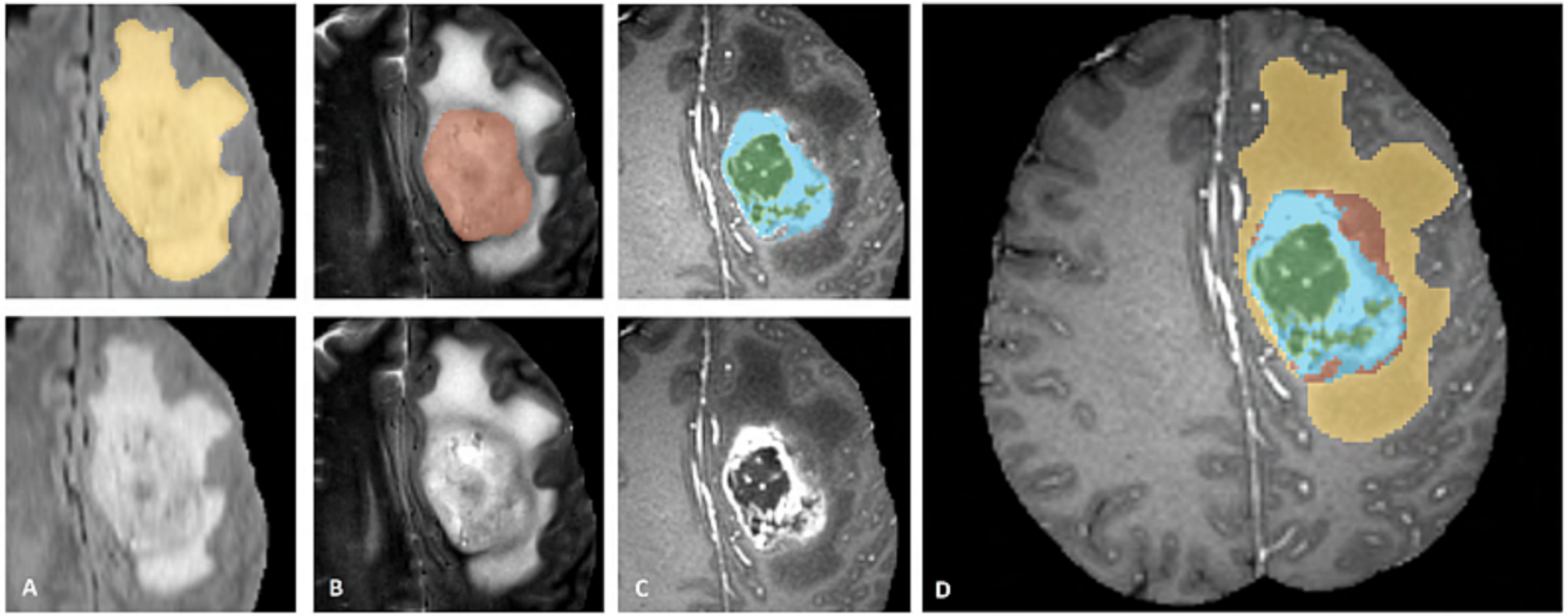
LSTM Multi-modal UNet for Brain Tumor Segmentation

ICIVC 2019, Xiamen China

Fan Xu, Haoyu Ma, Junxiao Sun, Rui Wu, Xu Liu, Youyong Kong*
Southeast University

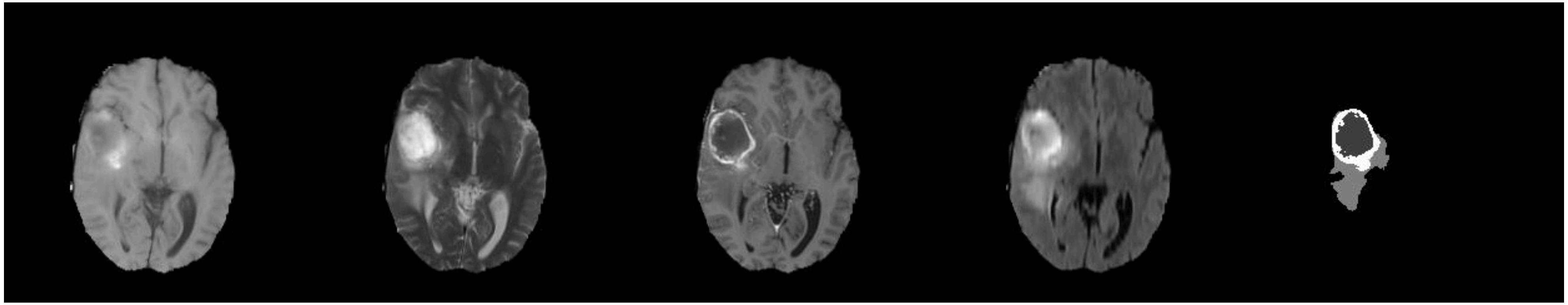
Brain Tumor Segmentation: Dataset and Task

- BraTS: Annual brain tumor segmentation challenge
- Generate segmentation of Whole Tumor, Tumor Core and Enhancing Tumor



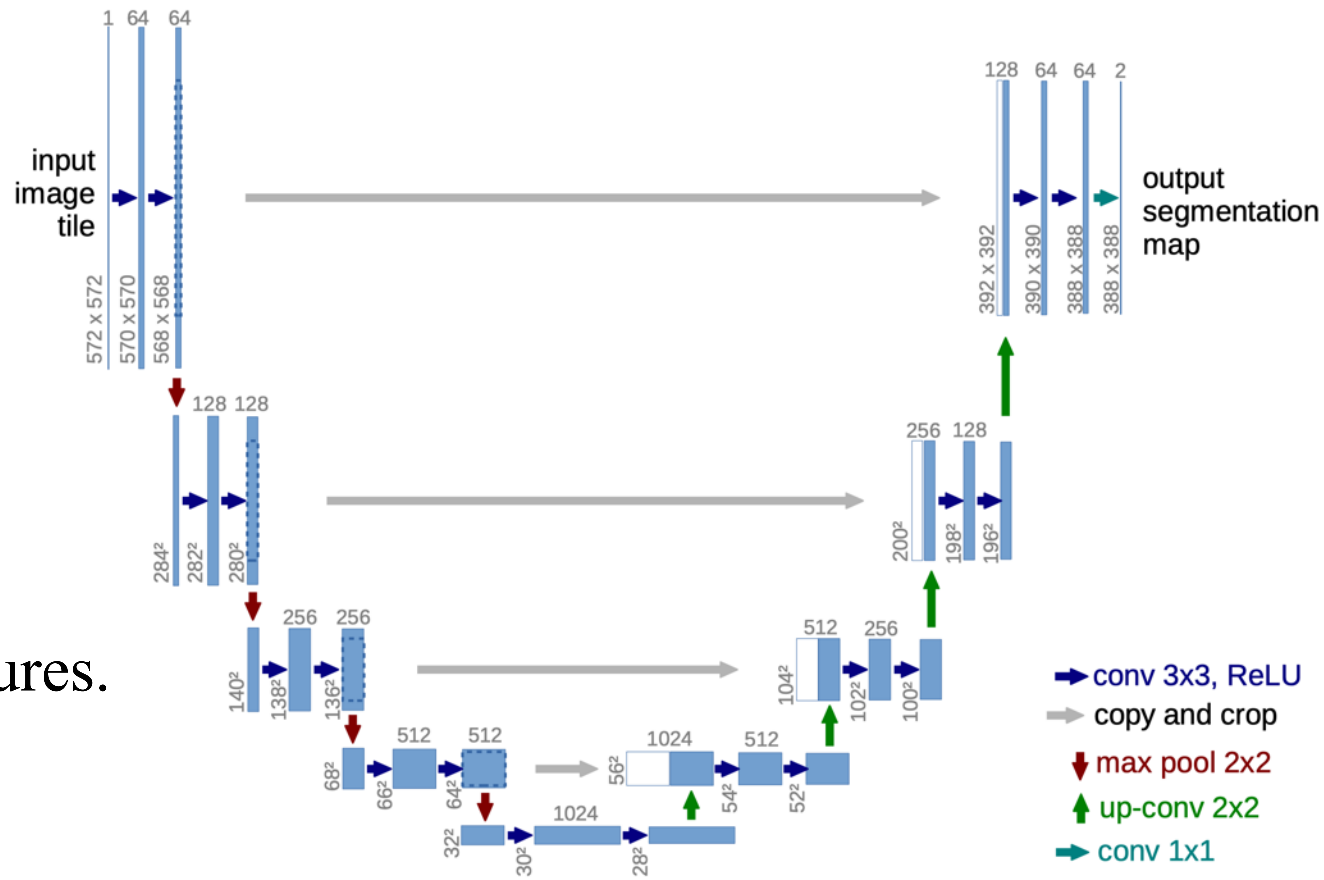
Brain Tumor Segmentation: Challenging points

- Different shape, size and location of brain tumor
- 3D images (size 155 x 240 x 240)
- Multi-modal Magnetic Resonance Imaging (MRI)



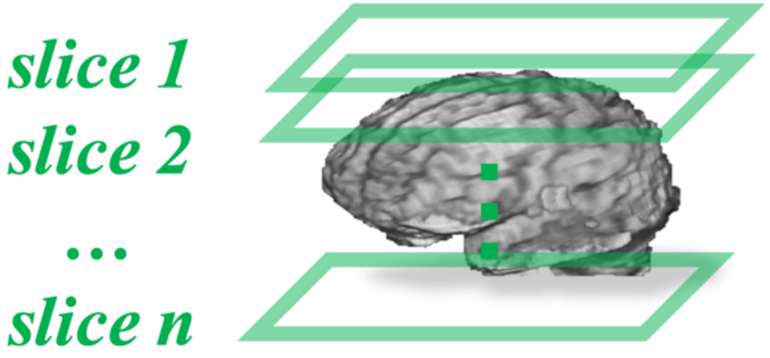
Flair T1 T1c T2 Label

Biomedical Image Segmentation Baseline

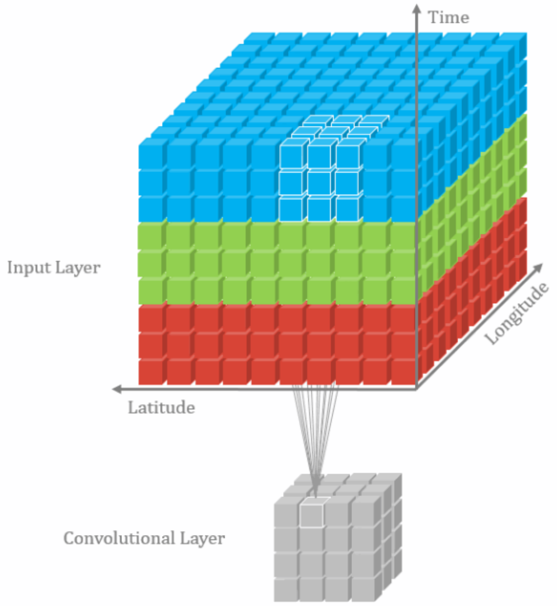


- U-Net
  - Down sample and up sample.
    - capture low and high level features.
  - Skip-connections.
    - transfer information during the compression process

Treatment of 3D Image



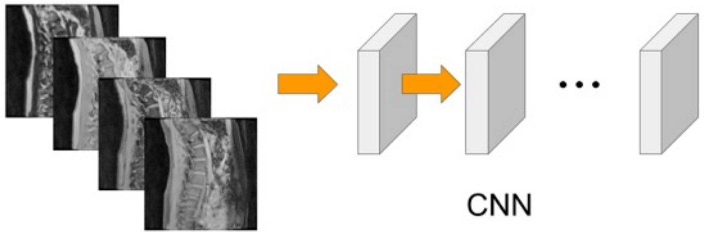
- Split 3D data into several 2D slices
  - Use 2D image based model
  - Neglect the depth information



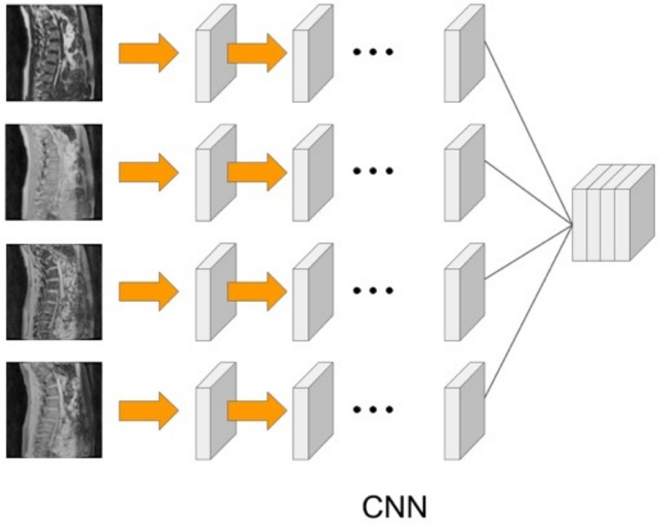
- Apply 3D Convolution
  - Model correlation between slices
  - Require larger number of parameters
- Use RNN/LSTM to capture the temporal information
  - Regard depth as temporal
  - 3D image = 2D video

Fusion strategies of multi-modal images

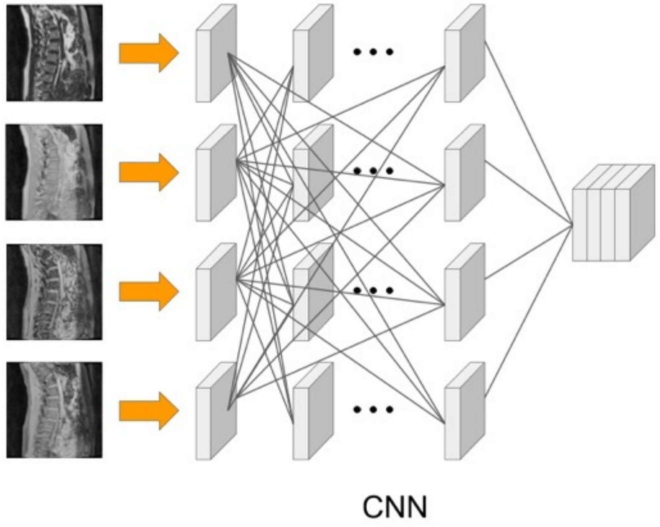
Early fusion



Late fusion



Hyper dense connection



Performance

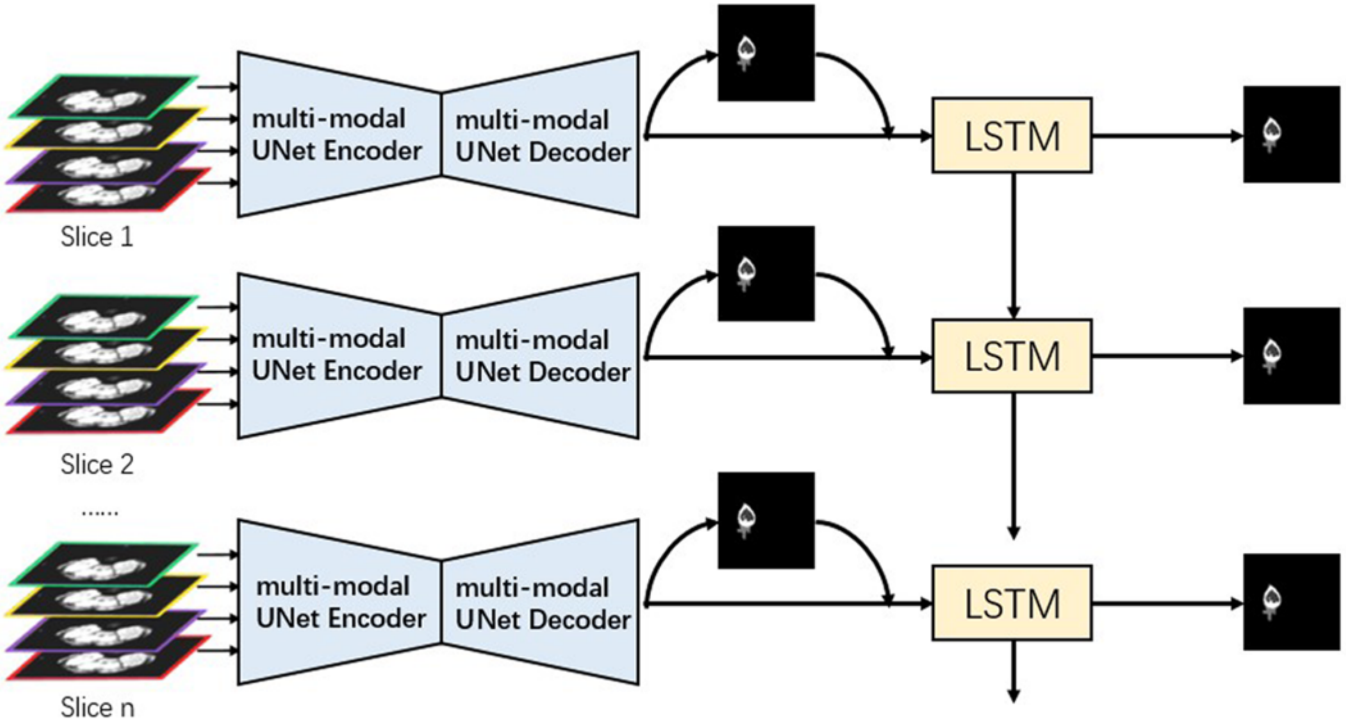
Hyper dense connection > late fusion > early fusion

What is LSTM multi-modal UNet?

- LSTM multi-modal UNet = Hyper Dense Connection UNet + convLSTM

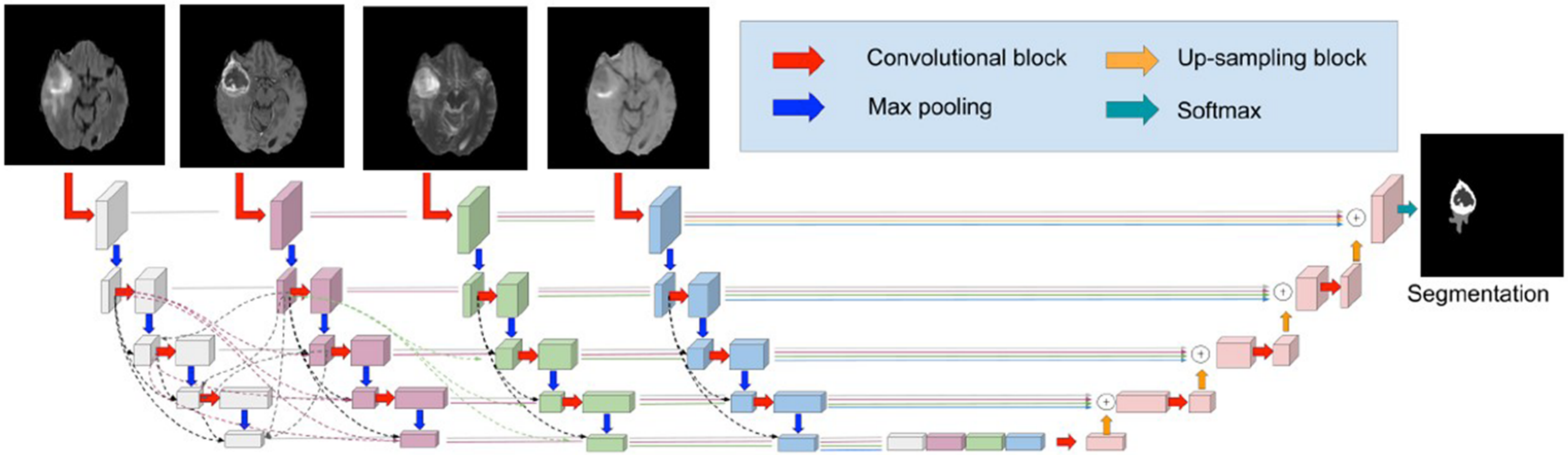
- Hyper Dense connections
  - Leverage multi-modal data

- convLSTM
  - Exploit depth information



Multi-modal UNet Architecture

- UNet-based encoder and decoder
- Multiple encoding paths for multimodal data
- **Hyper dense connections**



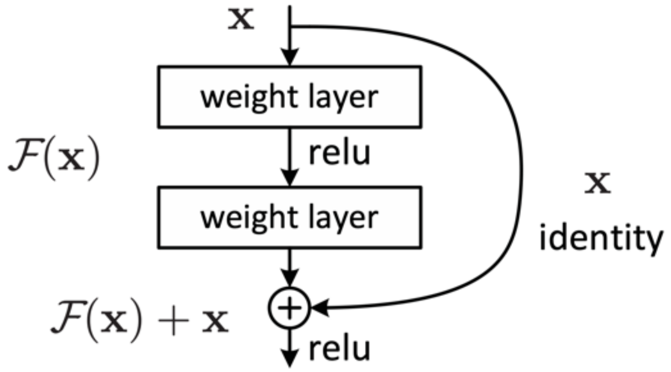
Hyper Dense Connections

- ResNets:

$$x_l = H_l(x_{l-1}) + x_{l-1}$$

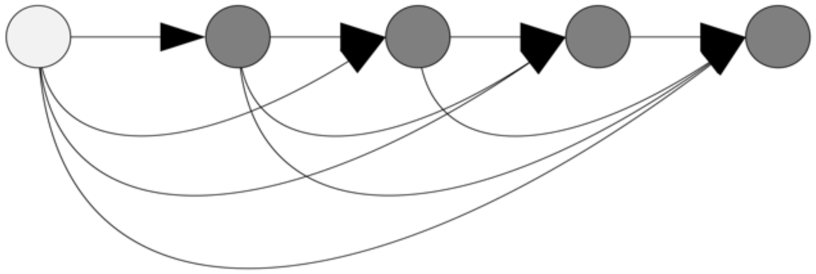
H_l mapping function

x_l output of l^th layer



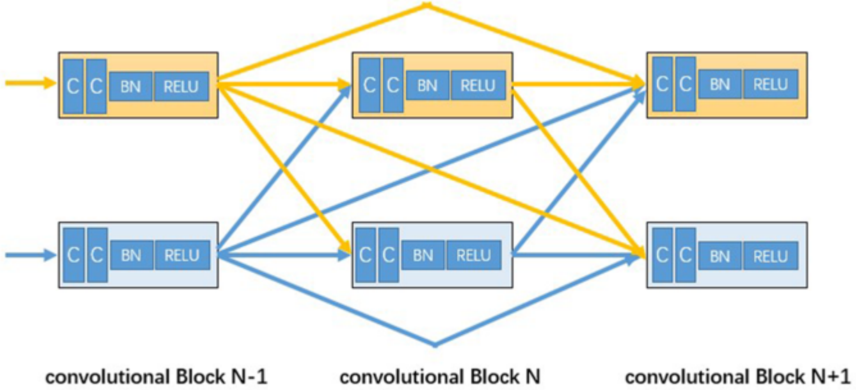
- Dense Net:
  - Concatenate all previous layers features

$$x_l = H_l([x_{l-1}, x_{l-2}, \ldots, x_0])$$



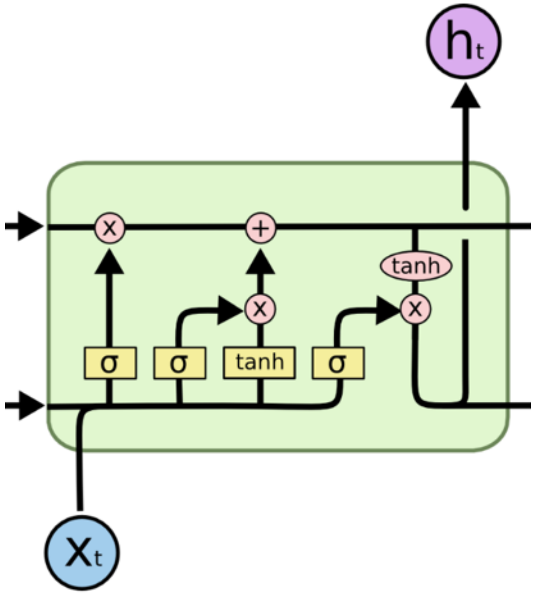
- Hyper Dense Net:
  - Concatenate all previous layers features in all paths

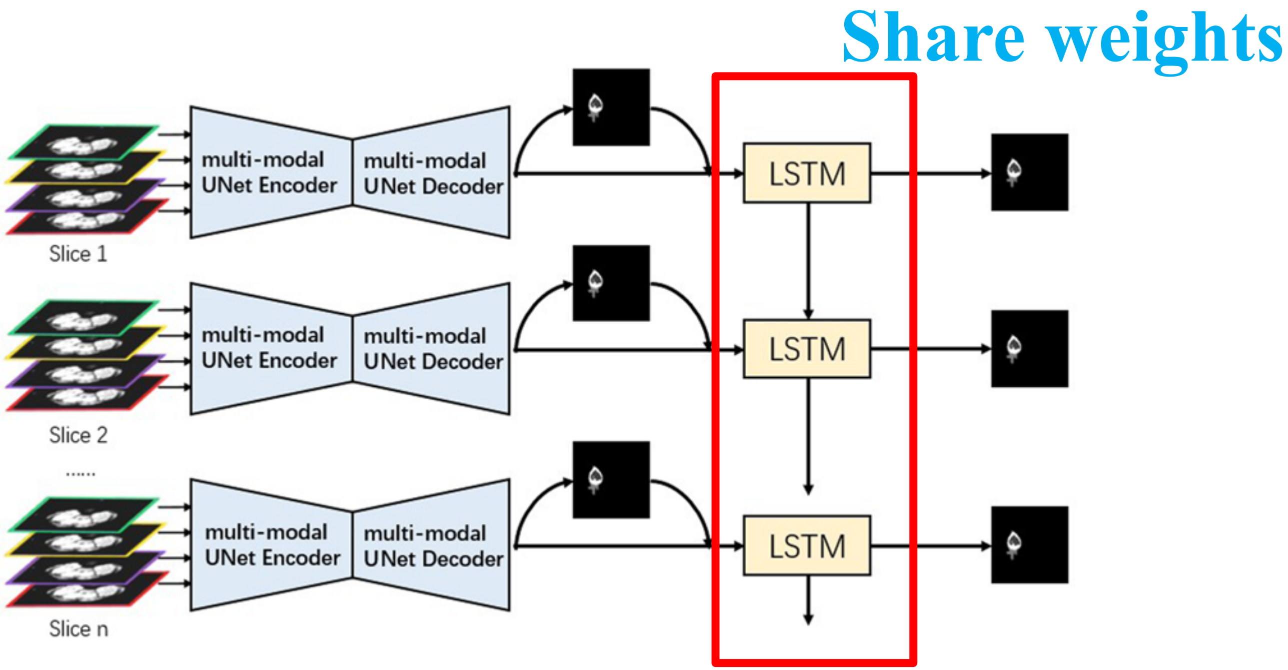
$$x_l^p = H_l^p([x_{l-1}^1, x_{l-1}^2, x_{l-2}^1, x_{l-2}^2, \ldots, x_0^1, x_0^2])$$



Slice Sequence Learning: convolutional LSTM

- Regard depth as temporal information
- Replace the matrix multiplication by a convolution operator
  - Reserves the spatial information for sequences
- Share weights for different slices





Model details

- Half channel number of UNet
  - Less number of model parameters
  - To prove parameters are not the decisive factor

COMPARE OF NETWORK SIZE

	Number of parameters	model size
U-Net	34530437	138.2MB
Ours	28713450	115.6MB

DETAIL INFORMATION OF NETWORK CHANNELS

	Name	**Feat maps(input)**	**Feat maps(output)**
	U Net		
Encoding	Conv layer 1	4×240×240	**64**×240×240
	Max pooling 1	64×240×240	64×120×120
	Conv layer 2	64×120×120	**128**×120×120
	Max pooling 2	128×120×120	128×60×60
	Conv layer	128×60×60	**256**×60×60
	Max pooling 3	256×60×60	256×30×30
	…		
	Multi-modal UNet		
Encoding (each mod)	Conv layer1	1×240×240	**32**×240×240
	Max pooling 1	32×240×240	32×120×120
	Conv layer 2	32×120×120	**64**×120×120
	Max pooling 2	64×120×120	64×60×60
	Conv layer 3	64×60×60	**128**×60×60
	Max pooing 3	128×60×60	128×30×30
	…		

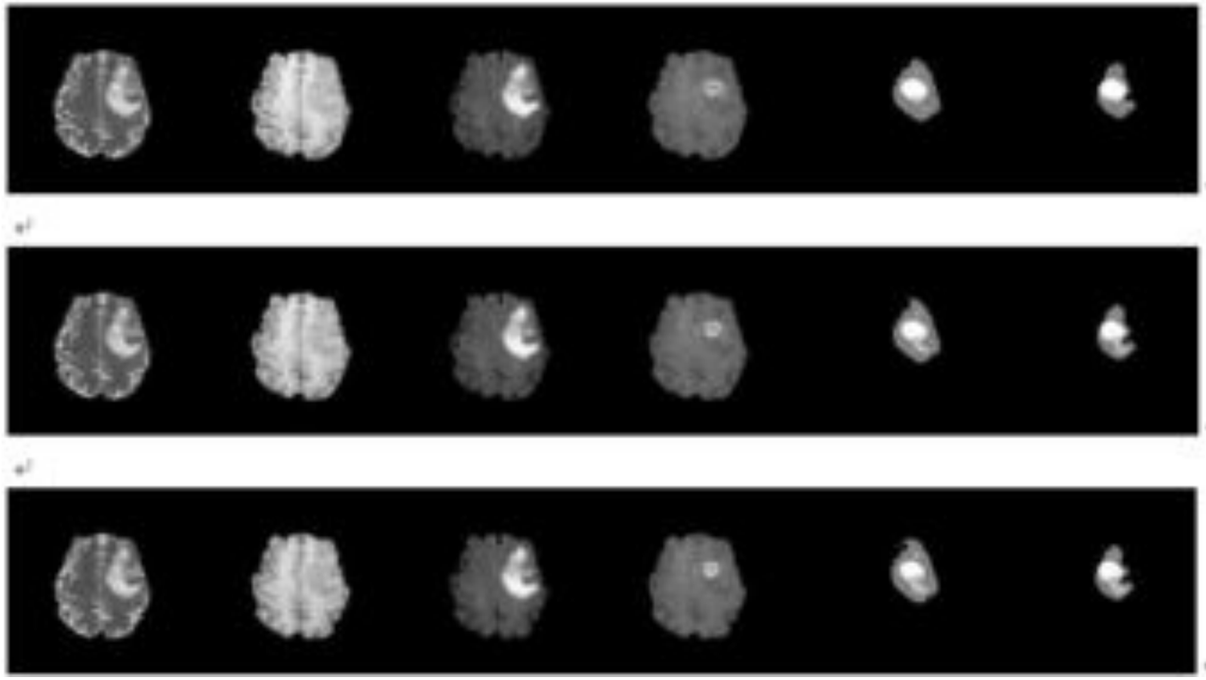
Experiments

- Dataset:
  - BraTS 2015 (224 subjects for training, 50 for testing)
- Optimizer:
  - Adam (default parameters)
- Loss:
  - cross entropy loss with median frequency balance
- Contrast model:
  - vanilla UNet
- Train from scratch
  - Same hyper parameters for both UNet and our model

Results

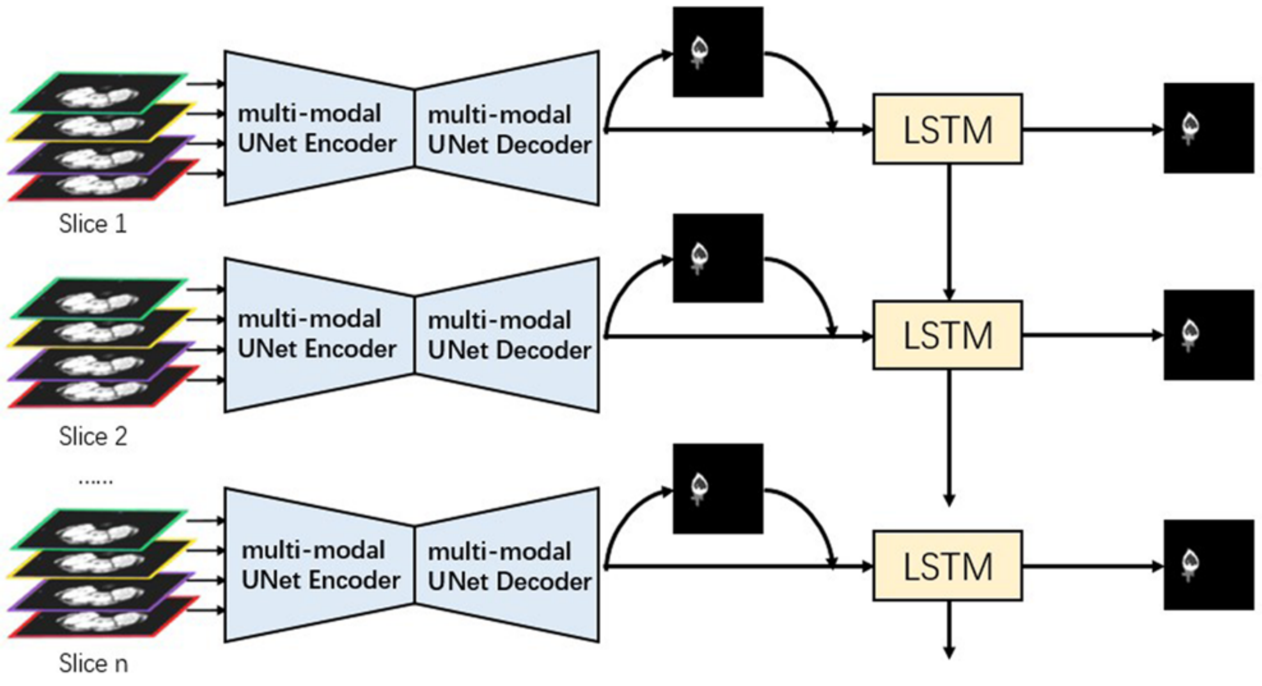
EVALUATION CRITERIA OF BRATS-2015

	Network	Complete	Core	Enhancing
Dice	UNet	0.7171	0.5989	**0.5022**
	ours	**0.7309**	**0.6235**	0.4254



Conclusion

- LSTM multi-modal UNet = Hyper Dense Connection UNet + convLSTM
- Exploit correlations between multimodal data and depth information
- Better performance on Brats 2015 with less parameters than UNet



Thanks for listening

Q & A