Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation

Konstantinos Kamnitsas, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, Ben Glocker

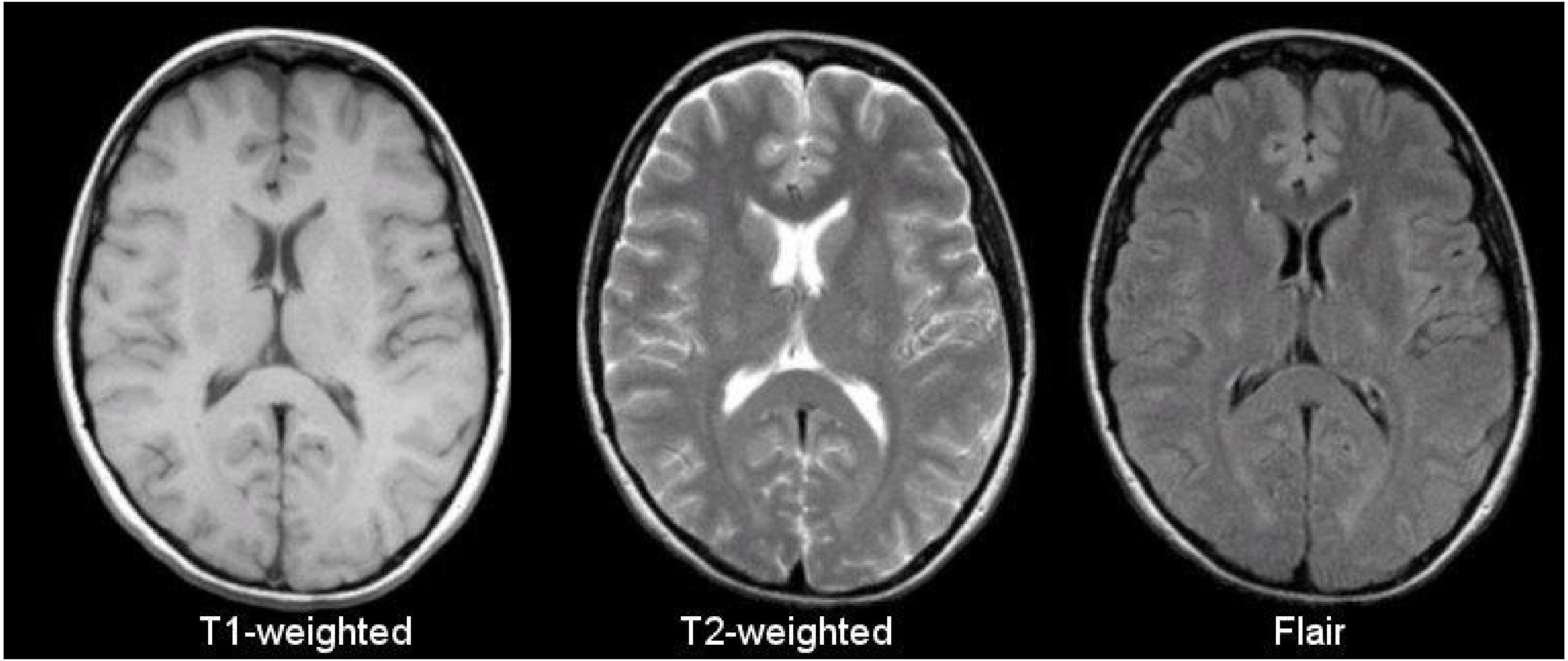
Team 12

Background

Magnetic Resonance Imaging (MRI)

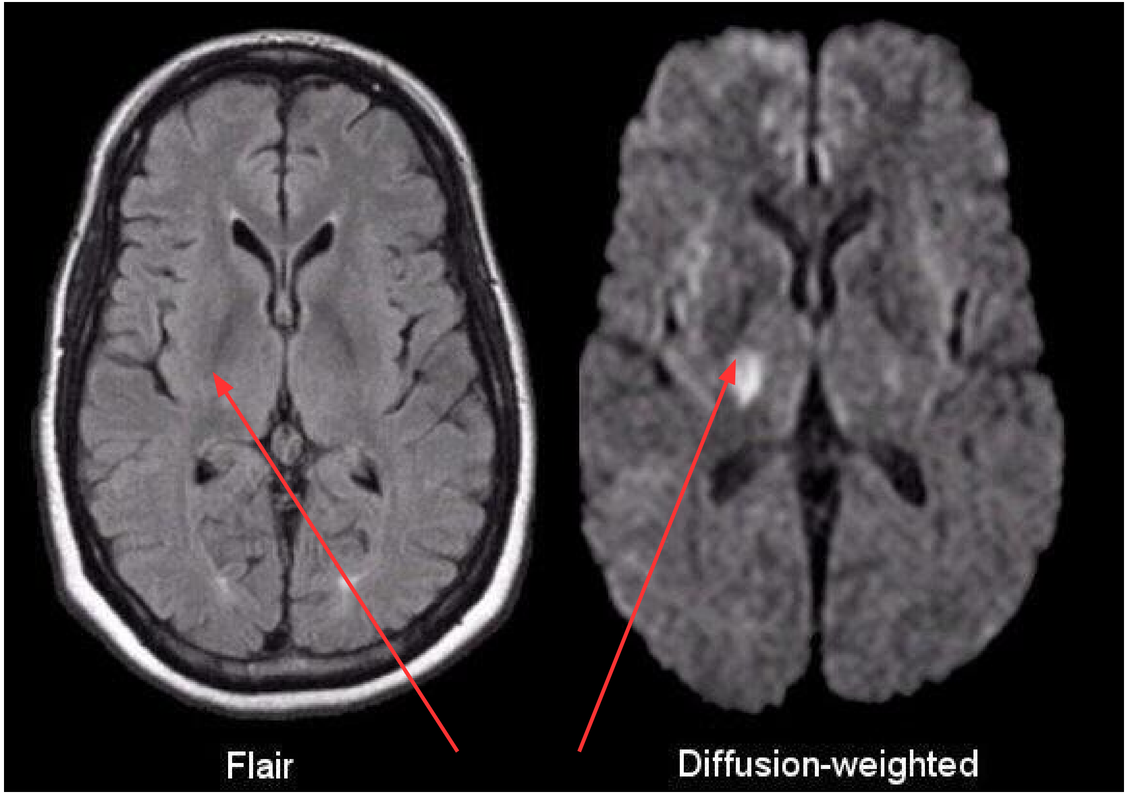
- Versatile, common imaging technique
- No risk of tissue damage or cancer from ionizing radiation
- Often used for diagnosing neurological conditions

MRI Sequence Types



Different sequence types highlight different tissues

Different sequences for different lesions



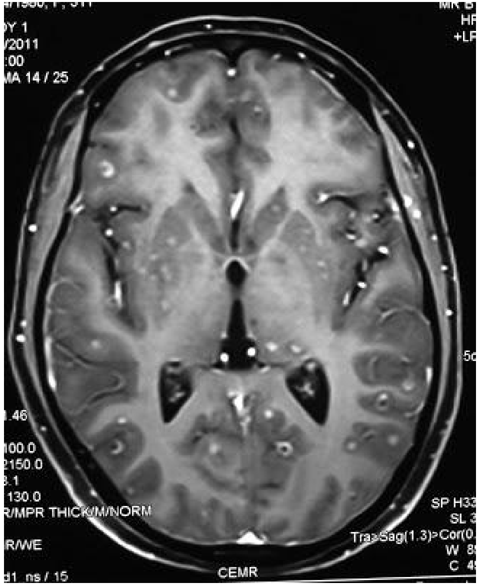
Infarction (obstruction of blood supply) only visible on right

Brain Lesions

- Lesions are sites damaged or changed by disease or trauma.
- Examining lesions on MRI allows doctors to identify and treat a patient's condition.

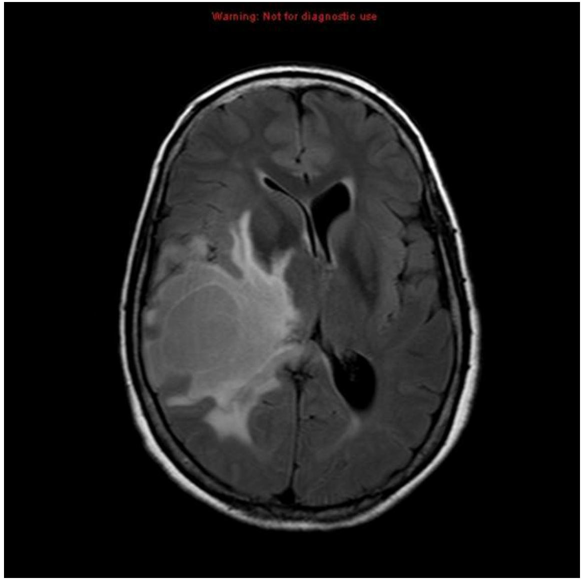
Many types of brain lesion are possible

Neurocysticerosis
(parasitic infection)



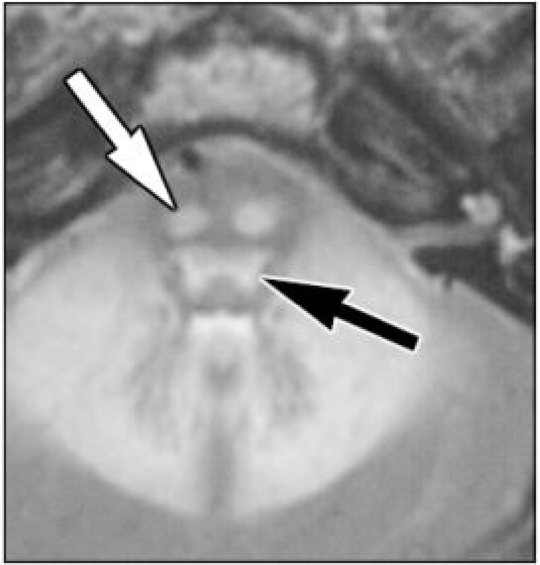
"Starry sky" punctate
lesions

Tumefactive demyelinating
lesion



Single, large mass distorting
other structures ("mass
effect")

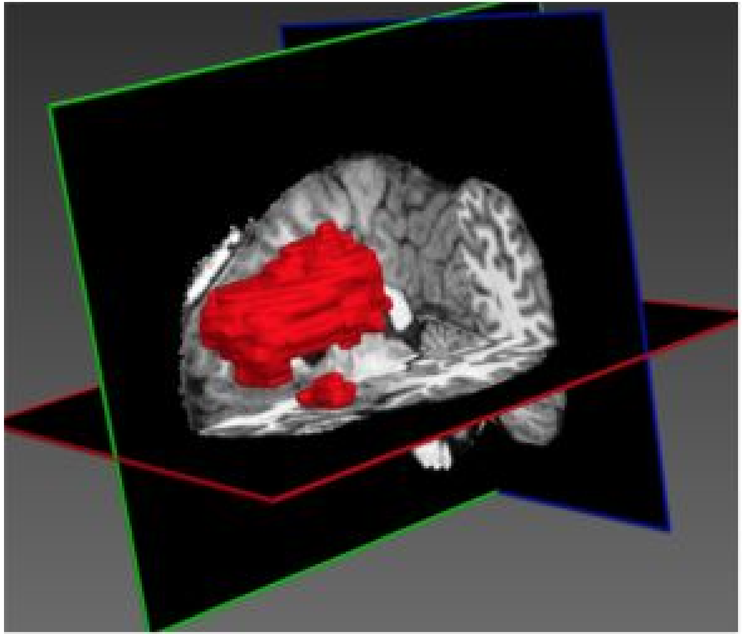
Heroin inhalation toxicity
("Chasing the dragon" sign)



"Image resembles
a bearded skull"

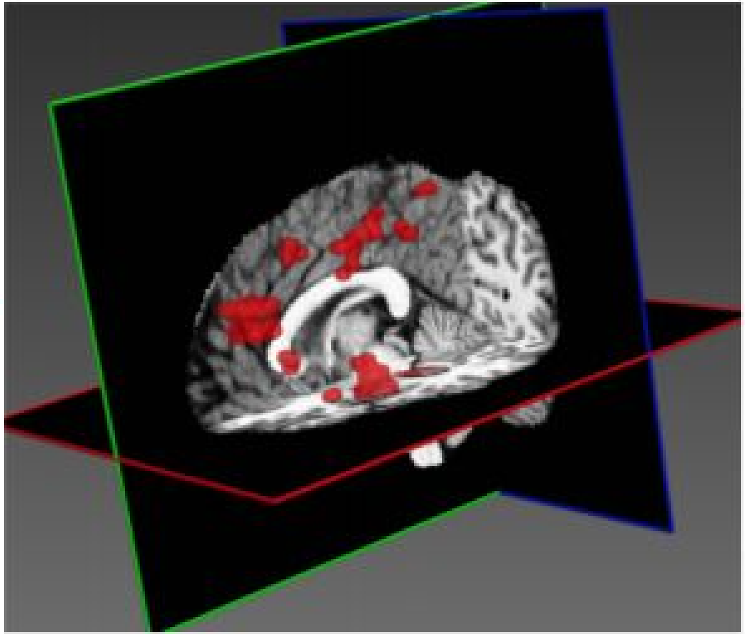
MRI example

Traumatic Brain Injury (TBI) lesions:



(a)

One large affected area

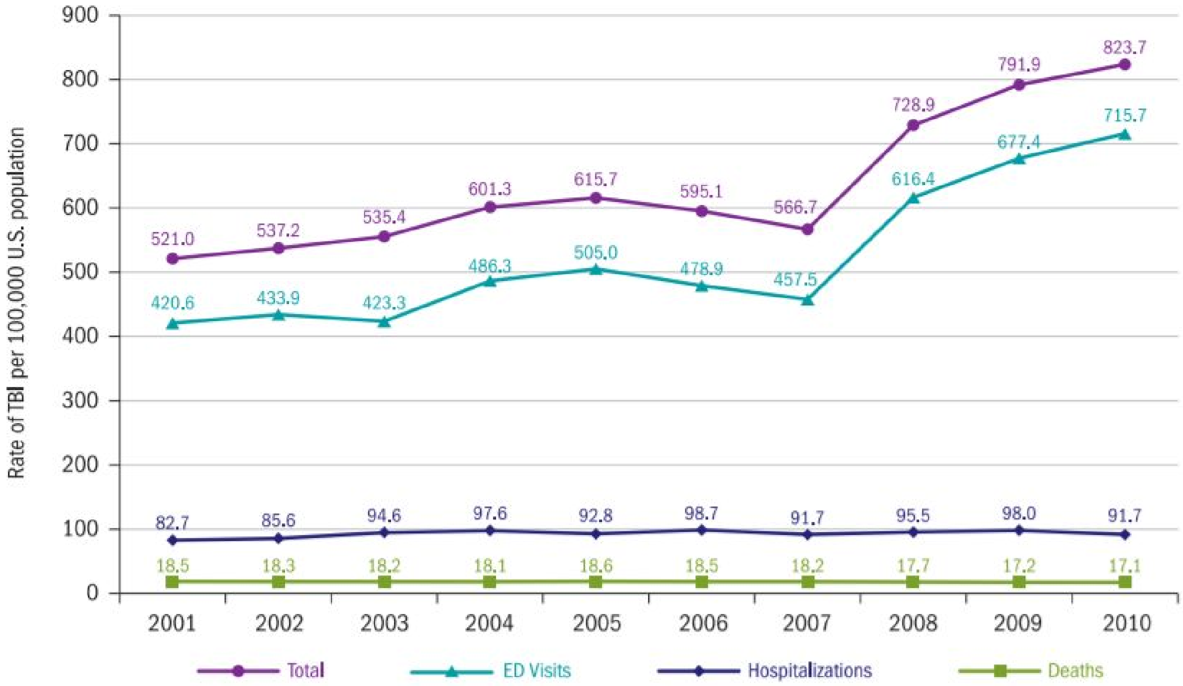


(b)

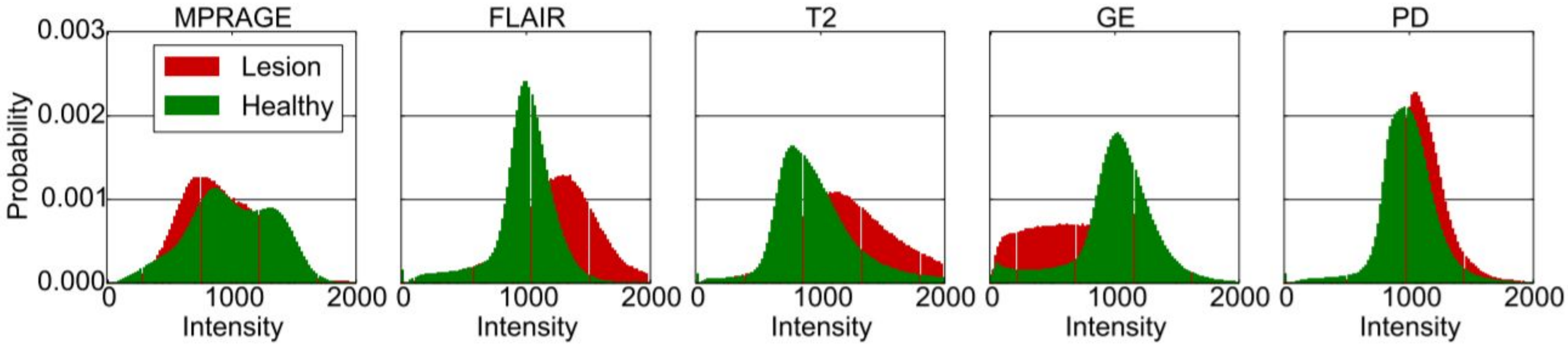
Many small affected areas

Incidence of TBI is rising

Figure 1. Annual age-adjusted rates of TBI-related Emergency Department (ED) visits, hospitalizations, and deaths—United States, 2001–2010



TBI lesions cannot be easily distinguished from healthy tissue



(e)

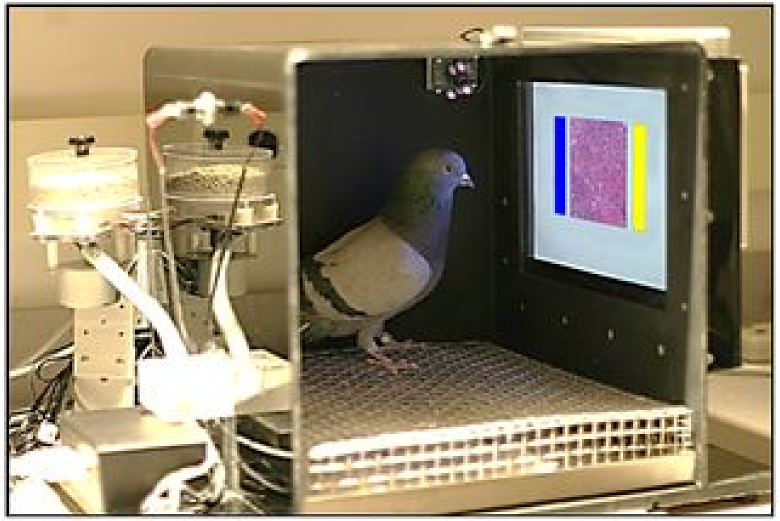
Identifying (or “segmenting”) lesions requires human expertise

- Location of lesion determines which brain functions are affected.
- Size and number of lesions is critical information for prognosis and treatment plan.
- Segmentation is slow, expensive, and introduces observer effects.

Why is this task so challenging?

- Lesions are heterogeneous
- Many tasks require differentiating different types of lesion
  - TBI: contusions, hemorrhages, edema
  - Cancer: proliferating vs. necrotic tissue
- Complex images – three-dimensional, possibly with several channels if multiple sequences are taken

Alternatives to manual segmentation



Trained pigeon, 2015
(yes, this is real)

Network Architecture

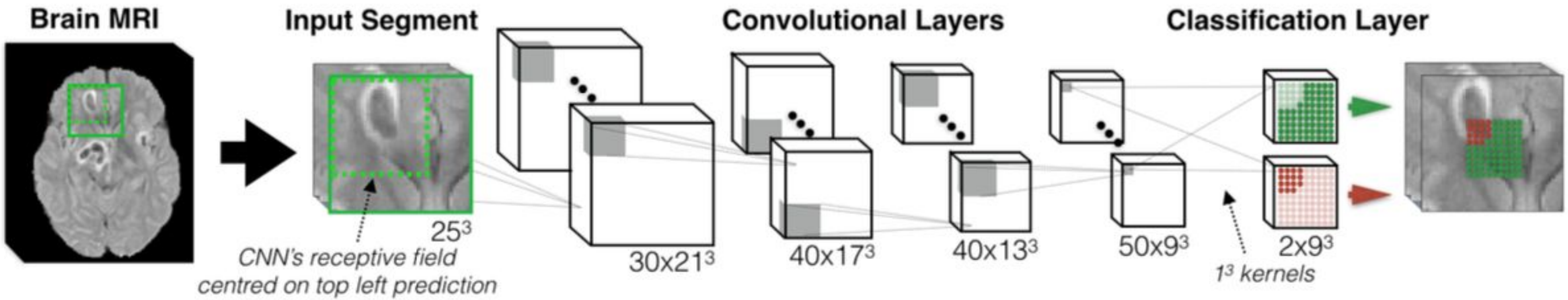
Fully Convolutional Network Model

Feature Map Convolution, m-th FM, l-th layer

$$\mathbf{y}_l^m = f\left(\sum_{n=1}^{C_{l-1}} \mathbf{k}_l^{m,n} \star \mathbf{y}_{l-1}^n + b_l^m\right)$$

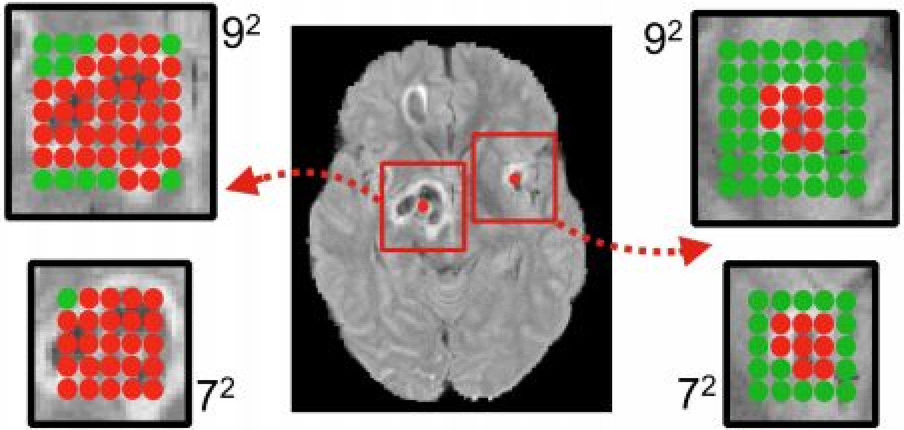
Classify each voxel in an image independently; taking the neighborhood into account (receptive field)

Dense-inference, output multiple predictions simultaneously, one for each stride of the receptive field on the input



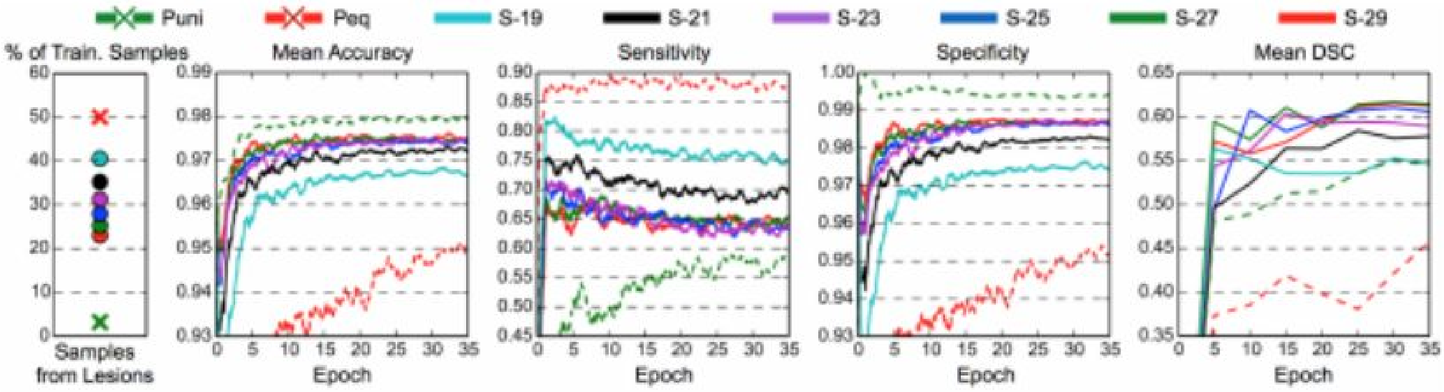
$(Number \times Size)$

Training Efficiency and Class Balance



- Because input patch outputs prediction for multiple voxels, the effective min-batch size is scaled up by the number of voxel prediction
- Training batch selected to alleviate class-imbalance: extracting segments with 50% probability being center on a foreground or background voxel.

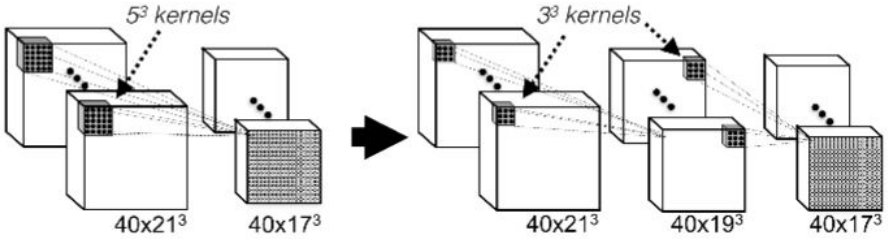
Dense Training Validation



- Two baselines: uniformly sampled / equally sampled from both classes
- Achieve balance of sensitivity / specificity between the two baseline approaches, leading to better overall performance

Building 11-Layer Deep Networks

- Use smaller kernels to improvement speed and prevent overfitting

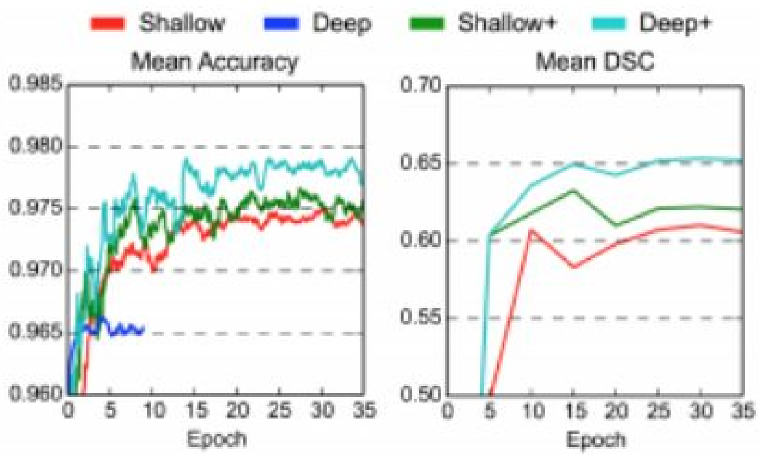


- Use ReLu-based networks to preserve gradient during backpropagation
- Batch Normalization to prevent internal covariate shift

Deep Network Validation

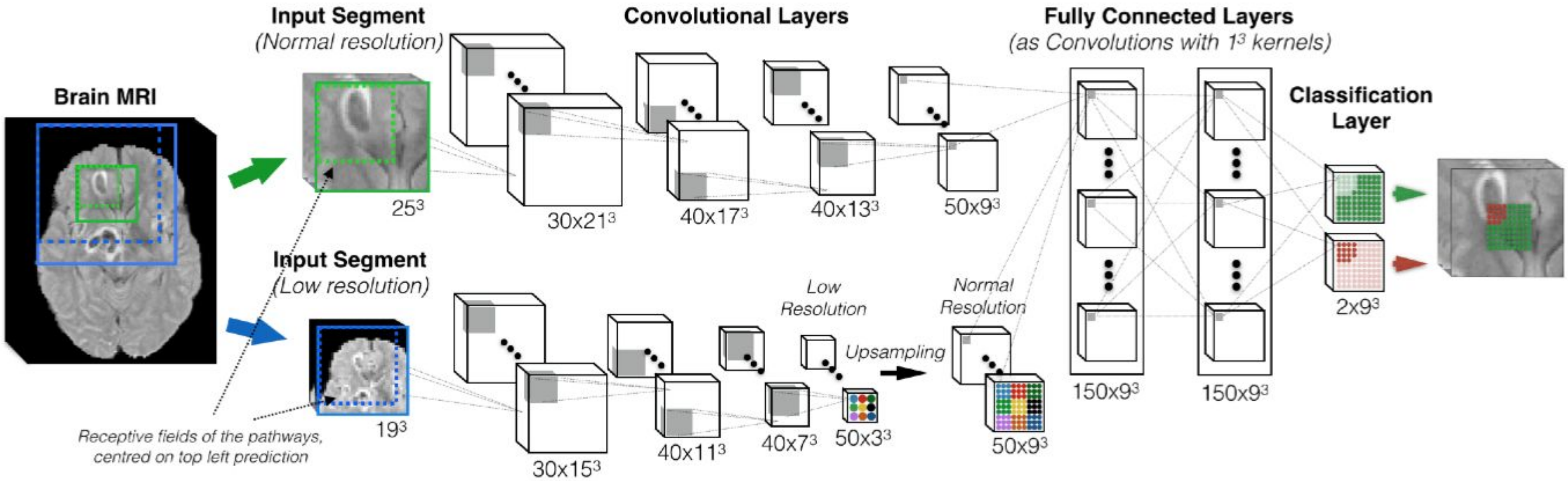
- Better Weight initialization + Batch Normalizations: “Shallow+” and “Deep+” networks.

- Significant performance improvement with the Deep+ network with similar computation times



Parallel Convolutional Pathways

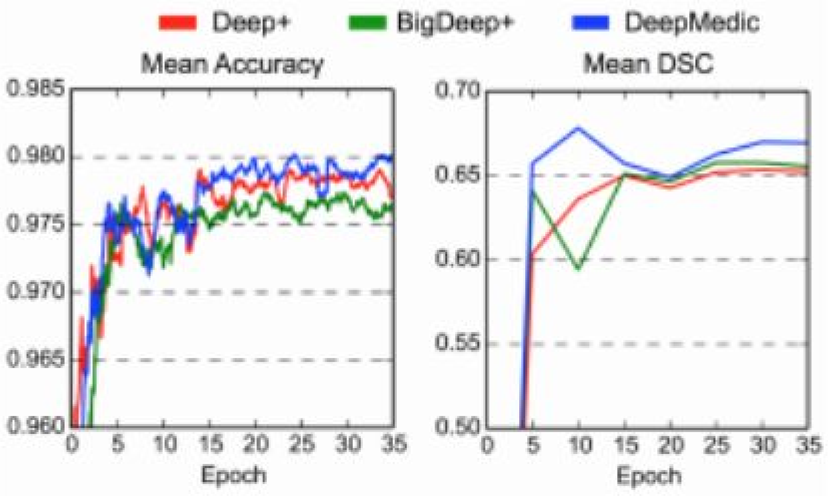
Added a second pathway that operates on down-sampled image to incorporate both local and larger contextual information



Dual Pathway Validation

- Better performance: Captures context in 3X of original image segment

- Only doubles the computational and memory requirements.



Fully Connected 3D CRF For Structure Prediction

Fully Connected CRF as post-processing step in order to ‘clean up’ the CNN results

$$E(\mathbf{z}) = \sum_i \psi_u(z_i) + \sum_{ij, i \neq j} \psi_p(z_i, z_j)$$

$$\psi_u(z_i) = -logP(z_i|\mathbf{I})$$

$$\psi_p(z_i, z_j) = \mu(z_i, z_j)k(\mathbf{f_i}, \mathbf{f_j})$$

$$k(\mathbf{f_i}, \mathbf{f_j}) = \sum_{m=1}^{M} w^{(m)}k^{(m)}(\mathbf{f_i}, \mathbf{f_j})$$

$$k^{(1)}(\mathbf{f_i}, \mathbf{f_j}) = exp\Big(- \sum_{d=\{x,y,z\}} \frac{|p_{i,d} - p_{j,d}|^2}{2\sigma_{\alpha,d}^2} \Big)$$

$$k^{(2)}(\mathbf{f_i}, \mathbf{f_j}) = exp\Big(- \sum_{d=\{x,y,z\}} \frac{|p_{i,d} - p_{j,d}|^2}{2\sigma_{\beta,d}^2} - \sum_{c=1}^{C} \frac{|I_{i,c} - I_{j,c}|^2}{2\sigma_{\gamma,c}^2} \Big)$$

Results

Evaluation on Clinical Data

- Model evaluated on 3 different data sets from open challenges
- Models
  - Vanilla (+CRF)
  - Ensemble (+CRF)
- Data
  - 5-fold cross validation
  - Data augmentation by flipping

Traumatic Brain Injuries

- Data:
  - 61 patients
  - Image resolution of 193x229x193
- Model
  - 2% dropout on conv layers, 50% on FC
  - Used Random Forest with 50 trees of max depth 30 as baseline

Traumatic Brain Injuries - Results

	DSC	Precision	Sensitivity	ASSD	Haussdorf
Random Forest	51.1(20.0)	50.1(24.4)	60.1(15.8)	8.29(6.76)	64.17(15.98)
DeepMedic	62.3(16.4)	65.3(18.8)	64.4(16.3)	4.24(2.64)	56.50(15.88)
DeepMedic+CRF	**63.0(16.3)****	67.7(18.2)	63.2(16.7)	4.02(2.54)	55.68(15.93)
Ensemble	64.2(16.2)	67.7(18.3)	65.3(16.3)	3.88(2.33)	54.38(15.45)
Ensemble+CRF	**64.5(16.3)***	69.8(17.8)	63.9(16.7)	3.72(2.29)	52.38(16.03)

Brain Tumor Segmentation

- Data
  - Training set: 220 patients with “high grade” and 54 patients with “low grade” glioma
  - Data set annotated by segmenting each into 4 classes of tissue
  - Test set 110 patients with unknown grades of glioma
  - Image resolution of 240x240x155
- Model
  - No dropouts since enough samples are provided
  - Compared against best submissions

Brain Tumor Segmentation - Training Results

	DSC			Precision			Sensitivity			
	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Cases
Ensemble+CRF	**90.1***	75.4	**72.8***	91.9	85.7	75.5	89.1	71.7	74.4	274
Ensemble	90.0	75.5	72.8	90.3	85.5	75.4	90.4	71.9	74.3	274
DeepMedic+CRF	**89.8****	75.0	**72.1***	91.5	84.4	75.9	89.1	72.1	72 .5	274
DeepMedic	89.7	75.0	72.0	89.7	84.2	75.6	90.5	72.3	72.5	274
bakas1	88	77	68	90	84	68	89	76	75	186
peres1	87	73	68	89	74	72	86	77	70	274
anon1	84	67	55	90	76	59	82	68	61	274
thirs1	80	66	58	84	71	53	79	66	74	267
peyrj	80	60	57	87	79	59	77	53	60	274

Brain Tumor Segmentation - Test Results

	DSC			Precision			Sensitivity		
	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.
DeepMedic	83.6	67.4	62.9	82.3	84.6	64.0	88.5	61.6	65.6
DeepMedic+CRF	**84.7****	67.0	62.9	85.0	84.8	63.4	87.6	60.7	66.2
Ensemble	84.5	66.7	63.3	83.3	86.1	63.2	88.9	59.9	67.3
Ensemble+CRF	**84.9****	66.7	**63.4***	85.3	86.1	63.4	87.7	60.0	67.4

Ischemic Stroke Lesion Segmentation

- Data
  - Training set: 28 patients
  - Test set: 36 patients
  - Images from two different clinics
  - Image resolution of 240x240x154
- Model
  - No mention of dropouts, used ensemble (+CRF)
  - Compared against best submissions

Ischemic Stroke Lesion Segmentation - Results

	DSC	Precision	Sensitivity	ASSD	Haussdorf
kamnk1(ours)	59(31)	68(33)	60(27)	7.87(12.63)	39.61(30.68)
fengc1	55(30)	64(31)	57(33)	8.13(15.15)	25.02(22.02)
halmh1	47(32)	47(34)	56(33)	14.61(20.17)	46.26(34.81)

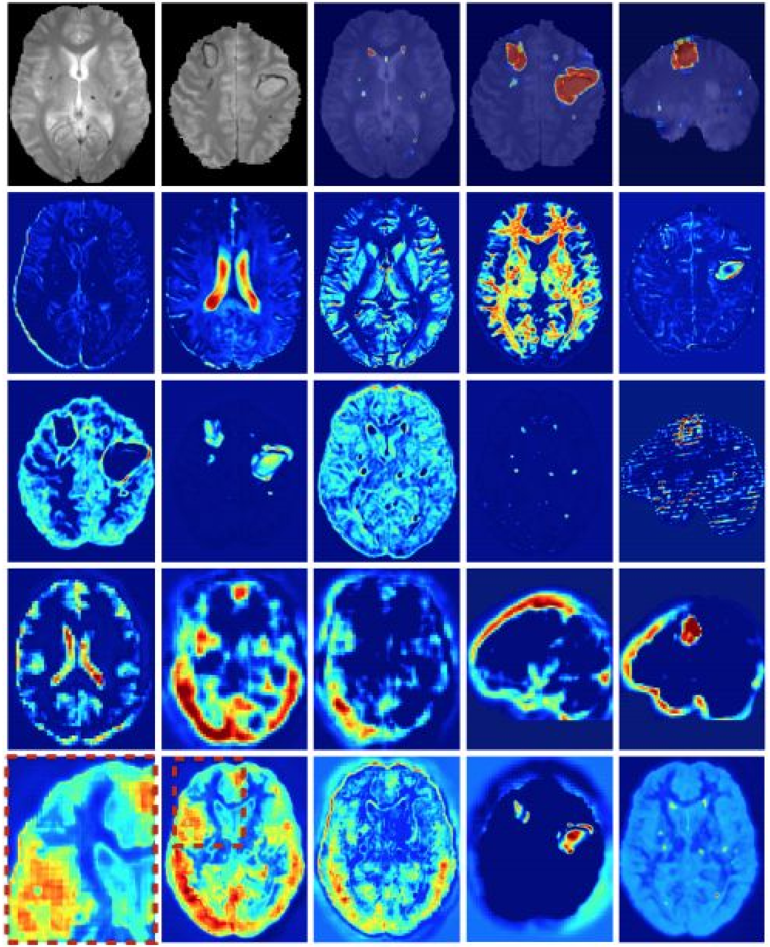
Implementation details

- Coded the CNN in Theano
- Ran on GTX Titan X
- Training took about 2-3 days depending on the dataset
- CRF is cpu code only. Extended from the work of Krahenbuhl and Koltun
- All source code available online

Conclusion

Conclusion

- Better Segmentation with high computational efficiency
- Proposed a solution to class imbalance in training data
- First application of CRF to 3d medical data
- Generic network capable of being repurposed for other similar tasks



Future Work

- Network alternatives:
  - CNN + RNN
  - Residual NN

- More hyperparameter tuning

- Sensitivity testing with dual channel architecture