

# MRI Tumor Segmentation with Densely Connected 3D CNN

**Lele Chen**, Yue Wu, Adora M. DSouze,  
Anas Z. Abidin, Axel Wismüller, and **Chenliang Xu**



# MRI Brain Tumor Segmentation

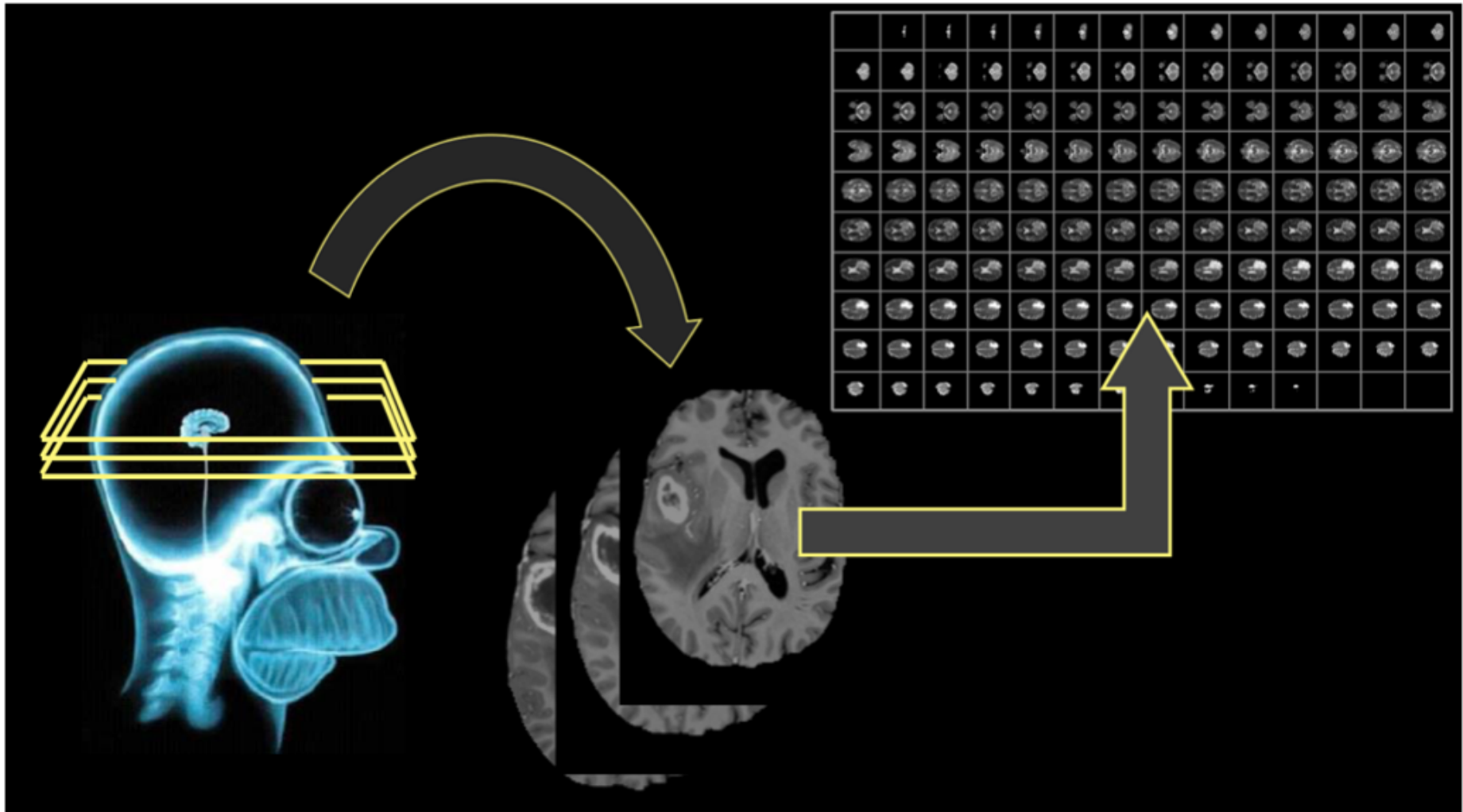


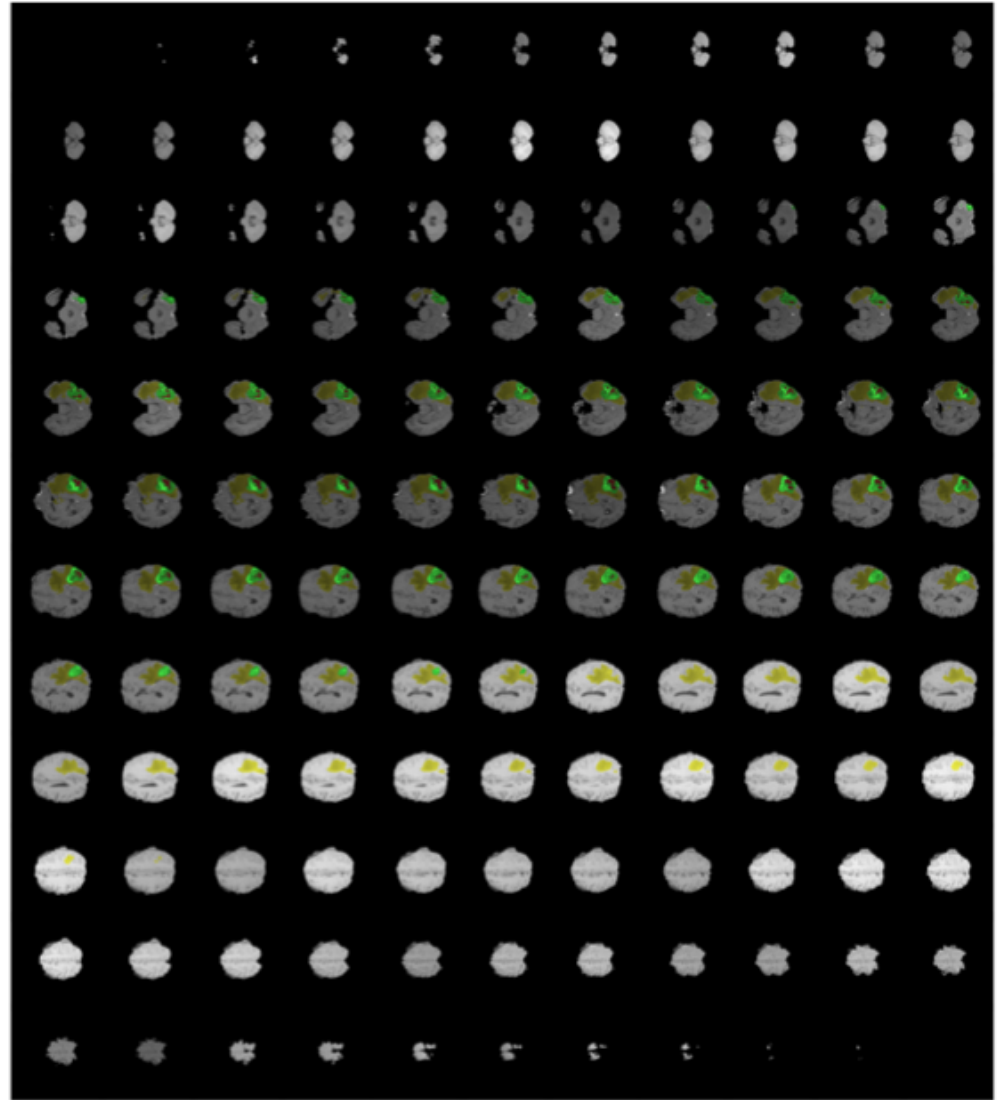
Image source: [https://github.com/naldeborgh7575/brain\\_segmentation](https://github.com/naldeborgh7575/brain_segmentation)

# MRI Brain Tumor Segmentation



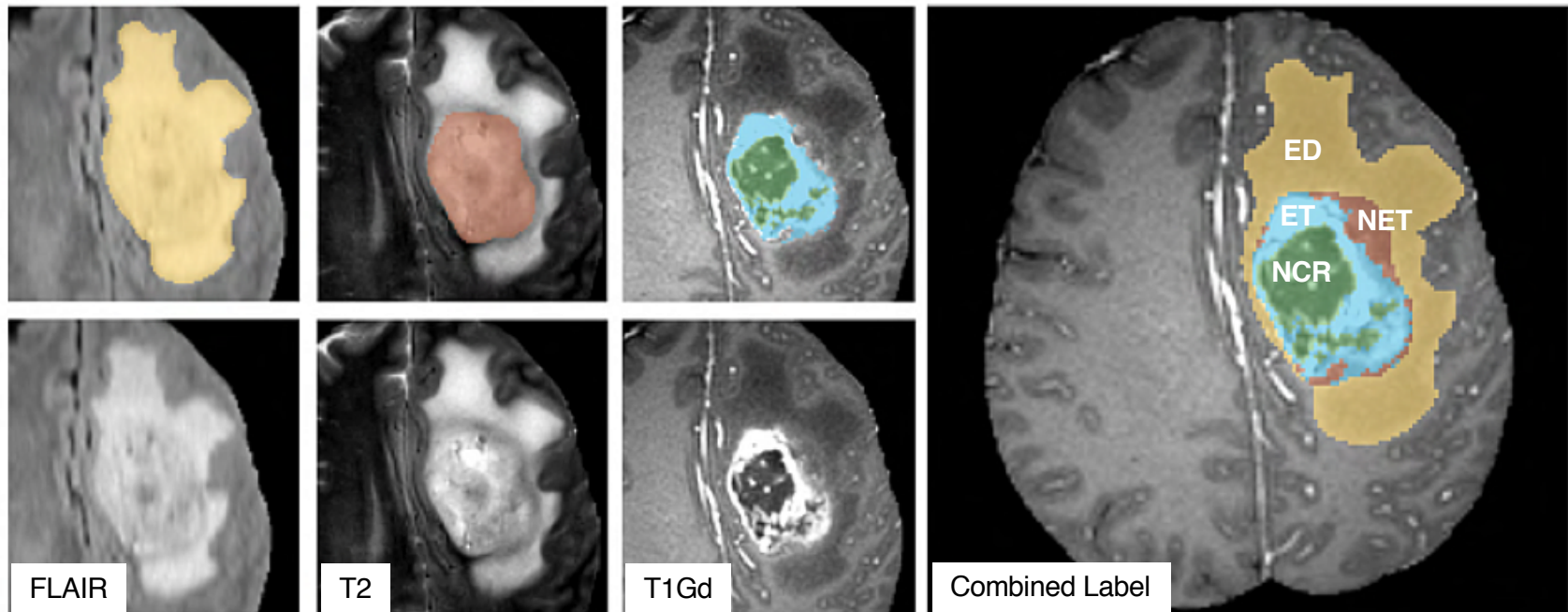
The lesion regions only occupy a small part in the whole 3D image.

- Necrotic (NCR) & NET
- Enhancing Tumor (ET)
- Edema (ED)



# MRI Brain Tumor Segmentation

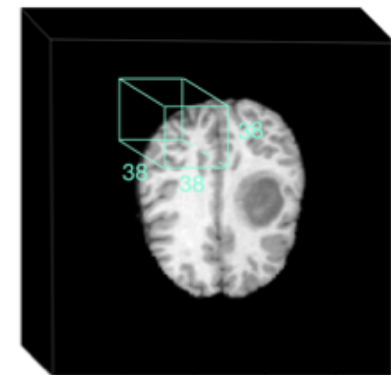
- Multimodal scans describe: a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes.



(Figure taken from the [BraTS IEEE TMI paper](#).)

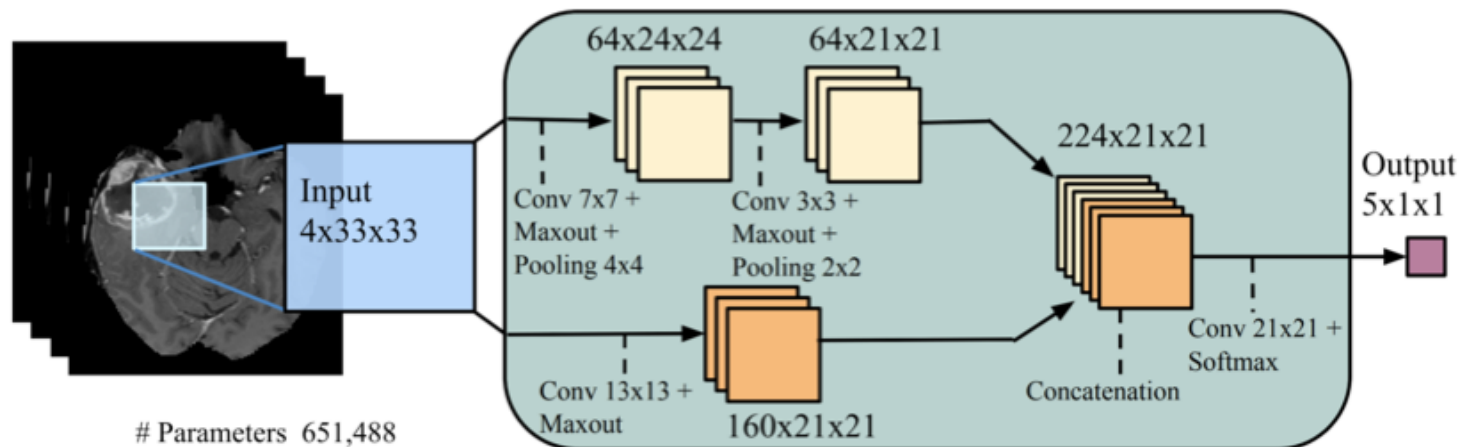
## Related Work

- A rich body of the literature in BraTS challenge.
- The lesion areas in MRI scans can be extremely small in many cases causing background domination.
- Patch-wise segmentation methods: extract small patches of the whole 2D/3D images with a pre-defined probability of being centered on lesion areas.
  - Solves background domination problem.
  - Mitigates the class imbalance problem.
  - Saves memory, so we can develop a deeper network.



# Related Work

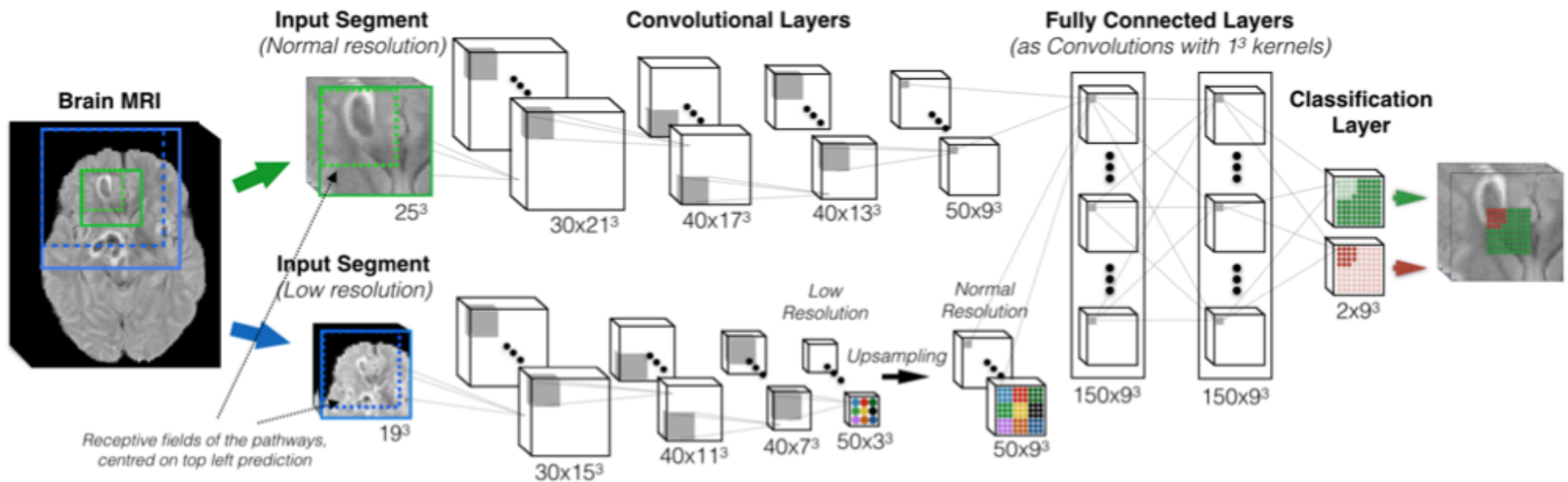
- Patch classification: [Pereira et al. TMI'16], [Zhao et al. BrainLes'16], [de Brébisson et al. CVPR'15], [Havaei et al. MIA'17]
  - Usually contains FC layers outputting the probabilities of a single voxel (center of patch) being in different categories.
  - Post-processing, e.g., CRF, to improve predictions.



[Havaei et al. MIA'17, Brain tumor segmentation with deep neural networks.]

# Related Work

- Patch segmentation:
  - Outputs a dense prediction of the labels for multiple voxels in that patch simultaneously (computationally efficient).
  - Optionally post-processing with CRF (gains <1%).
  - Our method is in the same vein of this line of research.



[Kamnitsas et al. MIA'17, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation.]

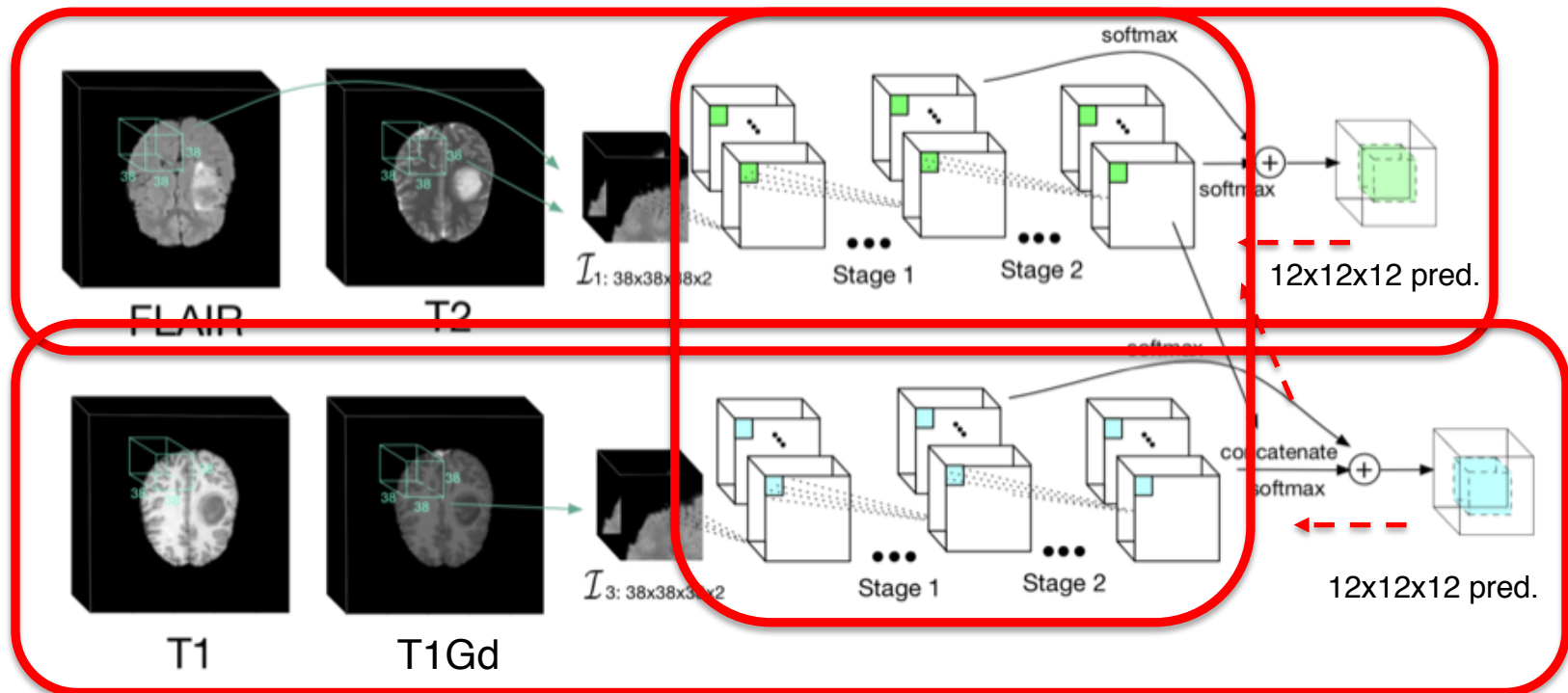
# Our Method

- We start with a 3D patch segmentation CNN and enhance it in three folds.
- 1) Incorporate the structured knowledge of multimodal scans in segmentation networks.
  - **A hierarchical segmentation structure**
- 2) Allow networks to go deeper but being parsimonious of parameters and encourage feature reuse.
  - **Generalize densely connected blocks to 3D images**
- 3) Being sensitive to the multi-scale contextual information when predicting lesion regions.
  - **Multi-scale receptive fields.**



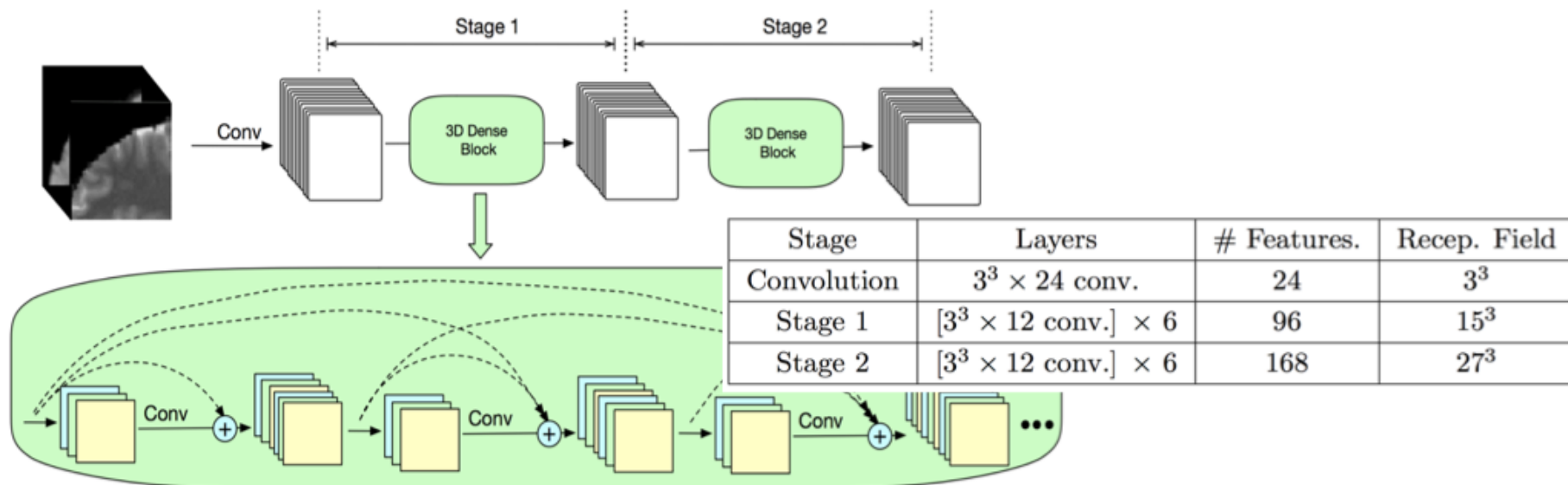
# Hierarchical Segmentation Structure

- The top pathway segments whole tumor from the background.
- The bottom pathway concatenates features from the top pathway and performs a four-class segmentation.
- Two pathways share the same structure and are trained end-to-end together.



# Two-Stage Densely Connected 3D CNN

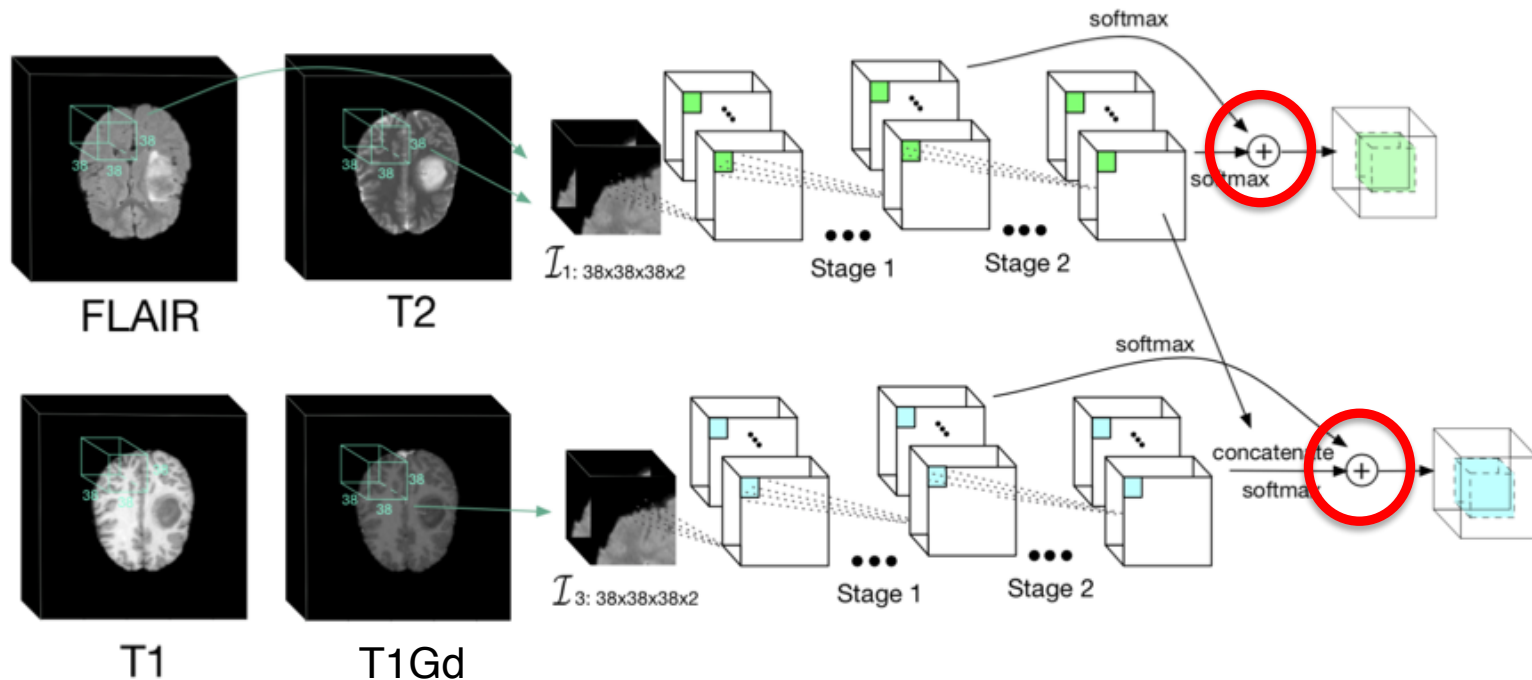
- We generalize DenseNet [Huang et al. CVPR'17] to 3D CNN.
- One layer in a densely connected block receives input from all previous layers:  $\mathbf{x}_\ell = H_\ell([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\ell-1}])$ .
- Define  $H_\ell(\cdot)$  as a composition:
  - BN-ReLU-Conv(1x1x1)-BN-ReLU-Conv(3x3x3).
- The feature dimension of each layer:  $d_l = d_0 + g \cdot l$ .



# Two-Stage Densely Connected 3D CNN

- An  $1^3 \times 168$  convolution operation follows the dense block of each stage merging features from all levels of abstraction.
- Scores at two stages (of different receptive fields) are merged.

Stage	Layers	# Features.	Recep. Field
Convolution	$3^3 \times 24$ conv.	24	$3^3$
Stage 1	$[3^3 \times 12 \text{ conv.}] \times 6$	96	$15^3$
Stage 2	$[3^3 \times 12 \text{ conv.}] \times 6$	168	$27^3$



# Experiments: Data and Setup

- Dataset: BraTS 2017
  - Missed the evaluation server at the time of submission.
  - Training set: 162 HGG patients (80%).
  - Testing set: 48 HGG patients (20%).
- Dice metric: 
$$\text{Dice} = \frac{2TP}{FN + FP + 2TP}$$
- Data preprocessing
  - Employ N4ITK bias correction.
  - Randomly sample voxels with 50% being in lesion regions.
  - Extract 38x38x38 3D patches; concatenate patches from FLAIR and T2, and concatenate patches from T1 and T1Gd as input to our two-pathway architecture.

# Experiments: Training and Testing

- Networks predict 12x12x12 masks from 38x38x38 patches.
- Loss function: cross entropy over voxels.
- Optimizer: Adam is applied with the initial learning rate:  $5e-4$ .
- Training data:
  - For each epoch, we randomly extract 400 patches with 200 of them centered on the lesion regions and others on bg.
- Training Time:
  - It takes 2.5 hours to train for an epoch on a single NVIDIA GTX 1080 GPU. The network converges after around 5 epochs.
- Testing:
  - Sliding window approach to obtain a dense prediction for an MRI sequence.

# Experiments: Ablation Study

- Non-Dense Structure:
  - Linear chain topology.
  - The same size of receptive fields.
  - More parameters are introduced.

	Comp.	Core	Enh.
Non-Dense (See Tab. 4)	0.61	0.77	0.78
Non-Hierarchical	<b>0.74</b>	0.74	0.75
Single-Scale Recep. Field	0.62	0.64	0.62
Proposed	0.72	<b>0.83</b>	<b>0.81</b>

Stage	Layers
Convolution	$3^3 \times 24$ conv.
Stage 1	$3^3 \times 36$ conv.
	$3^3 \times 48$ conv.
	$3^3 \times 60$ conv.
	$3^3 \times 72$ conv.
	$3^3 \times 84$ conv.
	$3^3 \times 96$ conv.
Stage 2	$3^3 \times 108$ conv.
	$3^3 \times 120$ conv.
	$3^3 \times 132$ conv.
	$3^3 \times 144$ conv.
	$3^3 \times 156$ conv.
	$3^3 \times 168$ conv.

# Experiments: Ablation Study

- Non-Hierarchical Structure:
  - Discard the two-pathway architecture.
  - Concatenate all four types of scans as the inputs.
  - Network directly predicts the scores for all labels.

	Comp.	Core	Enh.
Non-Dense (See Tab. 4)	0.61	0.77	0.78
Non-Hierarchical	<b>0.74</b>	0.74	0.75
Single-Scale Recep. Field	0.62	0.64	0.62
Proposed	0.72	<b>0.83</b>	<b>0.81</b>

# Experiments: Ablation Study

- Non-Hierarchical Structure:
  - Discard the two-pathway architecture.
  - Concatenate all four types of scans as the inputs.
  - Network directly predicts the scores for all labels.
- Single-Scale Receptive Field:
  - Remove all layers in stage 2 and make predictions only with the output of stage 1.

	Comp.	Core	Enh.
Non-Dense (See Tab. 4)	0.61	0.77	0.78
Non-Hierarchical	<b>0.74</b>	0.74	0.75
Single-Scale Recep. Field	0.62	0.64	0.62
Proposed	0.72	<b>0.83</b>	<b>0.81</b>



# Experiments: Compare with Others

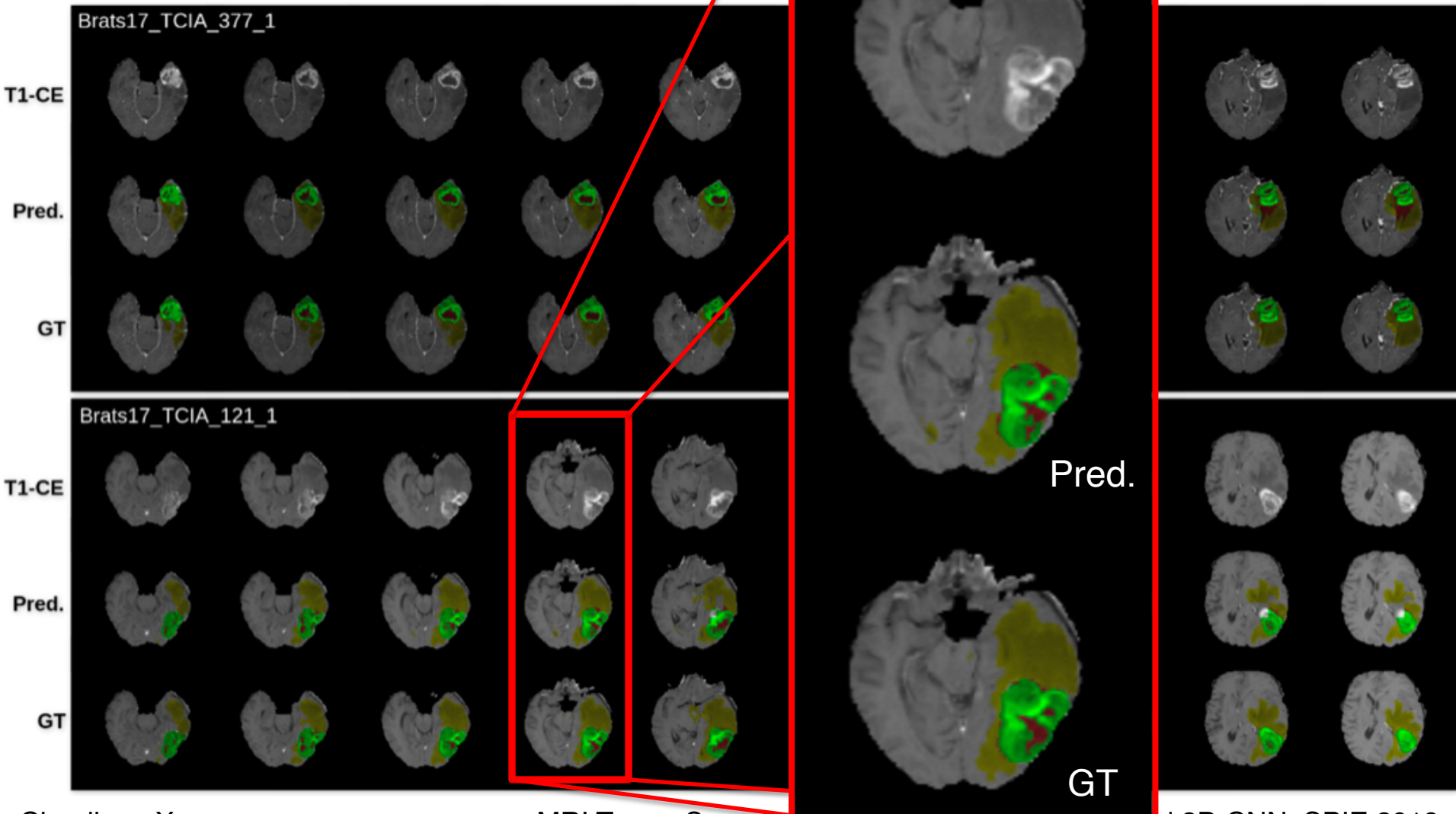
- Kamnitsas et al. is a strong 3D patch segmentation method.
- Both Pereira et al. and Zhao et al. apply 2D CNN models that take 33x33 patches as inputs and predict the label of the center voxel. Zhao et al. further use CRF as post-processing.

	Comp.	Core	Enh.
Pereira et al. <sup>4</sup>	0.84	0.72	0.62
Kamnitsas et al. <sup>3</sup>	<b>0.90</b>	0.75	0.73
Zhao et al. <sup>5</sup>	0.87	<b>0.83</b>	0.76
Ours	0.72	<b>0.83</b>	<b>0.81</b>

(Numbers are reported by respective authors' papers.)

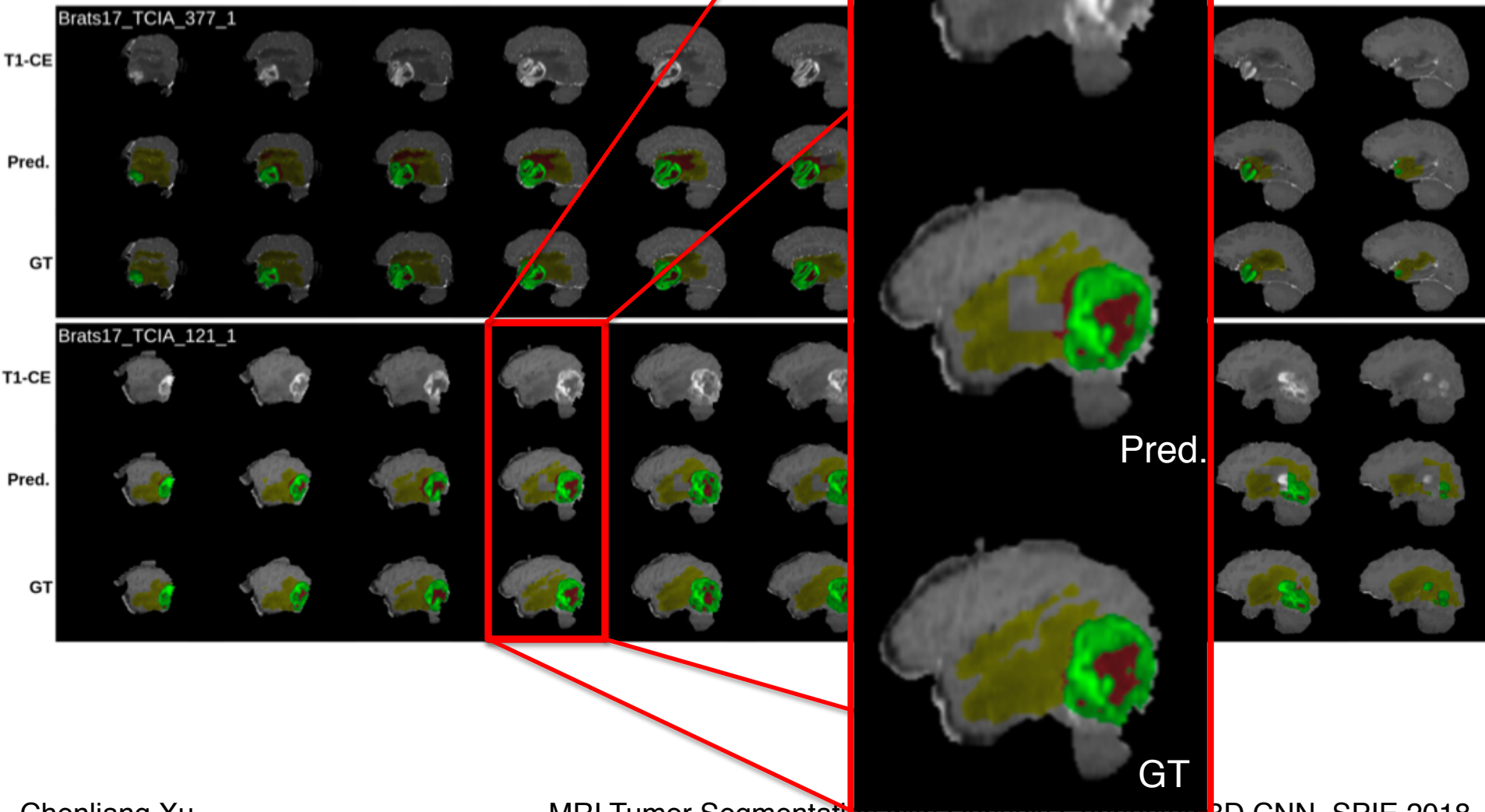
# Experiments: Qualitative Results

- View from the axial plane.



# Experiments: Qualitative Results

- View from the sagittal plane.



# Conclusion

- A hierarchical segmentation structure to incorporate the structured knowledge of multimodal scans.
- Generalize densely connected blocks to 3D images, which improves feature extraction and help optimization.
- Leverage multi-scale receptive fields to better model contextual information in prediction.

## Thank you! Questions?

- Our code is available at: <https://github.com/lelechen63/MRI-tumor-segmentation-Brats>.