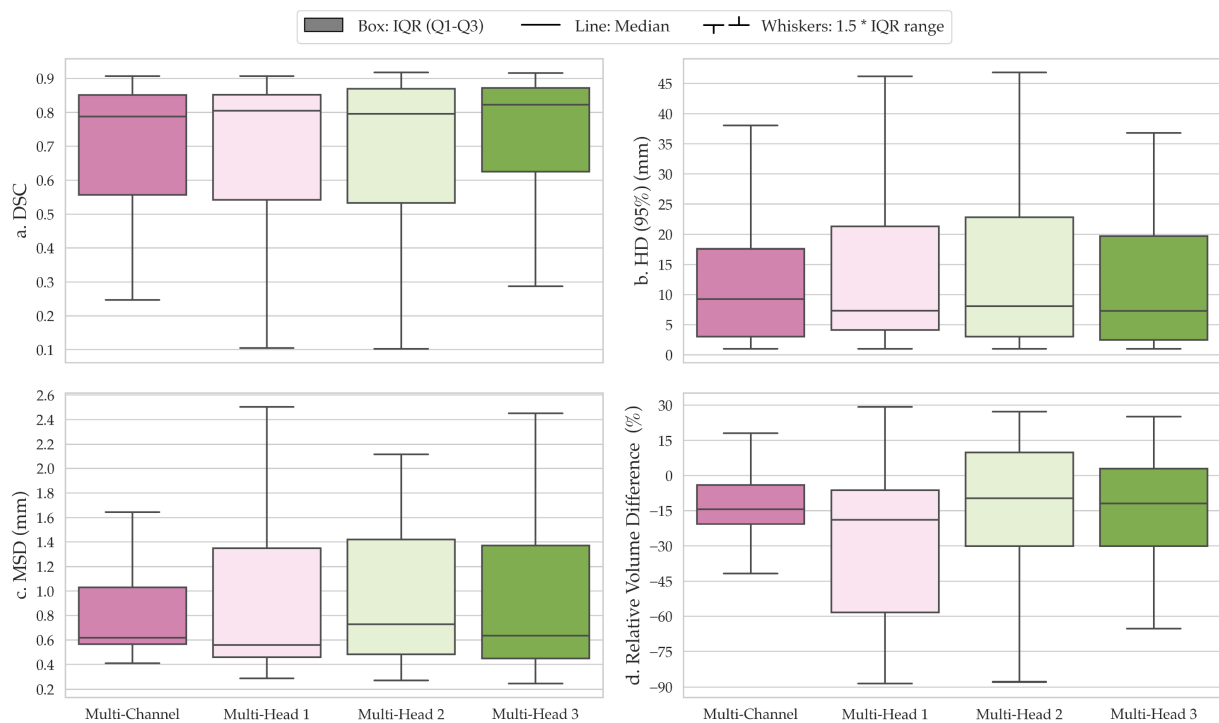
### 3. Results

The baseline multi-channel model was evaluated on various input channels and loss functions. Multi-channel input trained with Dice loss achieved the best mean segmentation performance in terms of DSC across all test cases, with consistent results across input channels. The average DSC values for multi-channel input ( $T_2W$ , b1000 DWI, ADC) using Dice, combined Dice and cross-entropy, and Tversky losses were 0.672, 0.661, and 0.664, respectively (Figure 2).



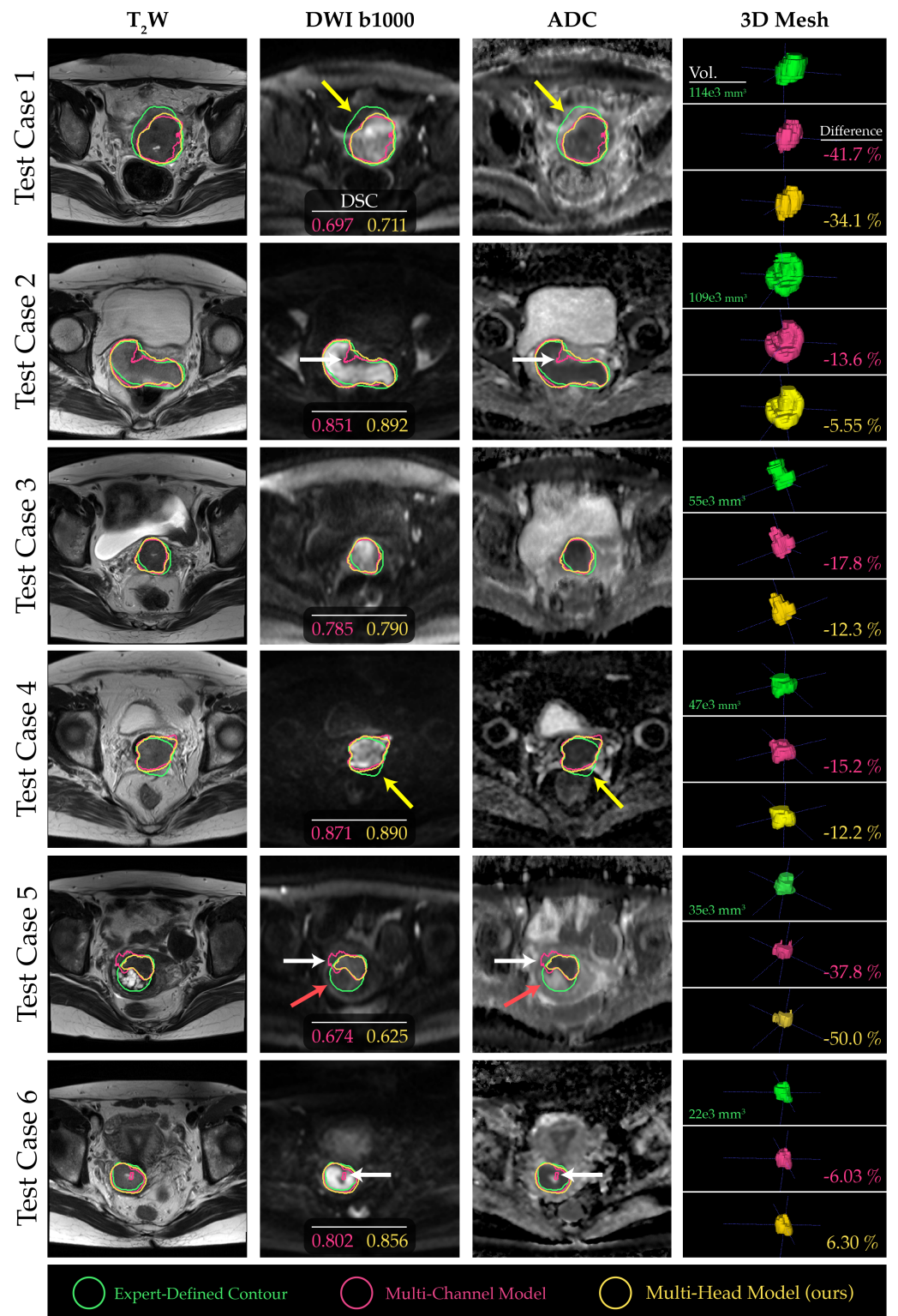
**Figure 2.** Heatmaps of mean (a) Dice similarity coefficient (DSC), (b) 95th percentile Hausdorff distance (HD), (c) mean surface distance (MSD), and (d) relative percentage volume difference for the multi-channel residual U-Net model trained across different input images and loss functions from the test data. The white color within the volume difference colourmap indicates the optimal metrics. The model trained on all  $T_2W$ , b1000 DWI, and ADC channels, and with Dice loss, achieved the best overall segmentation performance across all experiments.

The comparison of segmentations obtained from the baseline multi-channel and the proposed multi-head models revealed that multi-head model 3 exhibited superior performance across the quantitative metrics analyzed (Figure 3a–d). Specifically, the median DSC values for the multi-channel model, multi-head model 1, multi-head model 2, and multi-head model 3 were 0.788 (confidence interval (CI), 0.568–0.776), 0.805 (CI, 0.538–0.769), 0.796 (CI, 0.537–0.776), and 0.823 (CI, 0.595–0.797), respectively (Figure 3a). However, the differences in performance between the proposed multi-head models were not significantly different from the multi-channel model ( $p > 0.05$ ). On average, all models underestimated tumor volume compared with the contours delineated by the clinician, with median relative percentage volume differences for each model of  $-14.4\%$ ,  $-18.9\%$ ,  $-9.7\%$ , and  $-12.0\%$ , respectively (Figure 3d). Multi-head model 3 demonstrated the best quantitative scores compared with other experimental architectures; therefore, the segmentation contours predicted by this model were compared with those obtained from the baseline multi-channel model (Figure 4).



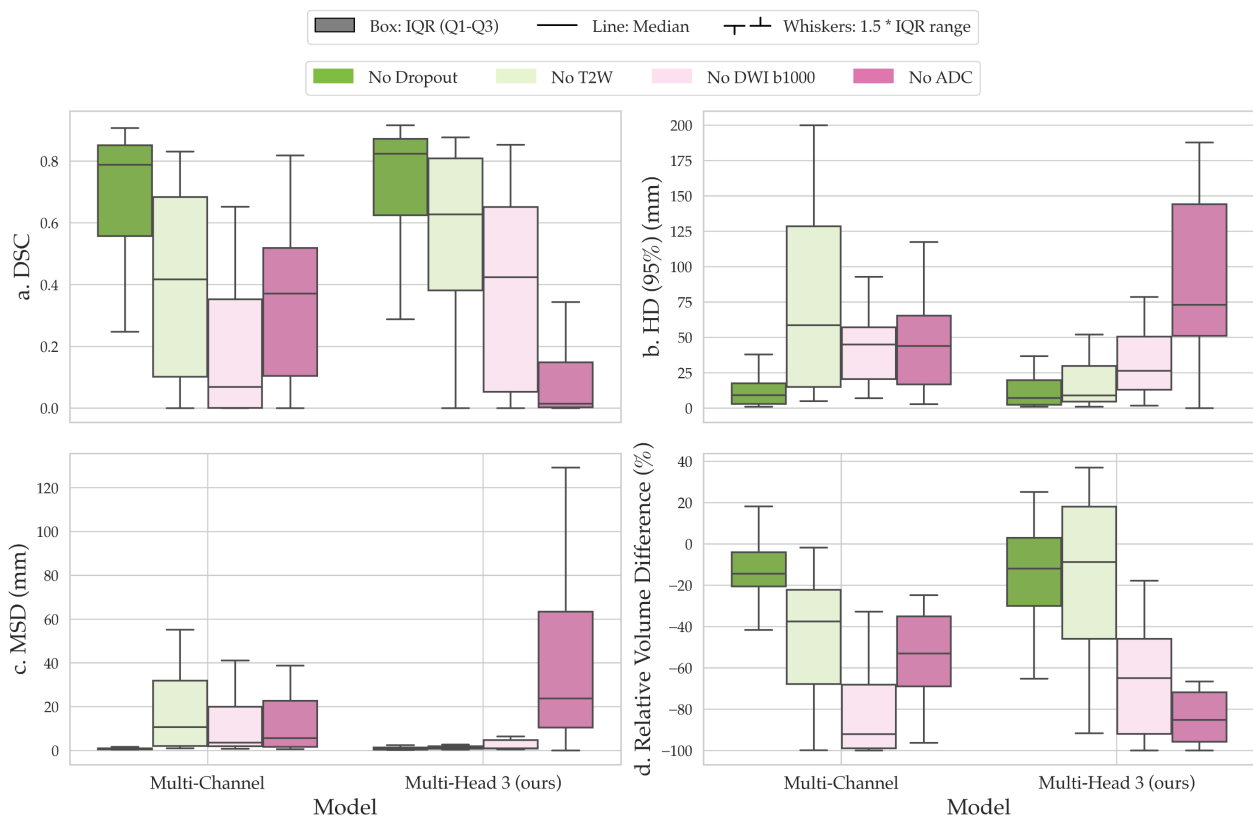
**Figure 3.** Comparison of segmentation performance based on (a) Dice similarity coefficient (DSC), (b) 95th percentile Hausdorff distance (HD), (c) mean surface distance (MSD), and (d) relative percentage volume difference between the baseline multi-channel model and three variations of the dilated multi-head model. Overall, the multi-head model with dilated convolutions and separate b1000 DWI/ADC feature aggregation (multi-head model 3) achieved the best performance among all models. IQR: interquartile range.



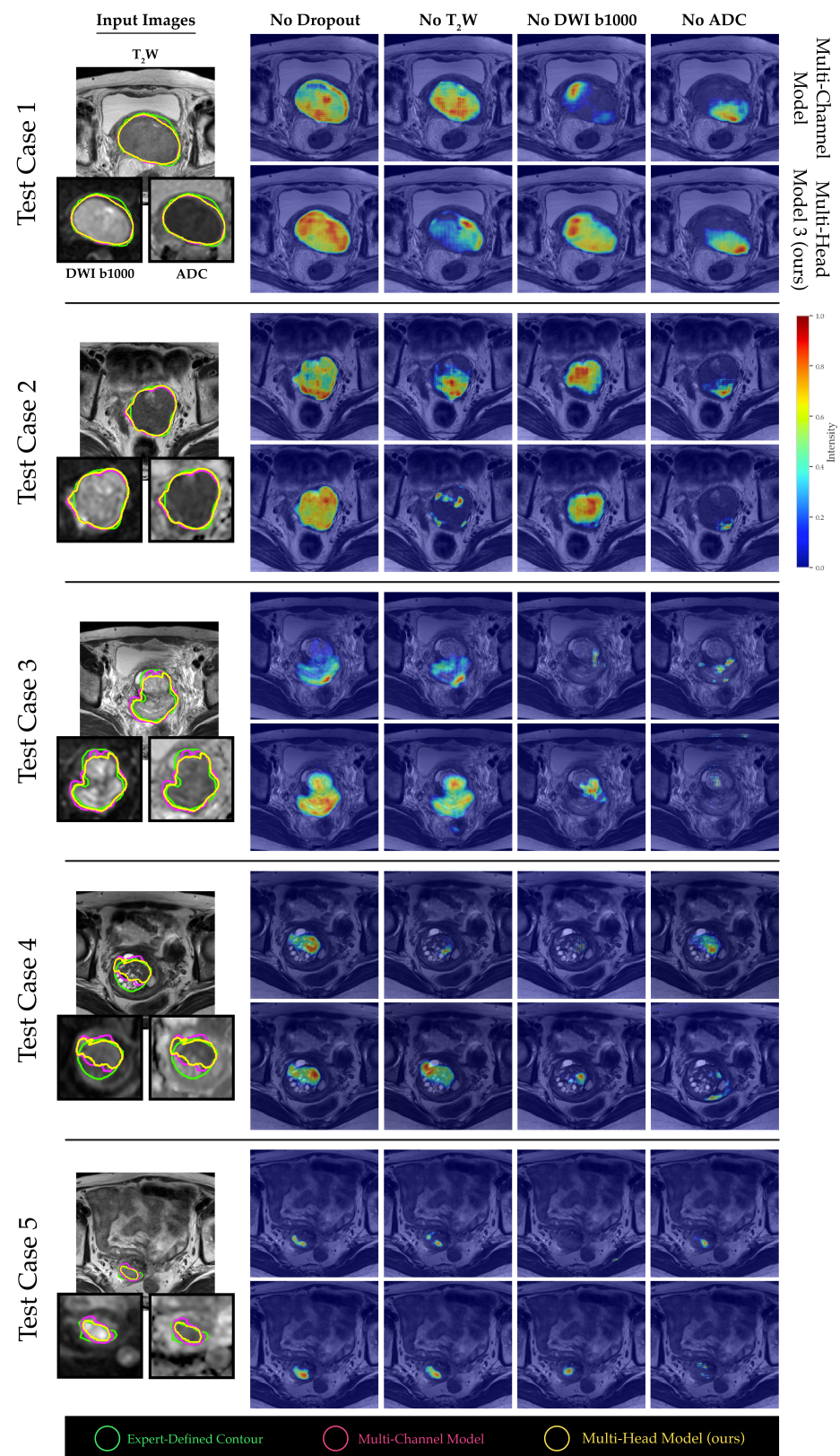


**Figure 4.** Comparative evaluation of the baseline multi-channel and proposed multi-head models for 6 example test cases, arranged in descending order of tumor size. The yellow arrows indicate regions where distortion in DWI and subjectivity in tumor location made contour propagation from T<sub>2</sub>W images challenging. Conversely, the white arrows highlight regions where the proposed model performed better by capturing boundaries and contextual information compared to the baseline multi-channel model. The red arrows highlight tumor regions within the ground-truth contours that did not exhibit strong impeded diffusion, thus remaining undetected by the DL models.

A channel sensitivity analysis was performed to compare the performance metrics between the multi-channel and our proposed multi-head model. One finding was the significant dependence of the multi-head model on the ADC channel. Specifically, when the ADC channel was excluded from the input, the tumor segmentation performance of the multi-head model deteriorated markedly, as evidenced in Figure 5. The saliency maps generated from both models revealed that b1000 DWI images had a relatively minor impact on the overall tumor segmentation performance for larger tumor volumes. In contrast, the T<sub>2</sub>W and ADC images were more crucial, as visualized in Figure 6, test cases 1 and 2. However, for smaller and more challenging tumor masses, the absence of ADC channels had a more pronounced adverse effect on the final outcome (see Figure 6, test cases 3–5).



**Figure 5.** Comparison of the sensitivity of the multi-channel and proposed multi-head models to different input channels based on quantitative metrics, including the (a) Dice similarity coefficient (DSC), (b) 95th percentile Hausdorff distance (HD), (c) mean surface distance (MSD), and (d) relative percentage volume difference. Our channel dropout analysis revealed a strong dependence of the proposed multi-head model on the ADC channel, indicating its importance in achieving accurate segmentation performance. IQR: interquartile range.



**Figure 6.** The normalized 3D GRAD-CAM saliency maps generated from the penultimate layer of the models, highlighting the most relevant regions for segmentation under different channel dropout conditions at test time. To improve visualization, the b1000 DWI and ADC images for smaller tumors were scaled.

#### 4. Discussion

DWI is a critical functional imaging technique for the detection and localization of tumors. This is particularly evident in ADC maps, where regions depicting impeded water molecule diffusion are indicative of increased cell density, often signifying a more aggressive disease [44]. However, using DWI and ADC maps in conjunction with T<sub>2</sub>W MRI sequences presents several challenges, such as voxel misalignment due to distortion and soft-tissue deformations between scans, variations in tumor delineations between the two modalities and the absence of standardized protocols for mpMRI. In this study, we proposed a novel multi-head framework that utilizes both b1000 DWI (including ADC) and T<sub>2</sub>W MRI for cervical cancer segmentation—a process well-suited to potential biomarker extraction. Our framework includes separate encoding heads that extract contextual information about tumors using dilated (or atrous) convolutions and shared residual connections. We have demonstrated that our technique provides a more robust and boundary-aware segmentation of cervical tumors when compared with the baseline multi-channel architecture that is commonly used in previous studies [32,35]. Our findings using the multi-channel training approach are comparable to previous reports on the segmentation of cervical tumors in multiparametric MRI [29,32,35]. However, it is challenging to make direct comparisons of our results due to the lack of public databases and resources for cervical cancer segmentation in MRI [45] and the use of different datasets.

Dilated convolutions are a crucial component of several successful segmentation architectures, including DeepLabv3+ [46], DeepLab [47], residual enhanced U-Net and atrous convolution (AtResU-Net) [48], and 3D deeply supervised fully convolutional network with concatenated atrous convolution (3D DSA-FCN) [49]. While most of these techniques utilize dilated convolutions throughout their architectures' encoding steps, we propose a multi-head framework that uses these operations for separate contextual and representational learning in b1000 DWI and ADC images for only the first block of the encoder. This approach ensures that the training parameters are not increased drastically compared with baseline multi-channel architectures, and the model is easier to interpret with conventional methods for future explainability studies. Moreover, lighter networks are better suited to online MR-guided treatments, where segmentation and planning are performed live on the scan taken on the day before radiation treatment [50], and speed is of the essence. In this study, we trained our models using anisotropic sub-volumes to maintain a greater focus on the plane of acquisition (axial) for 2D MRI. Our presented methodology, while primarily focused on its current application, possesses a versatile architecture that can be adapted and extrapolated for broader utilities. For example, it can be extended to 3D MRI scans, which are more commonly used for radiation therapy.

Channel sensitivity and saliency mapping of our experimental model indicated that our algorithm was more sensitive to ADC maps, which potentially makes it more robust to changes in acquisition protocol in MRI scanners. This approach could also serve as a strategy for more generalizable and cross-disease detection models [51]. However, this dependence on b1000 DWI and ADC images, as demonstrated in this study, may result in underestimation of the predicted tumor volume for malignancies with heterogeneous tumor mass diffusion. Although our findings indicate a stronger dependence on ADC input channels, further investigations are required for a more comprehensive understanding of tumor segmentation outcome using MRI and DWI images from multiple centres. Moreover, the subjectivity associated with inter-operator variability presents another drawback, with more reliable segmentations only attainable through the use of consensus contours. The decision to include specific areas within the ground-truth contour of the tumor is a discretionary choice made by the clinical expert annotating the images—a decision that relies heavily on their professional training and experience. It is important to note that this limitation may bias the findings from this study. Hence, future studies should aim to employ consensus ground-truth contours and evaluate the segmentation outcome through a number of expert human reader assessments to ensure the accuracy and reliability of the results for clinical decision-making.

## 5. Conclusions

Our proposed multi-head framework that combines b1000 DWI, ADC, and T<sub>2</sub>W MRI for cervical cancer segmentation has demonstrated improved accuracy and robustness compared to conventional multi-channel architectures. The use of dilated convolutions in only the first block of the encoder improves contextual learning with no significant parameter increase compared with conventional U-Net models. However, both the dependence on b1000 DWI and ADC channels and inter-operator variability are potential limitations that need to be addressed in future studies. Potential solutions may include architectural improvements or the use of consensus ground-truth contours and expert human reader assessments. Overall, our approach could offer a more accurate solution for monitoring disease progression and treatment response. This has the potential to enhance clinical decision-making in the diagnosis and treatment of patients with cervical cancer, as well as other pelvic malignancies.

**Author Contributions:** Conceptualization, R.K., M.D.B. and D.-M.K.; Methodology, R.K. and M.D.B.; Writing—Original Draft Preparation, R.K.; Supervision, S.C., J.M.W., G.L., C.M., M.D.B. and D.-M.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study represents independent research funded by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London, and by the Royal Marsden Cancer Charity. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. Gigin Lin received research funding from the Ministry of Science and Technology Taiwan (MOST 110-2628-B-182A-018).

**Institutional Review Board Statement:** The Chang Gung Memorial Hospital Institutional Review Board approved this study, and informed consent was waived (IRB 202000609B0C501).

**Informed Consent Statement:** Informed consent was waived for this study.

**Data Availability Statement:** The data for this study are unavailable due to privacy or ethical restrictions.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study, the writing of the manuscript, or the decision to publish.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Singh, D.; Vignat, J.; Lorenzoni, V.; Eslahi, M.; Ginsburg, O.; Lauby-Secretan, B.; Arbyn, M.; Basu, P.; Bray, F.; Vaccarella, S. Global estimates of incidence and mortality of cervical cancer in 2020: A baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. *Lancet Glob. Health* **2023**, *11*, e197–e206. [[CrossRef](#)] [[PubMed](#)]
3. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [[CrossRef](#)] [[PubMed](#)]
4. Togashi, K.; Nishimura, K.; Sagoh, T.; Minami, S.; Noma, S.; Fujisawa, I.; Nakano, Y.; Konishi, J.; Ozasa, H.; Konishi, I. Carcinoma of the cervix: Staging with MR imaging. *Radiology* **1989**, *171*, 245–251. [[CrossRef](#)] [[PubMed](#)]
5. Green, J.A.; Kirwan, J.M.; Tierney, J.F.; Symonds, P.; Fresco, L.; Collingwood, M.; Williams, C.J. Survival and recurrence after concomitant chemotherapy and radiotherapy for cancer of the uterine cervix: A systematic review and meta-analysis. *Lancet* **2001**, *358*, 781–786. [[CrossRef](#)] [[PubMed](#)]
6. Pollard, J.M.; Wen, Z.; Sadagopan, R.; Wang, J.; Ibbott, G.S. The future of image-guided radiotherapy will be MR guided. *Br. J. Radiol.* **2017**, *90*, 20160667. [[CrossRef](#)] [[PubMed](#)]
7. Koh, D.M.; Thoeny, H.C. *Diffusion-Weighted MR Imaging: Applications in the Body*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010.
8. Yoshida, S.; Takahara, T.; Kwee, T.C.; Waseda, Y.; Kobayashi, S.; Fujii, Y. DWI as an imaging biomarker for bladder cancer. *Am. J. Roentgenol.* **2017**, *208*, 1218–1228. [[CrossRef](#)]
9. Tsuruta, C.; Hirata, K.; Kudo, K.; Masumori, N.; Hatakenaka, M. DWI-related texture analysis for prostate cancer: Differences in correlation with histological aggressiveness and data repeatability between peripheral and transition zones. *Eur. Radiol. Exp.* **2022**, *6*, 1–12. [[CrossRef](#)]
10. Higaki, T.; Nakamura, Y.; Tatsugami, F.; Kaichi, Y.; Akagi, M.; Akiyama, Y.; Baba, Y.; Iida, M.; Awai, K. Introduction to the technical aspects of computed diffusion-weighted imaging for radiologists. *Radiographics* **2018**, *38*, 1131–1144. [[CrossRef](#)]

11. Abd elsalam, S.M.; Mokhtar, O.; Adel, L.; Hassan, R.; Ibraheim, M.; Kamal, A. Impact of diffusion weighted magnetic resonance imaging in diagnosis of cervical cancer. *Egypt. J. Radiol. Nucl. Med.* **2020**, *51*, 23. [[CrossRef](#)]
12. Subak, L.E.; Hricak, H.; Powell, C.B.; Azizi, E.; Stern, J.L. Cervical carcinoma: Computed tomography and magnetic resonance imaging for preoperative staging. *Obstet. Gynecol.* **1995**, *86*, 43–50. [[CrossRef](#)]
13. Romesser, P.B.; Tyagi, N.; Crane, C.H. Magnetic resonance imaging-guided adaptive radiotherapy for colorectal liver metastases. *Cancers* **2021**, *13*, 1636. [[CrossRef](#)] [[PubMed](#)]
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241; Proceedings, Part III 18.
15. Zhang, J.; Jin, Y.; Xu, J.; Xu, X.; Zhang, Y. Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation. *arXiv* **2018**, arXiv:1812.00352.
16. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [[CrossRef](#)] [[PubMed](#)]
17. Rehman, M.U.; Cho, S.; Kim, J.H.; Chong, K.T. Bu-net: Brain tumor segmentation using modified u-net architecture. *Electronics* **2020**, *9*, 2203. [[CrossRef](#)]
18. Rehman, M.U.; Cho, S.; Kim, J.; Chong, K.T. Brainseg-net: Brain tumor mr image segmentation via enhanced encoder–decoder network. *Diagnostics* **2021**, *11*, 169. [[CrossRef](#)] [[PubMed](#)]
19. Mubashar, M.; Ali, H.; Grönlund, C.; Azmat, S. R2U++: A multiscale recurrent residual U-Net with dense skip connections for medical image segmentation. *Neural Comput. Appl.* **2022**, *34*, 17723–17739. [[CrossRef](#)]
20. Lu, H.; She, Y.; Tie, J.; Xu, S. Half-UNet: A simplified U-Net architecture for medical image segmentation. *Front. Neuroinform.* **2022**, *16*, 911679. [[CrossRef](#)]
21. Ryu, J.; Rehman, M.U.; Nizami, I.F.; Chong, K.T. SegR-Net: A deep learning framework with multi-scale feature fusion for robust retinal vessel segmentation. *Comput. Biol. Med.* **2023**, *163*, 107132. [[CrossRef](#)]
22. Rehman, M.U.; Ryu, J.; Nizami, I.F.; Chong, K.T. RAAGR2-Net: A brain tumor segmentation network using parallel processing of multiple spatial frames. *Comput. Biol. Med.* **2023**, *152*, 106426. [[CrossRef](#)]
23. Bougourzi, F.; Distanto, C.; Dornaika, F.; Taleb-Ahmed, A. PDAAtt-Unet: Pyramid dual-decoder attention Unet for Covid-19 infection segmentation from CT-scans. *Med. Image Anal.* **2023**, *86*, 102797. [[CrossRef](#)] [[PubMed](#)]
24. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
25. Wang, H.; Xie, S.; Lin, L.; Iwamoto, Y.; Han, X.H.; Chen, Y.W.; Tong, R. Mixed transformer u-net for medical image segmentation. In *Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Virtual, 7–13 May 2022*; pp. 2390–2394.
26. Huang, L.; Zhu, E.; Chen, L.; Wang, Z.; Chai, S.; Zhang, B. A transformer-based generative adversarial network for brain tumor segmentation. *Front. Neurosci.* **2022**, *16*, 1054948. [[CrossRef](#)] [[PubMed](#)]
27. Rigaud, B.; Anderson, B.M.; Zhiqian, H.Y.; Gobeli, M.; Cazoulat, G.; Söderberg, J.; Samuelsson, E.; Lidberg, D.; Ward, C.; Taku, N.; et al. Automatic segmentation using deep learning to enable online dose optimization during adaptive radiation therapy of cervical cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2021**, *109*, 1096–1110. [[CrossRef](#)] [[PubMed](#)]
28. Ma, C.Y.; Zhou, J.Y.; Xu, X.T.; Guo, J.; Han, M.F.; Gao, Y.Z.; Du, H.; Stahl, J.N.; Maltz, J.S. Deep learning-based auto-segmentation of clinical target volumes for radiotherapy treatment of cervical cancer. *J. Appl. Clin. Med. Phys.* **2022**, *23*, e13470. [[CrossRef](#)] [[PubMed](#)]
29. Wang, B.; Zhang, Y.; Wu, C.; Wang, F. Multimodal MRI analysis of cervical cancer on the basis of artificial intelligence algorithm. *Contrast Media Mol. Imaging* **2021**, *2021*, 1673490. [[CrossRef](#)] [[PubMed](#)]
30. Lu, P.; Fang, F.; Zhang, H.; Ling, L.; Hua, K. AugMS-Net: Augmented multiscale network for small cervical tumor segmentation from MRI volumes. *Comput. Biol. Med.* **2022**, *141*, 104774. [[CrossRef](#)]
31. Zabihollahy, F.; Viswanathan, A.N.; Schmidt, E.J.; Lee, J. Fully automated segmentation of clinical target volume in cervical cancer from magnetic resonance imaging with convolutional neural network. *J. Appl. Clin. Med. Phys.* **2022**, *23*, e13725. [[CrossRef](#)]
32. Lin, Y.C.; Lin, C.H.; Lu, H.Y.; Chiang, H.J.; Wang, H.K.; Huang, Y.T.; Ng, S.H.; Hong, J.H.; Yen, T.C.; Lai, C.H.; et al. Deep learning for fully automated tumor segmentation and extraction of magnetic resonance radiomics features in cervical cancer. *Eur. Radiol.* **2020**, *30*, 1297–1305. [[CrossRef](#)]
33. Kano, Y.; Ikushima, H.; Sasaki, M.; Haga, A. Automatic contour segmentation of cervical cancer using artificial intelligence. *J. Radiat. Res.* **2021**, *62*, 934–944. [[CrossRef](#)]
34. Yoganathan, S.; Paul, S.N.; Paloor, S.; Torfeh, T.; Chandramouli, S.H.; Hammoud, R.; Al-Hammadi, N. Automatic segmentation of magnetic resonance images for high-dose-rate cervical cancer brachytherapy using deep learning. *Med. Phys.* **2022**, *49*, 1571–1584. [[CrossRef](#)] [[PubMed](#)]
35. Hodneland, E.; Kaliyugarasan, S.; Wagner-Larsen, K.S.; Lura, N.; Andersen, E.; Bartsch, H.; Smit, N.; Halle, M.K.; Krakstad, C.; Lundervold, A.S.; et al. Fully Automatic Whole-Volume Tumor Segmentation in Cervical Cancer. *Cancers* **2022**, *14*, 2372. [[CrossRef](#)] [[PubMed](#)]
36. Yoshizako, T.; Yoshida, R.; Asou, H.; Nakamura, M.; Kitagaki, H. Comparison between turbo spin-echo and echo planar diffusion-weighted imaging of the female pelvis with 3T MRI. *Acta Radiol. Open* **2021**, *10*, 2058460121994737. [[CrossRef](#)] [[PubMed](#)]

37. Donato, F., Jr.; Costa, D.N.; Yuan, Q.; Rofsky, N.M.; Lenkinski, R.E.; Pedrosa, I. Geometric distortion in diffusion-weighted MR imaging of the prostate—Contributing factors and strategies for improvement. *Acad. Radiol.* **2014**, *21*, 817–823. [[CrossRef](#)] [[PubMed](#)]
38. Kurman, R.J.; Carcangiu, M.L.; Herrington, C.S. *World Health Organisation Classification of Tumours of the Female Reproductive Organs*; International Agency for Research on Cancer: Lyon, France, 2014.
39. Saleh, M.; Virarkar, M.; Javadi, S.; Elsherif, S.B.; de Castro Faria, S.; Bhosale, P. Cervical cancer: 2018 revised international federation of gynecology and obstetrics staging system and the role of imaging. *Am. J. Roentgenol.* **2020**, *214*, 1182–1195. [[CrossRef](#)] [[PubMed](#)]
40. Tversky, A. Features of similarity. *Psychol. Rev.* **1977**, *84*, 327. [[CrossRef](#)]
41. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [[CrossRef](#)]
42. Cardoso, M.J.; Li, W.; Brown, R.; Ma, N.; Kerfoot, E.; Wang, Y.; Murrey, B.; Myronenko, A.; Zhao, C.; Yang, D.; et al. MONAI: An open-source framework for deep learning in healthcare. *arXiv* **2022**, arXiv:2211.02701.
43. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
44. McVeigh, P.Z.; Syed, A.M.; Milosevic, M.; Fyles, A.; Haider, M.A. Diffusion-weighted MRI in cervical cancer. *Eur. Radiol.* **2008**, *18*, 1058–1064. [[CrossRef](#)]
45. Kalantar, R.; Lin, G.; Winfield, J.M.; Messiou, C.; Lalondrelle, S.; Blackledge, M.D.; Koh, D.M. Automatic segmentation of pelvic cancers using deep learning: State-of-the-art approaches and challenges. *Diagnostics* **2021**, *11*, 1964. [[CrossRef](#)]
46. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), 8–14 September 2018; pp. 801–818.
47. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)]
48. Shah, N.A.; Gupta, D.; Lodaya, R.; Baid, U.; Talbar, S. Colorectal cancer segmentation using atrous convolution and residual enhanced unet. In *Proceedings of the Computer Vision and Image Processing: 5th International Conference, CVIP 2020, Prayagraj, India, 4–6 December 2020*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 451–462; Revised Selected Papers, Part I 5.
49. Wang, B.; Lei, Y.; Jeong, J.J.; Wang, T.; Liu, Y.; Tian, S.; Patel, P.; Jiang, X.; Jani, A.B.; Mao, H.; et al. Automatic MRI prostate segmentation using 3D deeply supervised FCN with concatenated atrous convolution. In Proceedings of the Medical Imaging 2019: Computer-Aided Diagnosis, San Diego, CA, USA, 16–21 February 2019; SPIE: Bellingham, WA, USA, 2019; Volume 10950, pp. 988–995.
50. Breto, A.L.; Spieler, B.; Zavala-Romero, O.; Alhuseini, M.; Patel, N.V.; Asher, D.A.; Xu, I.R.; Baikovitz, J.B.; Mellon, E.A.; Ford, J.C.; et al. Deep Learning for Per-Fraction Automatic Segmentation of Gross Tumor Volume (GTV) and Organs at Risk (OARs) in Adaptive Radiotherapy of Cervical Cancer. *Front. Oncol.* **2022**, *12*, 854349. [[CrossRef](#)]
51. Lin, Y.C.; Lin, Y.; Huang, Y.L.; Ho, C.Y.; Chiang, H.J.; Lu, H.Y.; Wang, C.C.; Wang, J.J.; Ng, S.H.; Lai, C.H.; et al. Generalizable transfer learning of automated tumor segmentation from cervical cancers toward a universal model for uterine malignancies in diffusion-weighted MRI. *Insights Imaging* **2023**, *14*, 14. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.