MULTI-SCALE SPATIAL FEATURE FUSION IN 3D CONVOLUTIONAL ARCHITECTURES FOR LUNG TUMOR SEGMENTATION FROM 3D CT IMAGES

M.Sc. Thesis Defense Presentation

Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology, Dhaka.

05 July 2022

Student Name: Muhammad Suhail Najeeb
Student ID: 1018062222

Supervisor:
Dr. Mohammed Imamul Hassan Bhuiyan
Professor, Dept. of EEE, BUET

Outline of Presentation

- Introduction & Background
  - Introduction
  - Related Work
  - Motivation
  - Objectives
- Proposed Methodology
  - Dataset, Data preprocessing & Augmentation
  - Baseline Model Architectures
  - 2D Architectures & Spatial Feature Extraction
  - Proposed Architectures
  - Segmentation Mask Generation
- Results and Discussion
  - Evaluation Metrics
  - 2D Networks:
    - Parameter Selection
    - Performance
  - 3D Networks
    - Parameter Selection
    - Detection Performance
    - Segmentation Performance
  - Computational Overhead
  - Comparison with other models
  - Visual Analysis
- Conclusion & Future Scope

Introduction & Background

Introduction

- Lung Cancer is the 2nd-most common form of Cancer, but the most threatening in terms of mortality.
- Every year lung cancer costs millions of lives
  - 1.80 million deaths in 2020, which is 18% of all cancer related deaths [9]
- Most lung cancer patients are diagnosed at an advanced stage due to late onset of symptoms and a lack of screening programs.
- Early diagnosis using low-dose CT scans can result in a 20% reduction in mortality from lung cancer. [12]
- **Delineation of the lung tumor volume**
  - Usually performed manually by an expert radiologist
  - Difficult, time-consuming, and error-prone task.
- Automated segmentation of the lung tumor volume is extremely important for Automated diagnosis of lung cancer.

Related Work

- Traditional computer-aided approaches: Morphological operations, connected components analysis, image processing [24-25].
- More streamlined approaches involved multiple-step processing – pre-processing of radiological images, segmentation, feature extraction, radiomics analysis, detection using machine learning, etc. [26-32]
- Traditional techniques can detect the existence of tumor, but lung tumor segmentation is a more challenging task and requires more advanced techniques like Deep Learning. [31]
- Several deep learning approaches have been utilized for segmentation tasks – pixel-wise classifier [45], Fully Convolutional Network, Dilated Convolutional Neural Networks [46], Convolutional encoder-decoder networks like SegNet [3], UNet [4] etc.
- Ronneberger et al. [4] – UNet architecture which revolutionized the field of biomedical image segmentation. Several biomedical segmentation networks such as UNet++[47], ResUNet++ [48], MultiResUNet [5], DRINet [49] etc. improved upon UNet for different medical image segmentation tasks.

Related Work

Lung/Lung Tumor Segmentation:

- Skourt et al. [50] – UNet network for lung CT segmentation
- Jiang et al. [51] – Multiple resolution connected feature streams for automatic lung tumor segmentation from CT images.
- Anthimopoulus et al. [52] – Dilated fully convolutional neural network for semantic segmentation of Interstitial Lung Disease (ILD)

Approaches based on the LOTUS Benchmark [REF]:

- Hossain et al. [56] – Hybrid 3D Dilated Convolutional Neural Network
- Kamal et al. [6] – Recurrent-3D-DenseUNet
- Farheen et al. [57] – Deeply Supervised MultiResUNet

Motivation

- Developing a fast, accurate, and efficient system for volumetric segmentation of lung tumors for the automatic diagnosis of lung cancer.
- Lung Tumor Segmentation is a Volumetric Segmentation task where existing segmentation networks present several limitations:
  - **2D Networks:** UNet [4] and its variants [5][48-49]. Only consider one-slice at a time, therefore are not able to process spatial context along the missing dimension.
  - **3D networks:** 3D-UNet [58], V-Net [59], VoxResNet [60], etc.
    - Utilizing the full 3D volume is computationally expensive
    - Down-sampling/taking patches leads to a loss of spatial context

Motivation

- What if we can utilize both 2D & 3D Data?
- Some approaches in literature make use of this, for example –
  - H-DenseUNet by Li et al. [61]: Hybrid approach with 2D & 3D subnetwork followed by Hybrid Feature Fusion layer.
  - Gan et al. [62] : Hybrid approach for lung tumor segmentation
  - Mahmud et al. [63] – Joint optimization strategy: deep 2D network followed by a shallow 3D network for Covid-19 Lesion Segmentation
- These approaches have only utilized high-level features with a late-fusion strategy – which disregards inter-slice relation of the features. None of these approaches consider the early fusion of multi-scale spatial features.
- Incorporating multi-scale spatial features (from 2D convolutional encoders) at the early stages of 3D segmentation networks can promote the learning of inter-feature relations corresponding to inter-slice information and has the potential for improving the volumetric segmentation performance.

Objectives

- To investigate different state-of-the-art 2D encoder-decoder segmentation networks [4] [5] for lung tumor segmentation and develop methodology to extract spatial features at multiple scales from 2D encoders.
- To develop 3D segmentation networks by incorporating multi-scale spatial feature fusion and 3D convolutions with minimal computational overhead and propose three novel architectures for volumetric segmentation – SFF-3D-UNet, SFF-3D-MultiResUNet, and SFF-Recurrent-3D-DenseUNet.
- To study and compare the segmentation of our proposed architectures with several state-of-the-art segmentation networks on a publicly available dataset [7] employing quantitative analysis in terms of both 2D and 3D dice coefficients and visual analysis for lung tumor segmentation.
- To study and compare the performance of lung tumor detection for our proposed architectures with state-of-the-art networks on a publicly available dataset.

Proposed Methodology

Lung Tumor Segmentation from 3D CT Scans using Multi-Scale Spatial Feature Fusion

Proposed Methodology

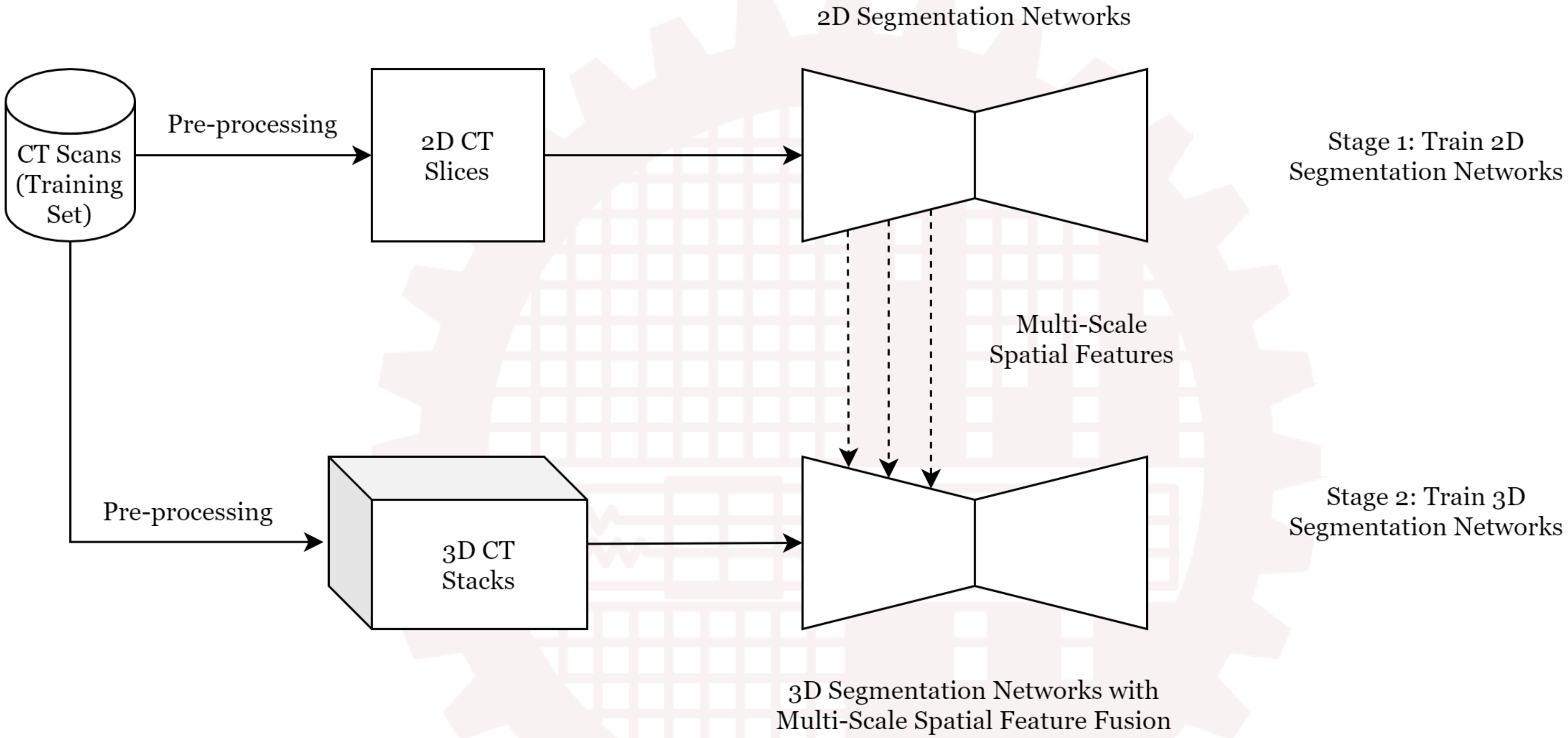


Figure 3: Brief overview of the proposed methodology

Dataset

- **LOTUS Benchmark (Lung-Originated Tumor Region Segmentation) [7]**
  - Prepared as part of the IEEE VIP Cup 2018 Challenge
  - Modified version of NSCLC-Radiomics Dataset
- Contains Computed Tomography (CT) scans of 300 lung cancer patients. CT scan resolution: 512 x 512
- Two different sources –
  - Siemens
  - CMS Imaging Inc.
- Annotations available for GTV, CTV, and PTV
- Segmentation Task: GTV (Gross-Tumor Volume)

		CT Scanner		**Number of Slices**	
Dataset	Patients	CMS Imaging Inc.	Siemens	Tumor	Non-Tumor
Train	260	60	200	4296 (13.7%)	26951 (86.3%)
Test	40	34	6	848 (18.9%)	3610 (81.1%)

Table 1: Dataset Statistics for the LOTUS Benchmark

Data Preprocessing

- Dataset provided in DICOM format
- PyDicom Library used to read DICOM scans
- Discrepancies present in the dataset associated with different CT scanners –
  - CMS Imaging Inc. : -1024 ~ 3071 HU
  - Siemens: 0 ~ 4095 HU
- HU Values were adjusted and normalized between 0 and 1
- The slices are resized using bilinear interpolation to a resolution of 256 x 256

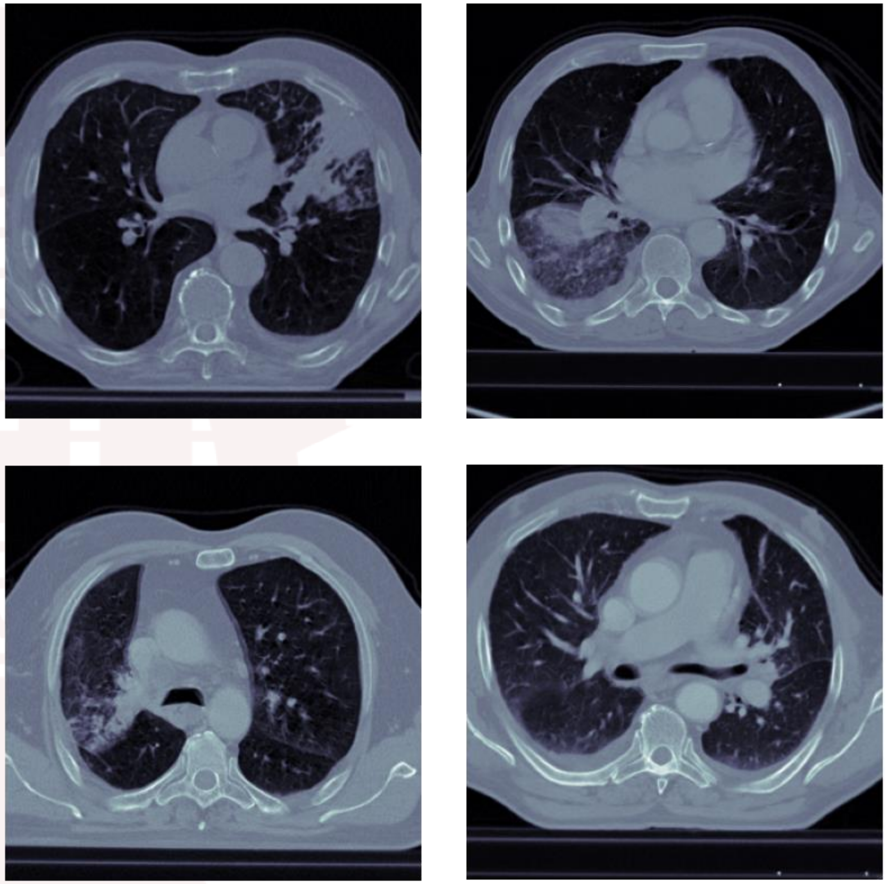


Figure 1: Sample Scans from the LOTUS Dataset

Data Augmentation

Data Augmentation is performed on-the-fly during training. One or more of the following data augmentations performed on each Training sample –

(a) Random Rotation
(b) Horizontal Flip
(c) Random Elastic Deformation
(d) Random Contrast Normalization
(e) Random Noise
(f) Blurring

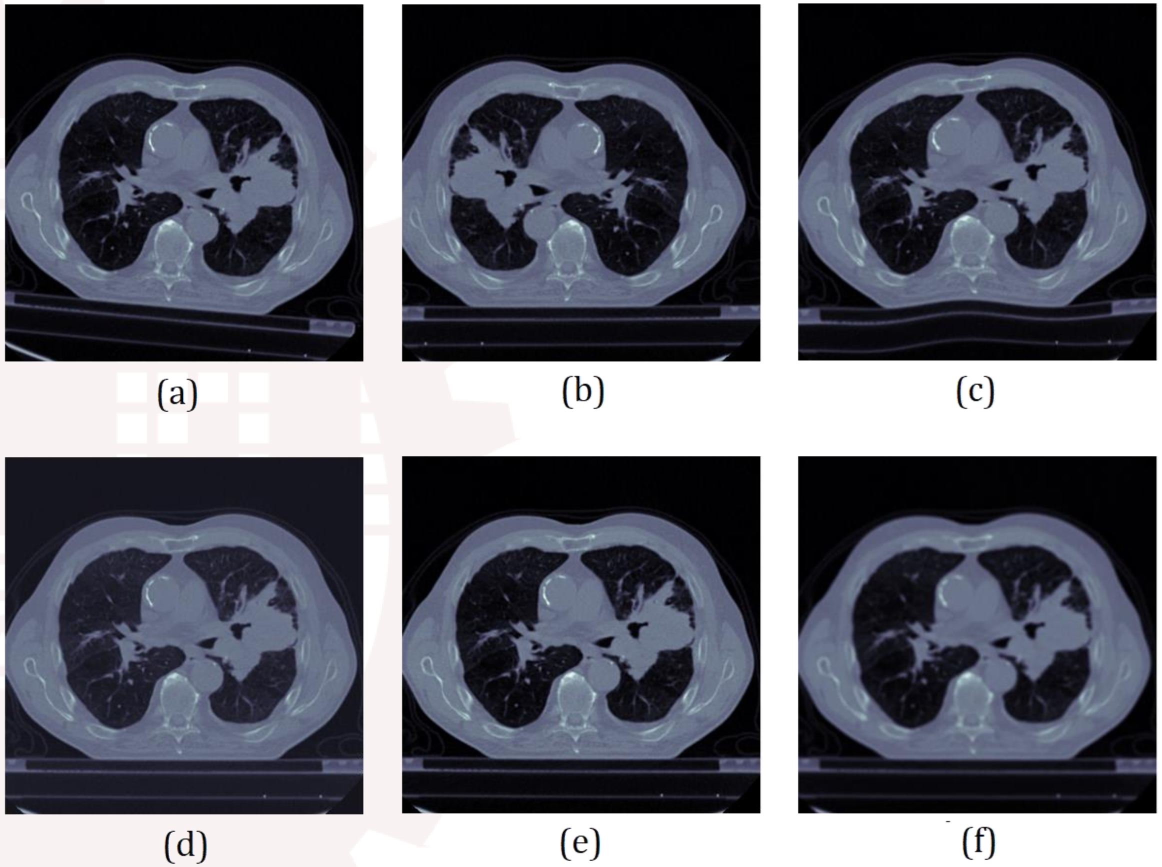


Figure 2: Illustration of different augmentations on a training sample

Baseline Model Architectures (1)

UNet Architecture: [REF]

- Convolutional Encoder-Decoder Segmentation Network (2D)
- Two paths: Contracting path/Expanding Path
- Successive 3x3 Convolution Operations followed by 2x2 Max Pooling in encoders
- 2x2 Upsampling Convolutions followed by Skip connections at decoders
- Four levels of encoder-decoders

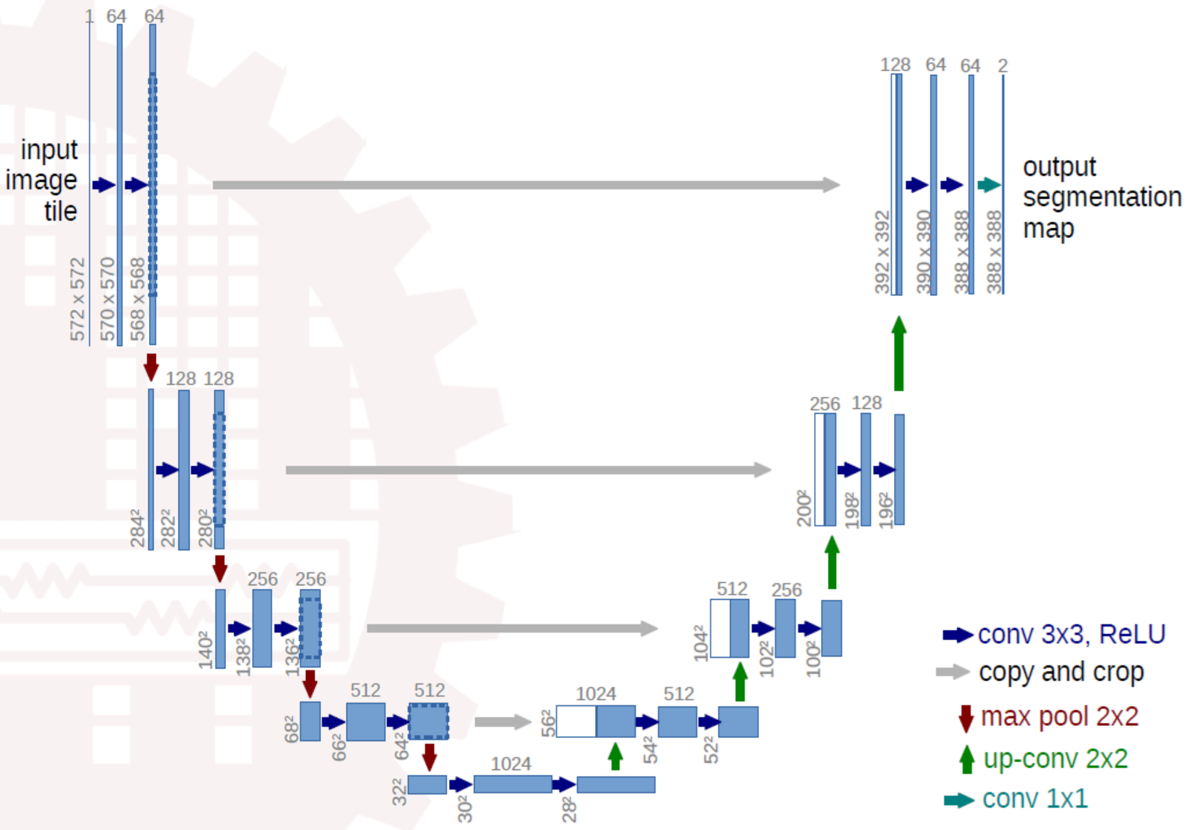


Figure 4: Architecture of the UNet

Baseline Model Architectures (2)

MultiResUNet [REF]

- MultiRes Block:
  - Modified encoder block with multiple 3x3 Convolution, joined by concatenation and a shortcut Connection. Helps the network deal with the variation of scale in medical images.
- Res Path:
  - Replaces shortcut connection with multiple convolution to bridge semantic gap between features

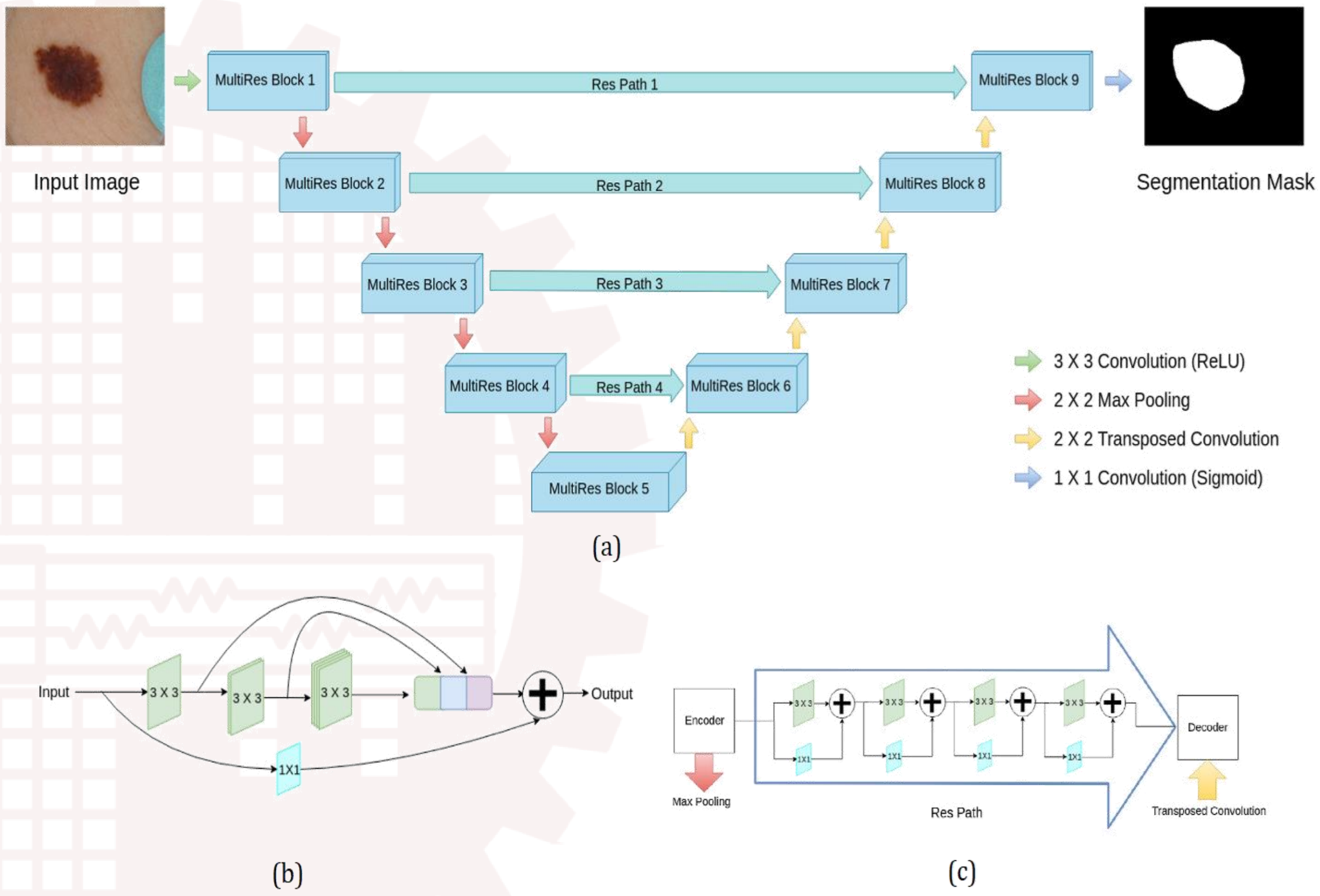


Figure 5: Architecture of the MultiResUNet

Baseline Model Architectures (3)

Recurrent-3D-DenseUNet [REF]

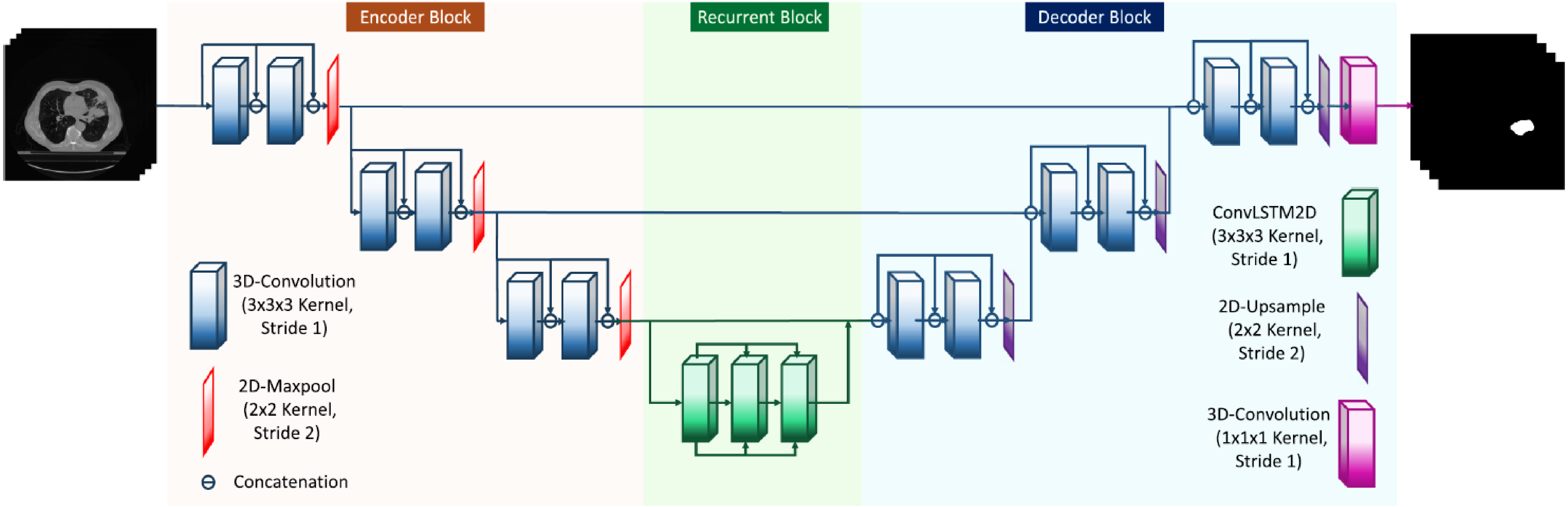


Figure 6: Architecture of the Recurrent-3D-DenseUNet

2D Architectures

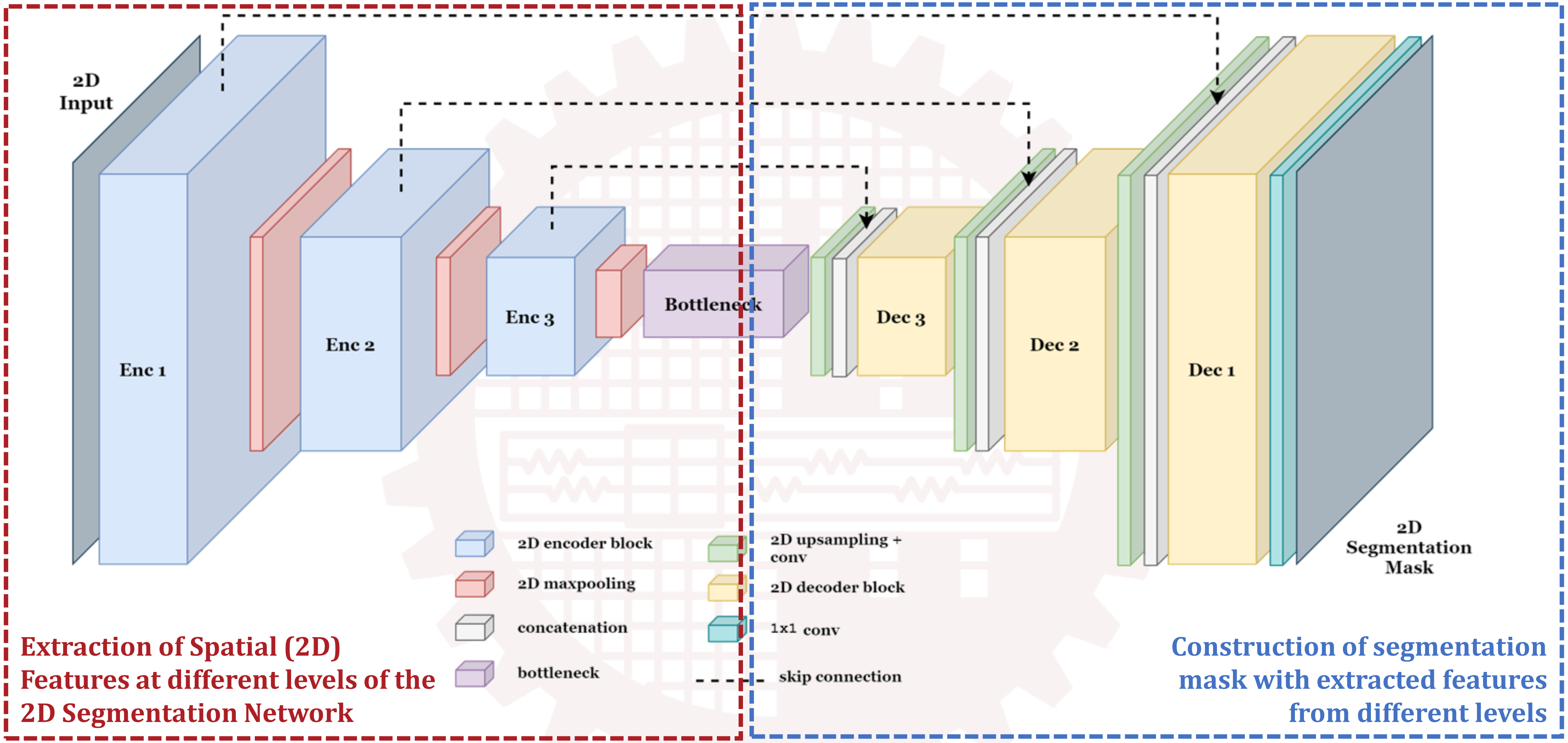
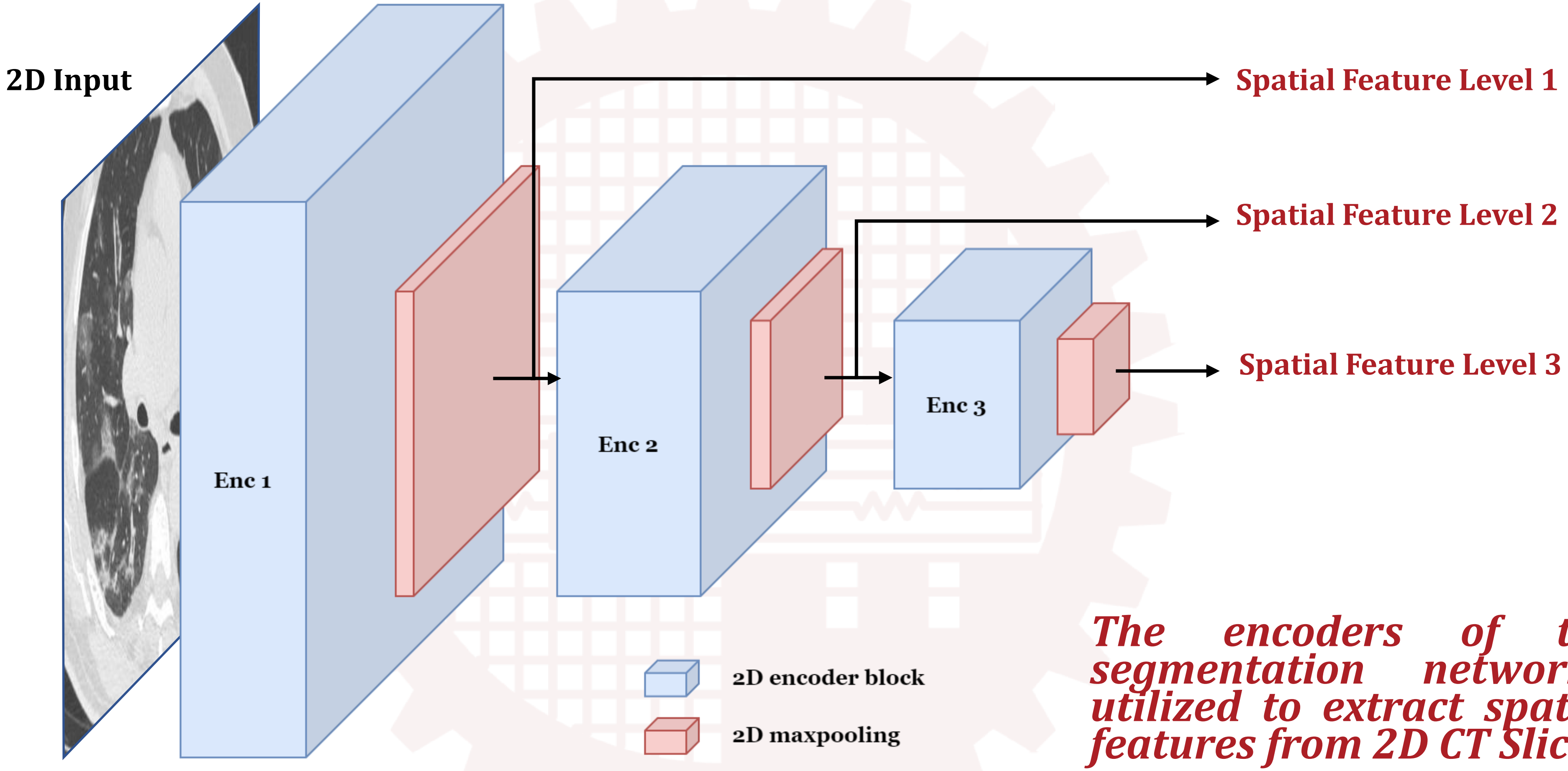


Figure 7: Basic Architecture of 2D Segmentation Networks

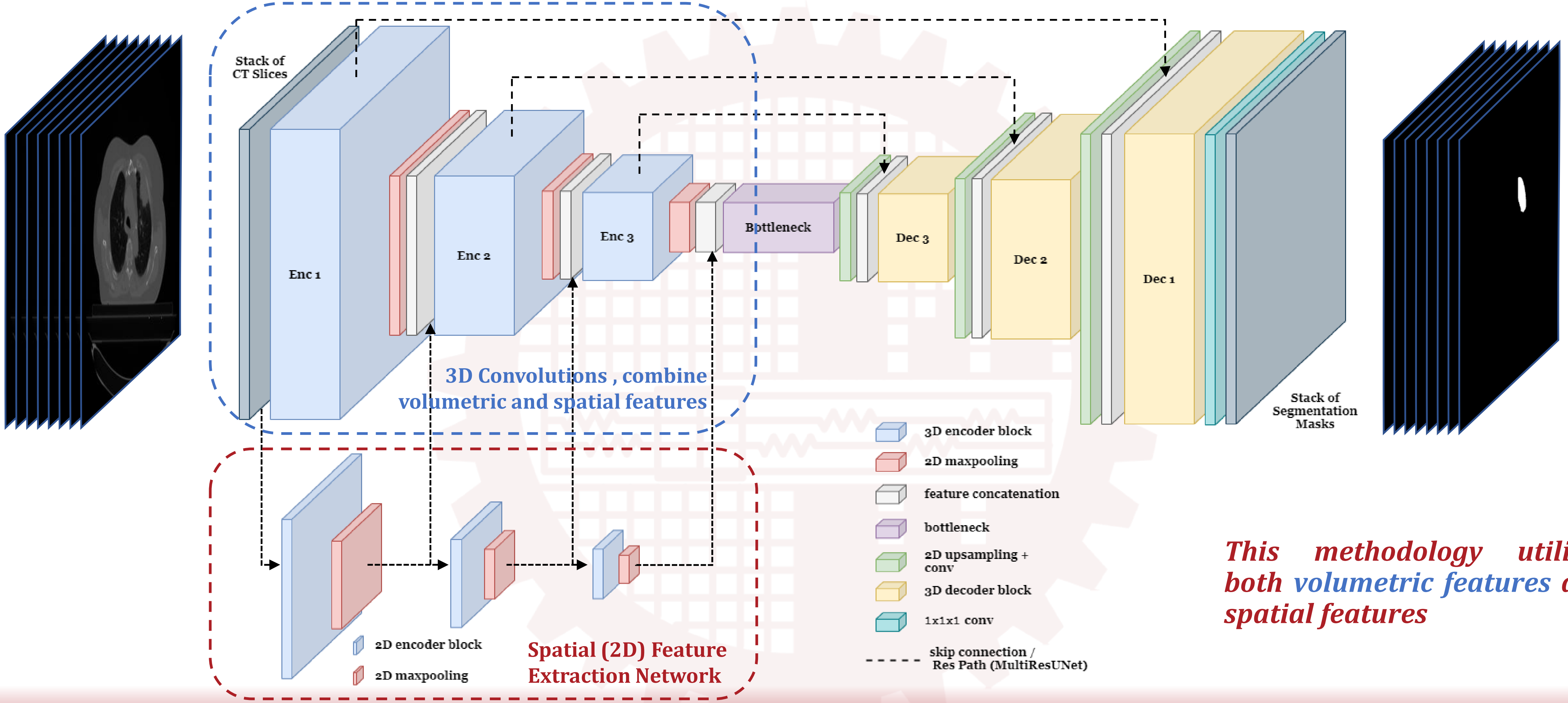
Multi-Scale Spatial Feature Extractor



The encoders of the 2D segmentation networks are utilized to extract spatial (2D) features from 2D CT Slices

Figure 8: Extraction of Multi-Scale Spatial Features

Proposed Methodology (Spatial Feature Fusion)



This methodology utilizes both volumetric features and spatial features

Figure 9: Brief Overview of Proposed Multi-Scale Spatial Feature Fusion

Proposed Architectures

- **SFF-3D-UNet**
  - Based on UNet [REF]
  - Replace 2D Convolutions with 3D Convolutions
  - Use 2D Maxpooling instead of 3D Maxpooling
  - Use 2D Upscaling instead of 3D Upscaling
  - Utilized pretrained 2D-UNet (3-level) for feature extraction
  - Incorporate Multi-Scale Spatial Feature Fusion
- SFF-3D-MultiResUNet
- SFF-Recurrent-3D-DenseUNet

SFF-3D-UNet Architecture

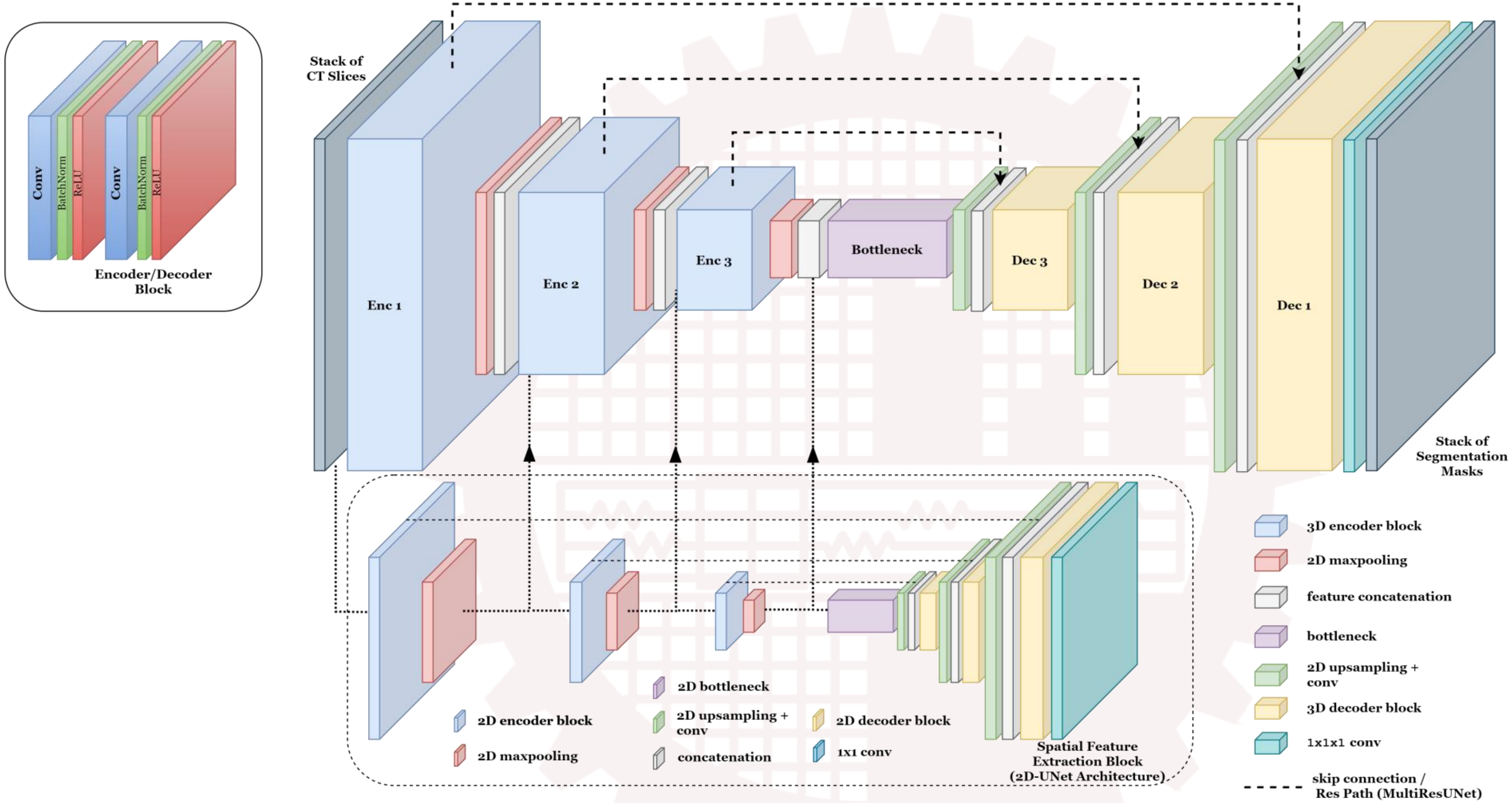


Figure 10: Architecture of Proposed SFF-3D-UNet

Proposed Architectures

- SFF-3D-UNet
- **SFF-3D-MultiResUNet**
  - Based on 2D-MultiResUNet [REF]
  - Replaced 2D MultiRes Block with 3D MultiRes Block (3D Convolutions instead of 2D Convolutions)
  - Replaced 2D ResPath with 3D ResPath (3D Convs)
  - Utilized 2D Maxpooling & 2D Upsampling
  - Utilized pretrained 2D-MultiResUNet (3-level) for feature extraction
  - Incorporate Multi-Scale Spatial Feature Fusion
- SFF-Recurrent-3D-DenseUNet

SFF-3D-MultiResUNet

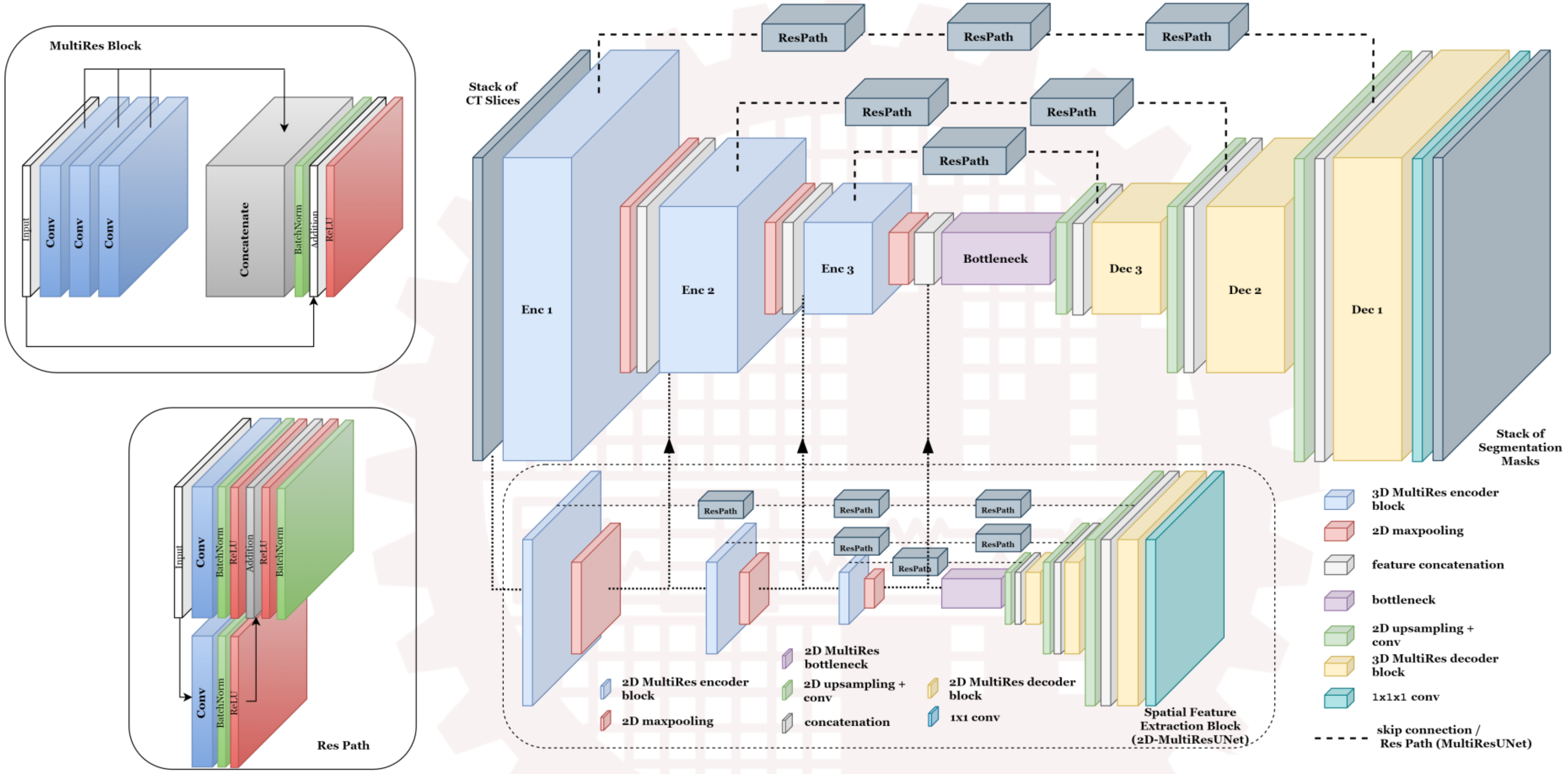


Figure 11: Architecture of Proposed SFF-3D-MultiResUNet

Proposed Architectures

- SFF-3D-UNet
- SFF-3D-MultiResUNet
- **SFF-Recurrent-3D-DenseUNet**
  - Based on Recurrent-3D-DenseUNet [REF]
  - Proposed 2D-DenseUNet architecture utilized for feature extraction
  - Incorporated Multi-Scale Spatial Feature Fusion

SFF-Recurrent-3D-DenseUNet

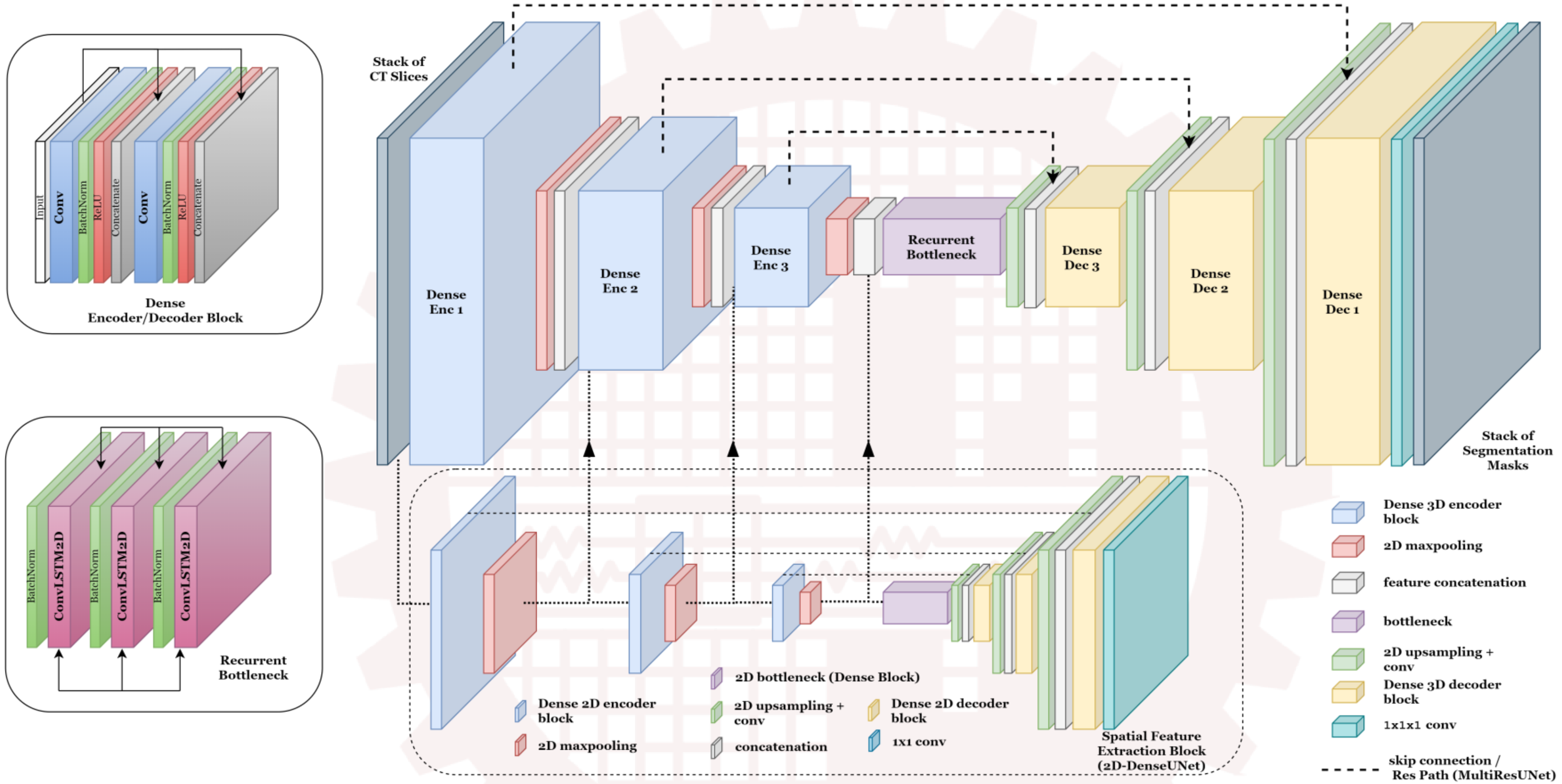


Figure 11: Architecture of Proposed SFF-Recurrent-3D-DenseUNet

Segmentation Mask Generation

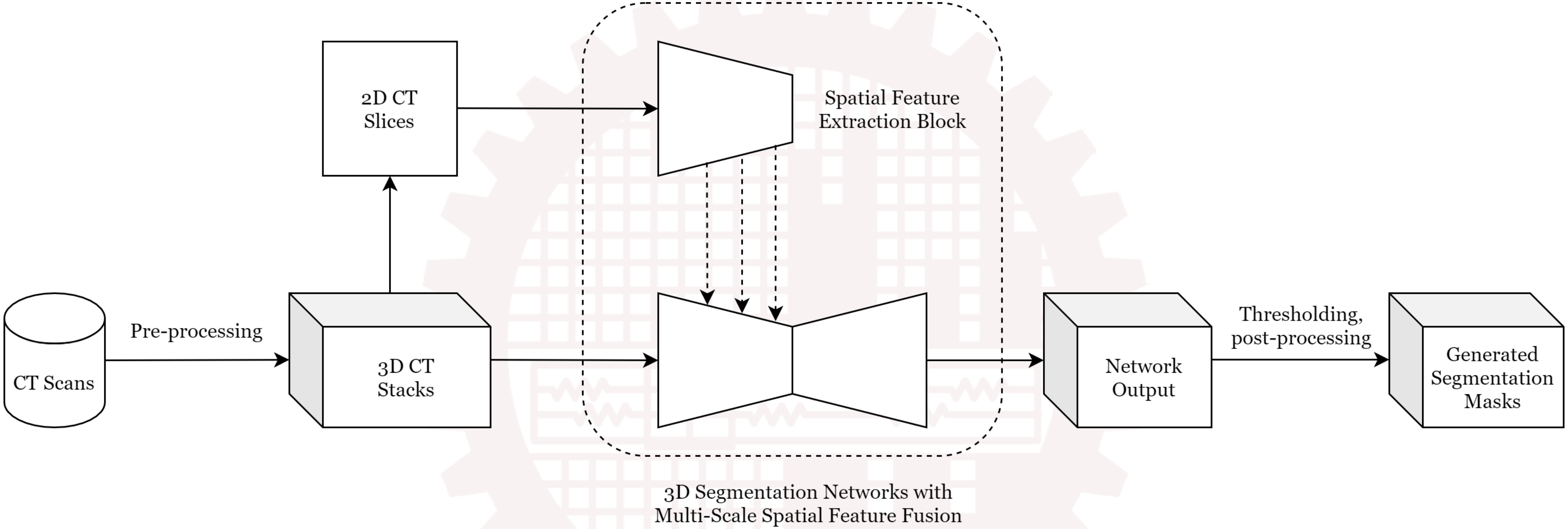


Figure 12: Brief outline of the segmentation mask generation process.

Segmentation Mask Generation

- 3D Segmentation networks: Produce results on 8 consecutive slices
- To produce final segmentation masks, overlapping stacks of CT slices are processed by the segmentation networks.
- Overlapping masks are averaged which serves as a post-processing step to remove noise.
- Segmentation mask values are within 0~1 where 0 signifies no tumor and 1 signifies tumor
- **Two-step thresholding approach is applied to generate final segmentation mask.**
  - **Step 1: Apply a threshold of 0.7 to filter out false-positive slices**
  - **Step 2: Apply a threshold of 0.5 to generate the final tumor volume**

Results & Discussion

Evaluation Metrics

Segmentation Performance:

- Dice Coefficient:
  - 2D Dice Score
  - 3D Dice Score

$$D = \frac{2 * |X \cap Y|}{|X| + |Y|}$$

- Overall Dice Coefficient (2D):
  - Dice coefficient calculated using above formula for True-Positive & True-Negative cases.
  - For True-Negatives (Model Successfully detects that no tumor is present), dice coefficient = 1
  - For False-Positives (Model mistakenly classifies tumor) dice coefficient = 0

- 3D Dice Coefficient: 3D Dice score of *predicted tumor volumes* with respect to the *tumor volumes present in the ground truth*.

Evaluation Metrics

Detection Performance:

- TP: True Positive FP: False Positive
- TN: True Negative FN: False Negative

- F1 Score:

$$F1_{score} = \frac{2 * TP}{2 * TP + FP + FN}$$

- MCC:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

2D Networks - Parameter Selection

Table 2: Dice coefficients (validation set) for different optimizers and learning rates of the 2D models

Model	Optimizer	Learning Rate	Dice Coefficient (Validation)
2D-UNet	SGD	0.1	0.5327
2D-UNet	SGD	0.01	0.4799
2D-UNet	Adam	0.01	0.5327
2D-UNet	Adam	0.001	0.5950
2D-MultiResUNet	Adam	0.01	0.6066
2D-MultiResUNet	Adam	0.001	0.5436
2D-DenseUNet	Adam	0.01	0.6088
2D-DenseUNet	Adam	0.001	0.5926

2D Networks - Performance

Table 3: Dice coefficients (test set) for the 2D models

Model	Opti- mizer	Learning Rate	Threshold	Dice Coefficient (Test Set)
2D-UNet	Adam	0.001	0.5	0.5886
			0.7	0.6510
2D-MultiResUNet	Adam	0.01	0.5	0.6706
			0.7	0.7158
2D-DenseUNet	Adam	0.01	0.5	0.6098
			0.7	0.6911

3D Networks – Parameter Selection

Table 4: Dice coefficients (validation set) for different learning rates for the 3D models

Model	Optimizer	Learning Rate	3D Dice (Validation)
3D-UNet	Adam	0.001	0.5942
3D-UNet	Adam	0.0001	0.5955
SFF-3D-UNet	Adam	0.001	0.6550
SFF-3D-UNet	Adam	0.0001	0.6568
3D-MultiResUNet	Adam	0.001	0.5615
SFF-3D-MultiResUNet	Adam	0.001	0.6175
Recurrent-3D-DenseUNet	Adam	0.0001	0.5979
SFF-Recurrent-3D-DenseUNet	Adam	0.001	0.6458

3D Networks – Segmentation

Table 5: Comparison of the dice coefficients for different models at different thresholds

Model	Threshold: 0.5		Threshold: 0.7		Two-step threshold	
	2D Dice	3D Dice	2D Dice	3D Dice	2D Dice	3D Dice
3D-UNet	0.7874	0.5460	0.8056	0.5102	0.8144	0.5440
SFF-3D-UNet	0.7914	0.5886	0.8178	0.5504	0.8275	0.5853
3D-MultiResUNet	0.8304	0.5844	0.8365	0.5368	0.8478	0.5803
SFF-3D-MultiResUNet	0.8437	0.5992	0.8555	0.5506	0.8669	0.5938
Recurrent-3D-DenseUNet	0.7777	0.5715	0.7984	0.5147	0.8080	0.5634
SFF-Recurrent-3D-DenseUNet	0.7916	0.5971	0.8143	0.5386	0.8276	0.5874

Significance of Thesholds

- Lower threshold (0.5) generates a more accurate delineation of 3D tumors but leads to more false-positives
- Higher threshold (0.7) reduces false-positives and improves the overall 2D Dice score. However, this reduces the volumetric segmentation performance (3D Dice score)
- The two-step thresholding approach offers a balance between false-positives and segmentation accuracy.
- **The two-step thresholding approach improves 2D dice score by 1.30% on average.**

Improvement in Segmentation

- In terms of 2D dice coefficient, the proposed models with SFF achieve performance improvements of –

2D Dice Score	Without SFF	With SFF	Improvement
3D-UNet	0.8144	0.8275	1.61%
3D-MultiResUNet	0.8478	0.8669	2.25%
Recurrent-3D-DenseUNet	0.8080	0.8276	2.42%

- In terms of 3D dice coefficient, the proposed models with SFF achieve performance improvements of -

3D Dice Score	Without SFF	With SFF	Improvement
3D-UNet	0.5440	0.5853	7.58%
3D-MultiResUNet	0.5803	0.5938	2.32%
Recurrent-3D-DenseUNet	0.5634	0.5874	4.28%

3D Networks - Detection

Table 6: Detection Performance (Test Set) of the different 3D models at different thresholds

Model	Thresh -old	TP	FP	TN	FN Score	F1	MCC
3D-UNet	0.5	603	488	3416	245	0.6219	0.5264
	0.7	549	367	3267	299	0.6224	0.5307
SFF-3D-UNet	0.5	639	520	3114	209	0.6367	0.5460
	0.7	583	358	3276	265	0.6517	0.5664
3D-MultiResUNet	0.5	611	319	3315	237	0.6872	0.6111
	0.7	546	241	3393	302	0.6678	0.5945
SFF-3D- MultiResUNet	0.5	633	283	3351	215	0.7176	0.6493
	0.7	577	179	3455	271	0.7194	0.6601
Recurrent-3D-DenseUNet	0.5	637	539	3095	211	0.6294	0.5367
	0.7	566	403	3231	282	0.6230	0.5295
SFF-Recurrent-3D-DenseUNet	0.5	664	526	3108	184	0.6516	0.5661
	0.7	606	365	3269	242	0.6663	0.5893

Improvement in Detection

- In terms of F1-Score, the proposed models with SFF achieve performance improvements of –

F1 Score	Without SFF	With SFF	Improvement
3D-UNet	0.6224	0.6517	4.71%
3D-MultiResUNet	0.6678	0.7194	7.73%
Recurrent-3D-DenseUNet	0.6230	0.6663	6.95%

- In terms of MCC, the proposed models with SFF achieve performance improvements of -

MCC	Without SFF	With SFF	Improvement
3D-UNet	0.5307	0.5664	6.73%
3D-MultiResUNet	0.5945	0.6601	11.03%
Recurrent-3D-DenseUNet	0.5295	0.5893	11.29%

Computational Overhead

Table 6: Comparison of different 3D Models in terms of computational overhead

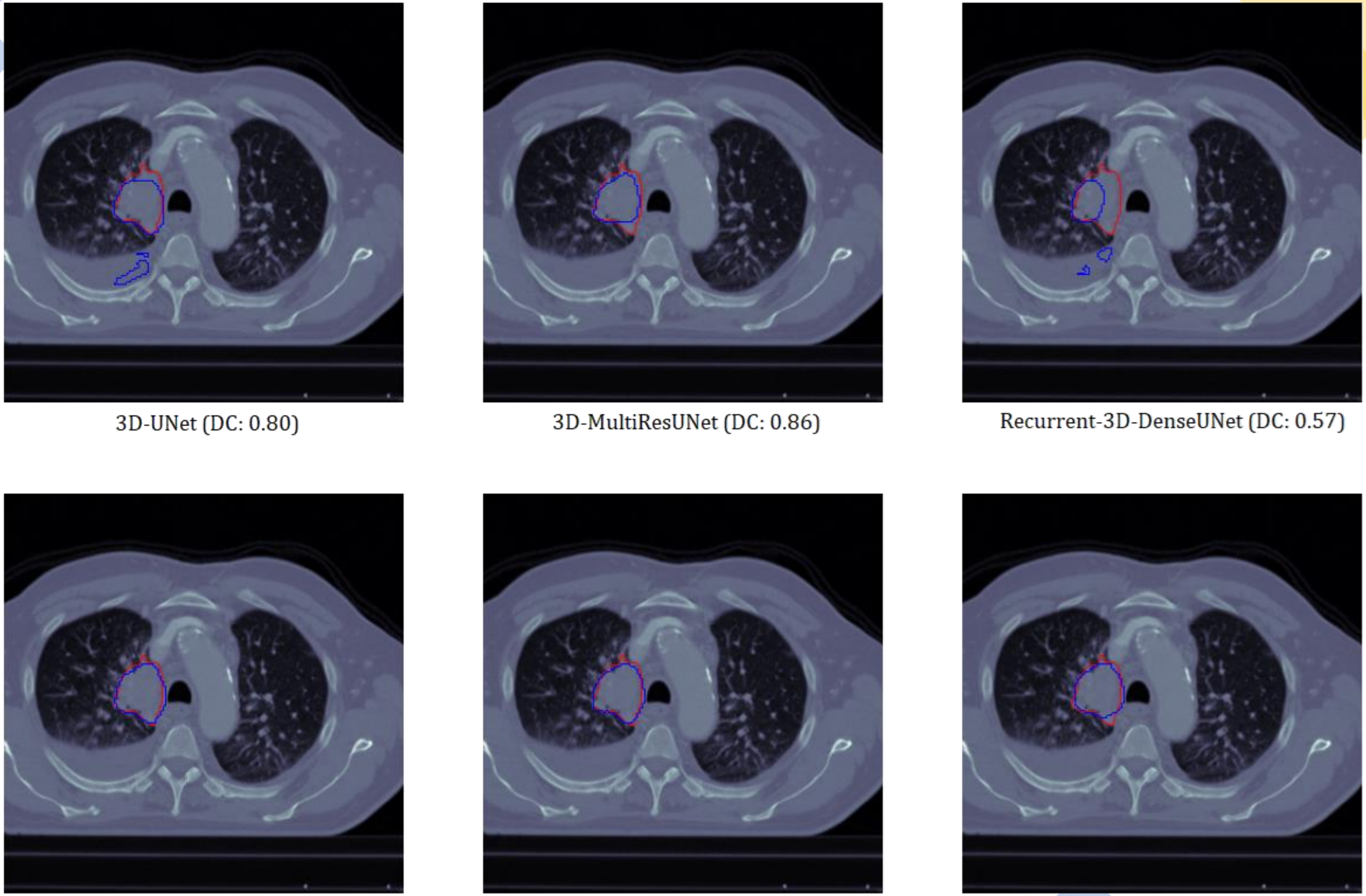
Model	Number of Parameters	Trainable Parameters	Epochs to Converge	Training time per epoch (min.)	Testing time (min.)	Dice Score (2D)
3D-UNet	5.433×10^6	5.430×10^6	29	13:43	4:02	0.8144
SFF-3D-UNet	6.882×10^6	6.591×10^6	20	14:18	4:10	0.8275
3D-MultiResUNet	4.297×10^6	4.285×10^6	30	22:18	6:27	0.8478
SFF-3D-MultiResUNet	5.135×10^6	4.932×10^6	17	22:48	7:02	0.8669
Recurrent-3D-DenseUNet	19.220×10^6	19.216×10^6	29	29:04	8:01	0.8080
SFF-Recurrent-3D-DenseUNet	25.012×10^6	24.551×10^6	28	32:55	9:47	0.8276

Comparison with Other models

Table 6: Comparison of different 3D Models in terms of computational overhead

Model	Mean Dice Coefficient (2D)	Number of parameters
2D-LungNet [56]	0.6267	1.30×10^5
3D-LungNet [56]	0.6577	4.03×10^5
3D-DenseNet [6]	0.6884	14×10^6
Recurrent-3D-DenseUNet [6]	0.7228	19.22×10^6
Deeply-Supervised-MultiResUNet [57]	0.8472	7.28×10^6
SFF-3D-UNet	0.8275	6.59×10^6
SFF-3D-MultiResUNet	0.8669	5.13×10^6
SFF-Recurrent-3D-DenseUNet	0.8276	25.01×10^6

Visual Analysis



3D-UNet (DC: 0.80)

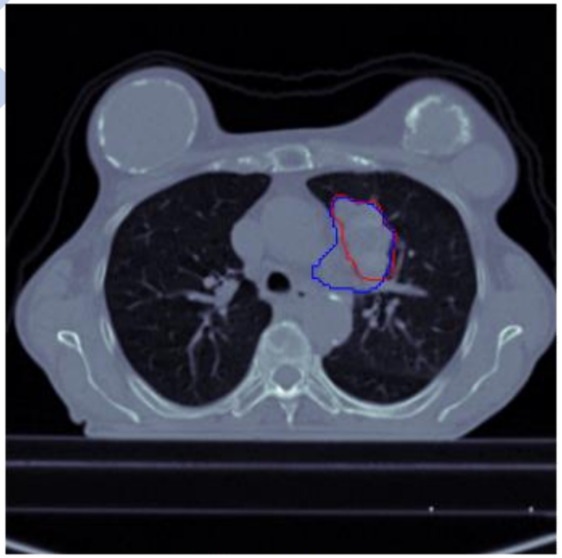
3D-MultiResUNet (DC: 0.86)

Recurrent-3D-DenseUNet (DC: 0.57)

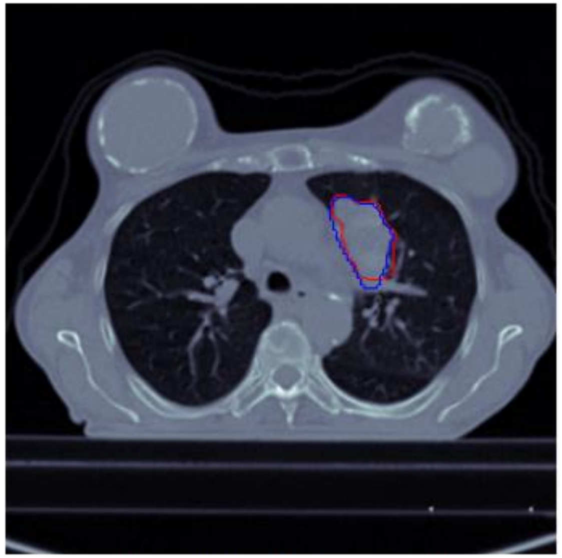
SFF-3D-UNet (DC: 0.92)

SFF-3D-MultiResUNet (DC: 0.92)

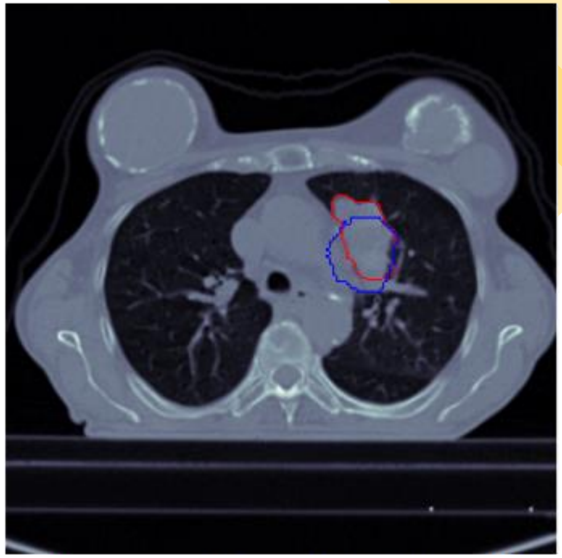
SFF-Recurrent-3D-DenseUNet (DC: 0.90)



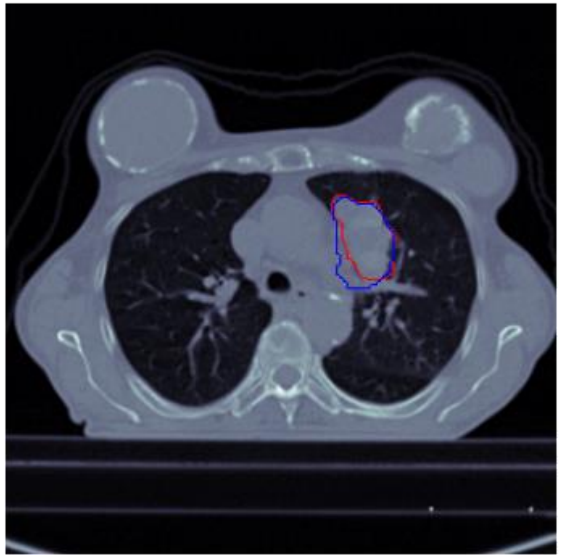
3D-UNet (DC: 0.76)



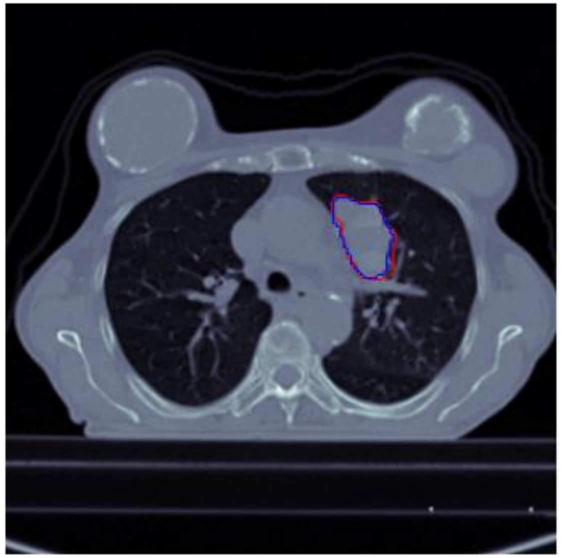
3D-MultiResUNet (DC: 0.89)



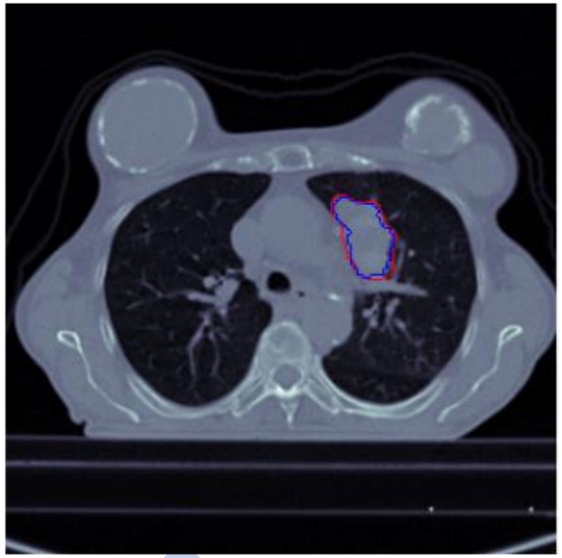
Recurrent-3D-DenseUNet (DC: 0.70)



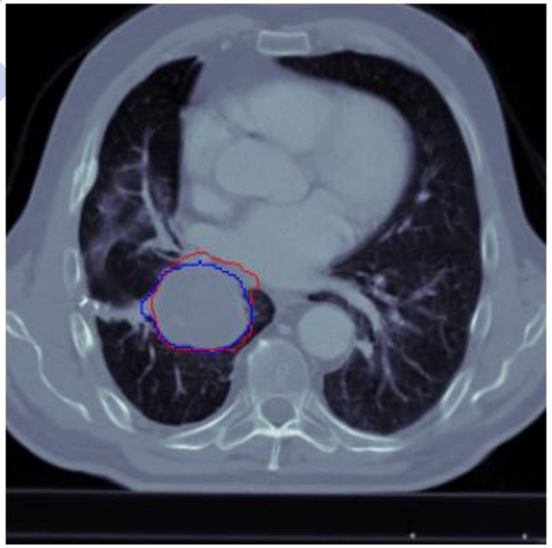
SFF-3D-UNET (DC: 0.86)

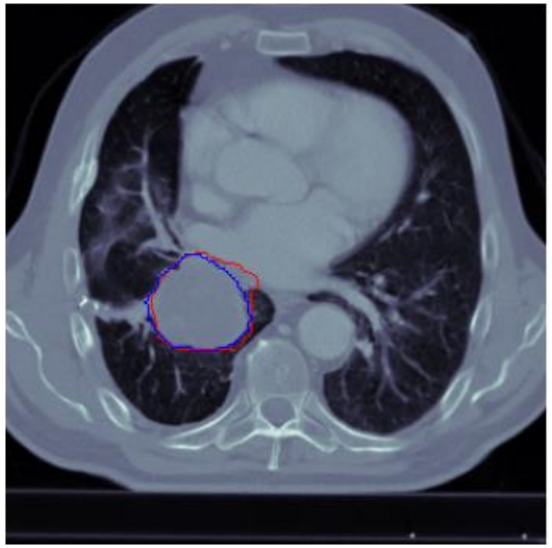


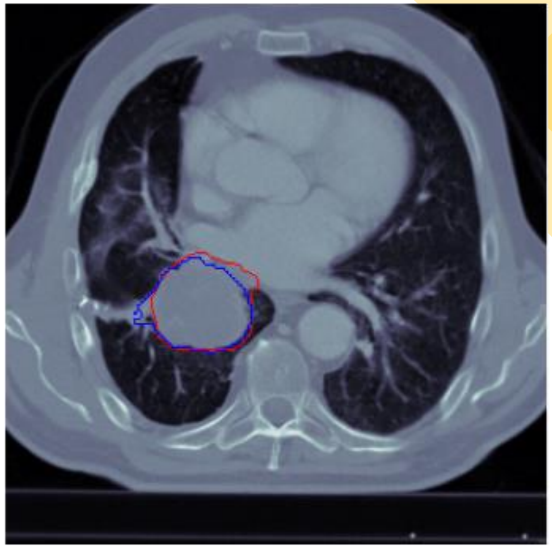
SFF-3D-MultiResUNet (DC: 0.92)

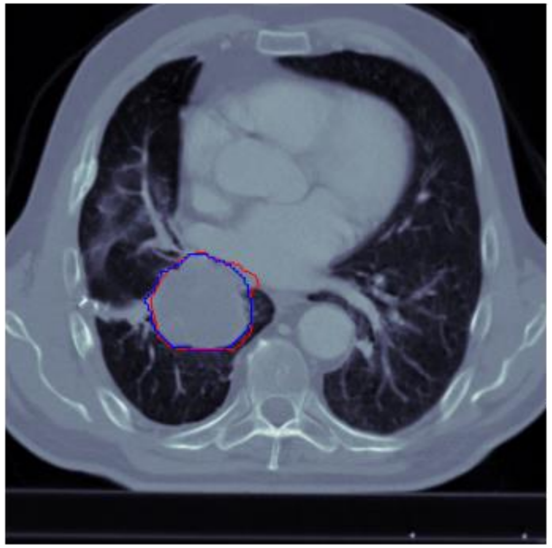


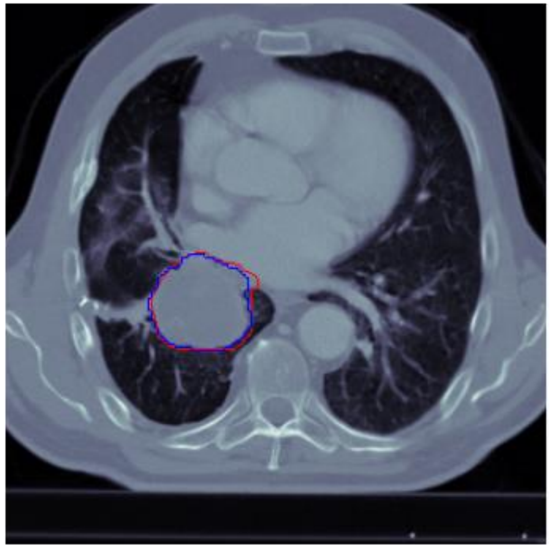
SFF-Recurrent-3D-DenseUNet (DC: 0.90)

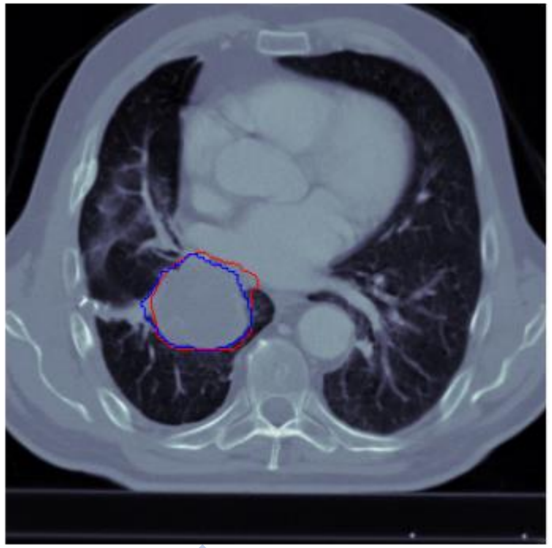

3D-UNet (DC: 0.90)

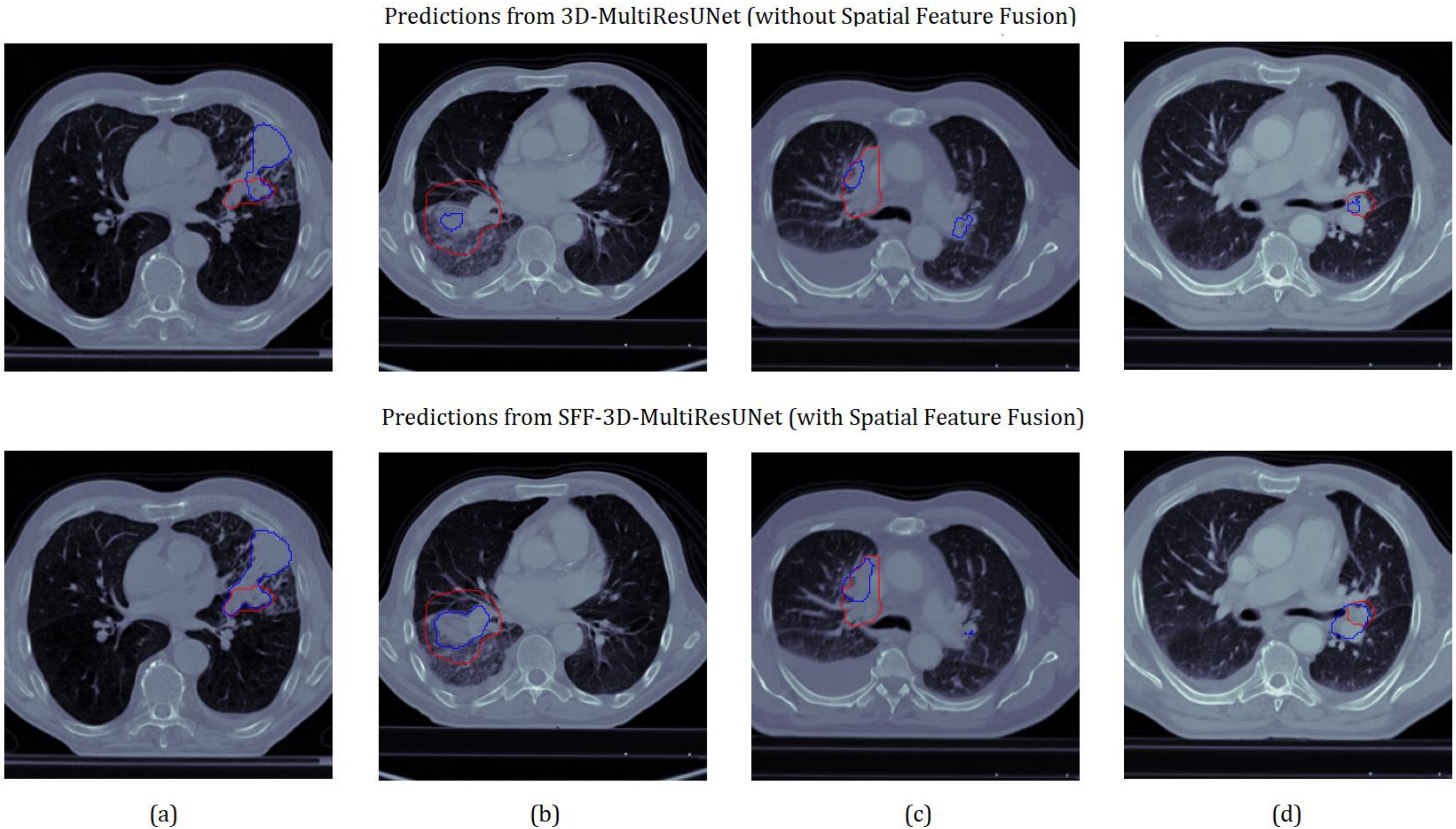

3D-MultiResUNet (DC: 0.92)


Recurrent-3D-DenseUNet (DC: 0.90)


SFF-3D-UNet (DC: 0.95)


SFF-3D-MultiResUNet (DC: 0.95)


SFF-Recurrent-3D-DenseUNet (DC: 0.92)


Predictions from 3D-MultiResUNet (without Spatial Feature Fusion)
Predictions from SFF-3D-MultiResUNet (with Spatial Feature Fusion)
(a)
(b)
(c)
(d)

Conclusion

- We have proposed three novel architectures which incorporate multi-scale spatial feature fusion and improve lung tumor segmentation performance with minimal computational overhead.
- Our proposed architectures achieved performance improvements of 1.61%, 2.25%, and 2.42% respectively in terms of 2D Dice Coefficient.
- Our proposed architectures also achieved performance improvements of 7.58%, 2.32%, and 4.28% in terms of 3D Dice Coefficient.
- Our proposed best model SFF-3D-MultiResUNet outperforms all approaches in the literature to achieve the best overall 2D Dice Coefficient (0.8669) on the LOTUS Benchmark.
- Our proposed approach can speed up lung cancer diagnostic process and has the potential of saving lives.

Future Scope of Work

- Explore architectural improvements of the baseline architectures to improve the performance of the overall pipeline.
- Implement advanced training strategies like deep supervision to improve the training of the baseline models.
- Extend our proposed approach to other biomedical image segmentation domains to improve volumetric segmentation performance.
- We plan to continue further research on this topic and explore different avenues to improve the overall pipeline and achieve better results.

Question & Answer Session

Thank You!