

# UNet-based Segmentation of Kidney and Kidney Tumor in 3D Computed Tomography Images

Yichi Zhang

## Overview

Kidney cancer affects over 400,000 individuals annually worldwide, with more than 150,000 deaths recorded in 2022 alone [2]. The primary treatment, such as partial nephrectomy, requires precise details about tumor size, shape, and location. Therefore, automated segmentation of kidneys and kidney tumors in CT images is crucial for accurate diagnosis and treatment planning. It reduces reliance on manual nephrometry scoring systems, which are time-consuming and prone to interobserver variability.

This study aimed to develop an effective model for segmenting kidneys and kidney tumors in 3D CT images. Using 210 training cases from the KiTS19 challenge, we trained three variants of the 3D UNet (plain, with residual connections, and with pre-activation residual connections). Extensive hyperparameter tuning was conducted to optimize data preprocessing and training procedures, followed by 5-fold cross-validation to compare different model architectures. The Residual 3D UNet was selected for its superior and consistent predictive capability in segmenting kidney tumors. Using this model, predictions were generated for 90 KiTS19 test cases, achieving a composite Dice score of 0.7569 in the challenge.

## Basic Setup

- Data preprocessing: Each 3D volume in the KiTS19 dataset is adjusted to a voxel spacing of  $3.22 \times 1.62 \times 1.62$  mm and standardized for intensity ranges. The volumes are then converted into a set of  $80 \times 160 \times 160$  patches.
- Model architecture: Plain 3d UNet (Figure1) residual and pre-activation residual 3D UNet (Figure2). In the Residual 3D UNet, residual block is conv-instnorm-ReLU-conv-instnorm-ReLU. In the pre-activation Residual 3D UNet, the order becomes instnorm-ReLU-convinstnorm-ReLU-conv.
- data augmentations: Spatial transformations including random rotations and scaling, colour adjustments such as random brightness and contrast adjustments, and Gaussian noise addition.
- Model training: The loss function combines weighted cross-entropy and Dice score. Adam optimizer is used for stochastic gradient descent (SDG).
- Inference method: Patch-based approach with overlapping sliding windows, averaging predicted probabilities for overlapped parts.

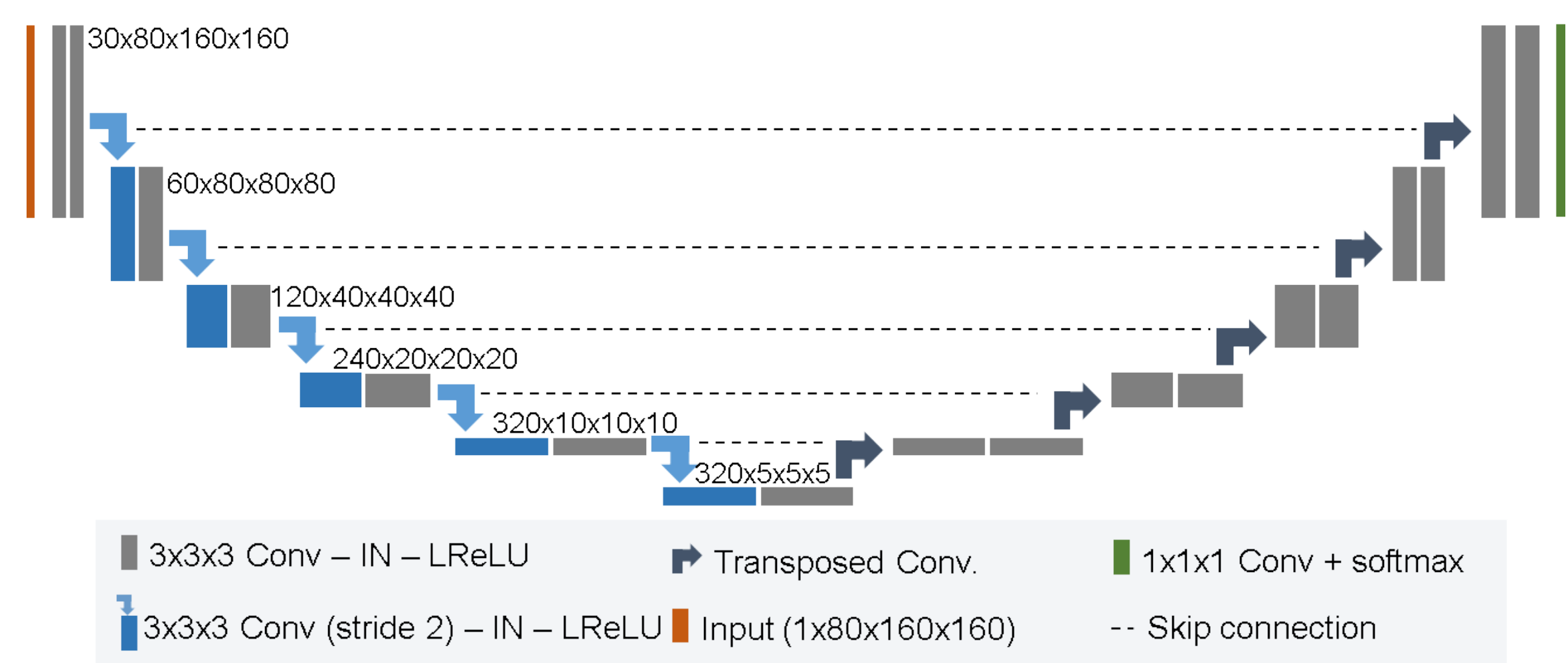


Figure 1: Plain 3D UNet [1]

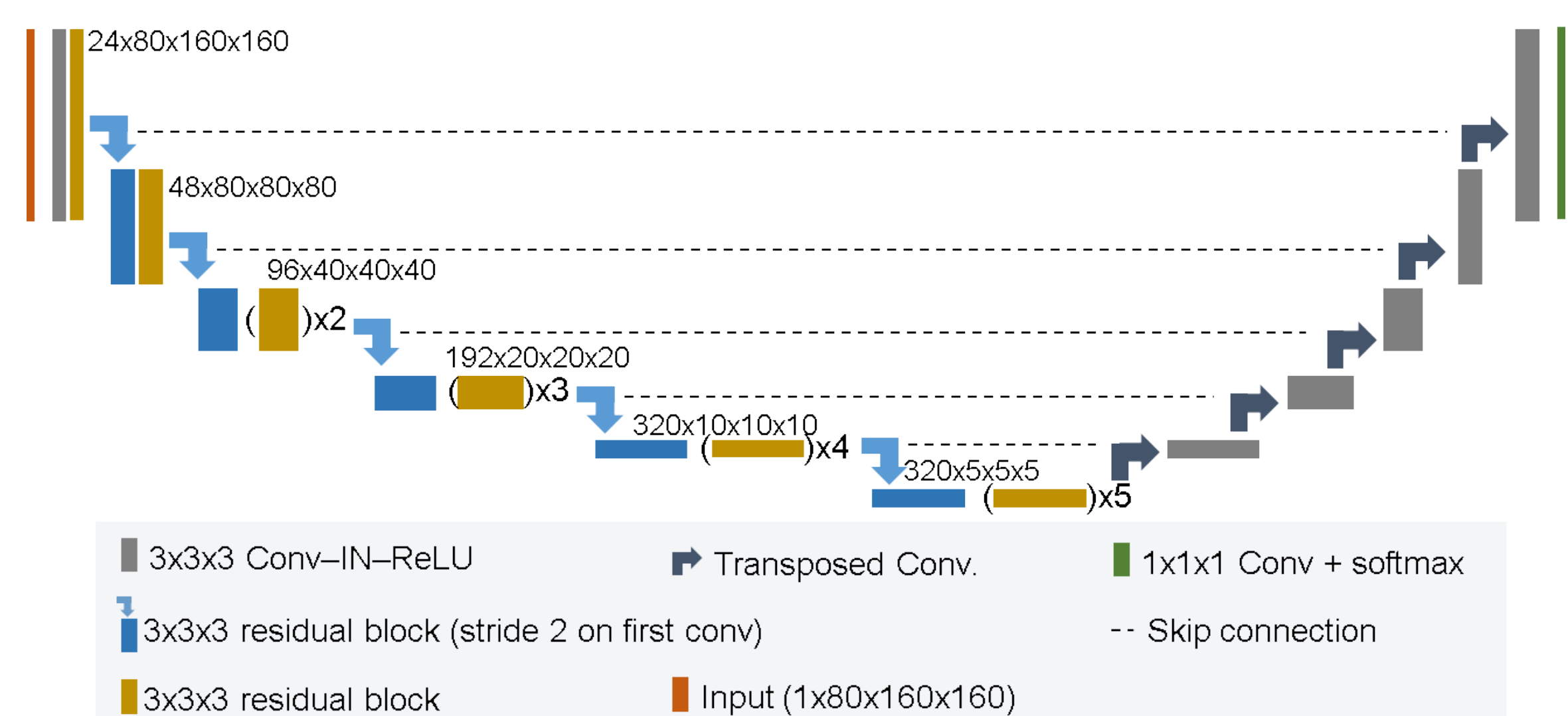


Figure 2: Residual 3D UNet [1]

## Hyper-parameter tuning

By comparing how the training process evolves with different choices made in data preprocessing and training, we opted for the ones that ensure training efficiency and generalization ability on the validation set. The optimal method for creating a batch in stochastic gradient descent (SGD) in our case is shown in Figure 3. A small batch size of 2 was chosen for its implicit regularization effect. This batch is formed by selecting 2 patches generated using overlapped sliding windows from 2 cases, applying 2 random data augmentations to each, and ultimately inputting 6 images into the model.

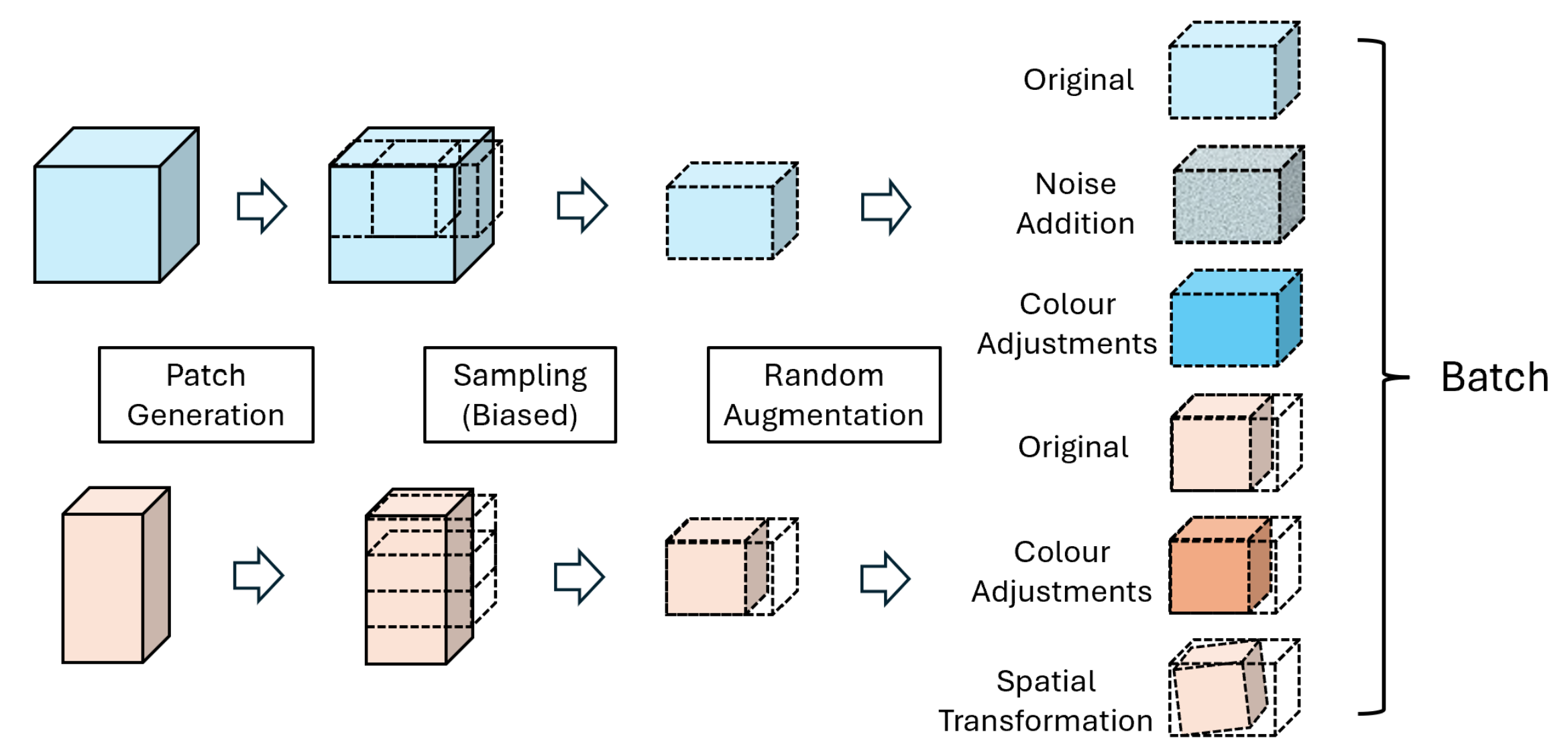


Figure 3: How a batch is created

It was also found learning rate reduction should occur judiciously, following a comprehensive exploration of the parameter space. Therefore, the learning rate is reduced by a factor of 0.2 only when the exponential moving average of the training loss does not show improvement over the last 30 epochs.

## Results

The means and standard deviations of the final validation tumor and kidney Dice scores from 5 folds after training three different 3D U-Net models for 600 epochs are shown below.

Table 1: Cross validation result		
Model	Tumour Dice	Kidney Dice
plain	0.4884 +- 0.0452	0.9339 +- 0.0100
residual	0.5490 +- 0.0195	0.9449 +- 0.0083
pre-activation	0.5269 +- 0.0376	0.9400 +- 0.0076

No further epochs were trained because in all folds, the tumor Dice scores ceased to improve after epoch 500, and there was a noticeable trend towards overfitting across all models. The ensemble of 5 residual 3D UNets was then created to generate predictions for 90 test cases. The scores being obtained using our generated predictions from the challenge are mean Kidney Dice 0.9360 and Tumor Dice 0.5778.

## Discussion

To enhance tumor segmentations and overall model performance, prioritizing reductions in false negatives (pixels for tumour misclassified as background) is crucial. Potential strategies include using Tversky loss and Focal Loss, supervisions at different resolutions, incorporating hard negative mining, training a dedicated model solely for tumour segmentation and refining the model architecture with components like attention layers.

## References

- [1] Fabian Isensee and Klaus H Maier-Hein. An attempt at beating the 3d u-net. *arXiv preprint arXiv:1908.02182*, 2019.
- [2] World Cancer Research Fund. Kidney cancer statistics, 2022.