Brain Tumor Segmentation with Deep Neural Networks

Christian Arndt

Original authors:
Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, Hugo Larochelle
Medical Image Analysis, Volume 35, 1 January 2017, Pages 18-31

Content

Introduction

Clinical Background:

- There are > 20.000 new brain cancer cases p.a. in the US.
- MRI as most common modality for tumor diagnosis.
- Often difficult to localize (gliomas and glioblastomas).

Why Deep Learning?

- Voxel values in MR images are not standardized.
- *Learning* features with CNNs (as opposed to classical machine learning).
- Improved diagnostics, growth rate prediction and treatment planning.

Convolutional Neural Networks (CNNs)

- Deep Neural Networks adapted to image data.
- Stanford University class: http://cs231n.github.io/

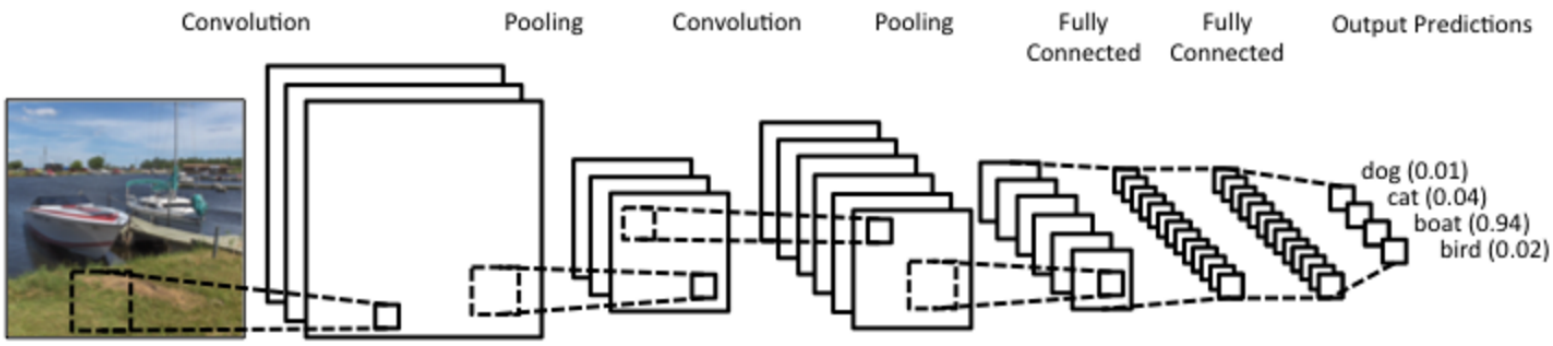

Figure: Schematic overview of a CNN architecture.
(Image from https://medium.com/@Aj.Cheng/convolutional-neural-network-d9f69e473feb, 12.05.2018)

Convolution (in images)

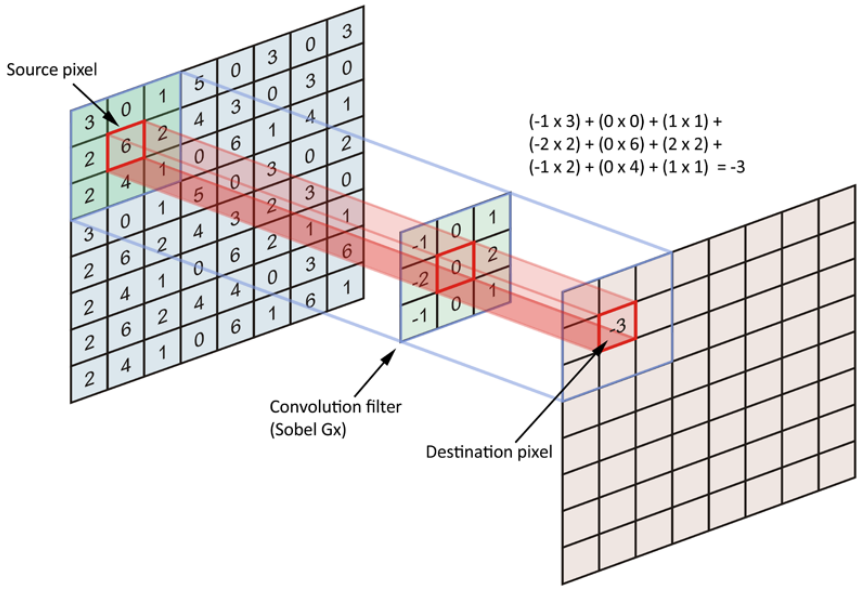

Figure: 2D-convolution: weighted sum over a local patch of data.
(Image from https://i.stack.imgur.com/YDusp.png, 12.05.2018)

Architecture

- BraTS 2013 dataset.
- Slice by slice segmentation due to bad depth resolution.
- Different image modalities of the same structures.

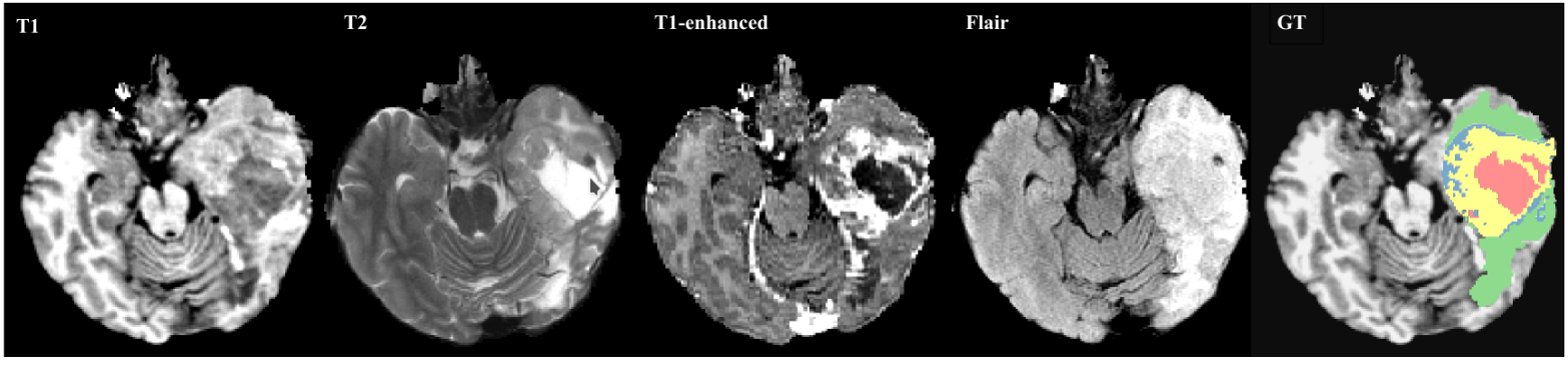

Figure: MRI sequences used as input channels and ground truth labels (p. 9, figure 4).

Architecture

Two-pathway architecture:

- One smaller (7×7) and one larger (13×13) receptive field.
- Prediction based on local region and larger context.

Cascaded architectures:

- Goal: model dependencies between spatially close labels by
- *input concatenation*,
- *local pathway concatenation*, and
- *pre-output concatenation*.

Two-pathway architecture

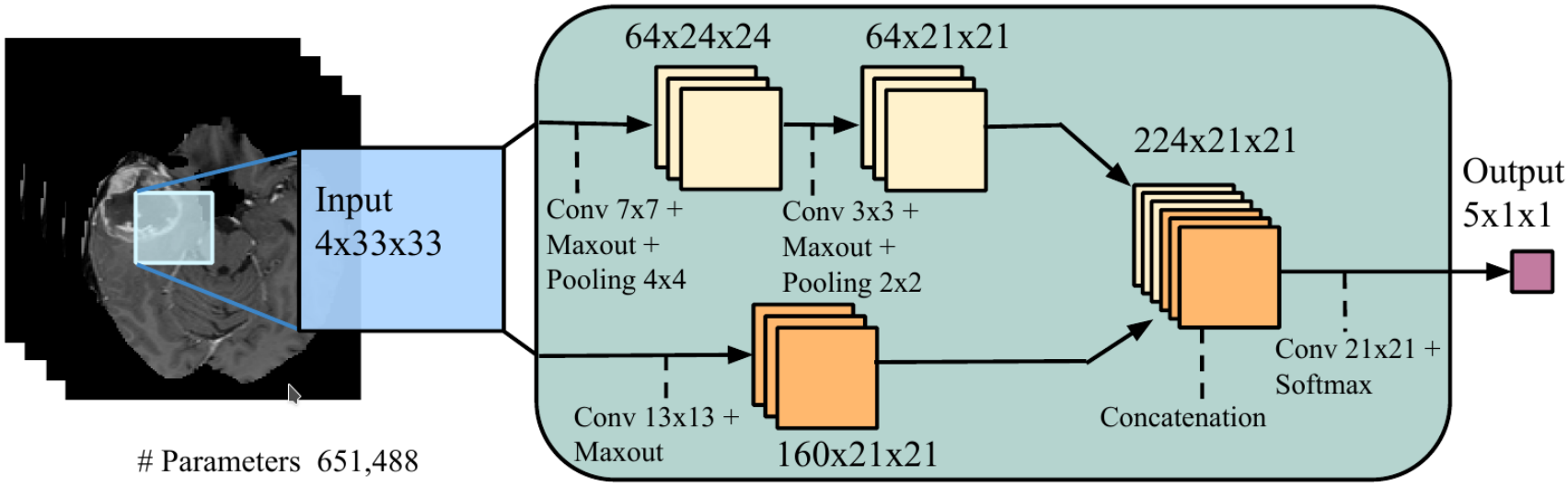

Figure: Two-pathway CNN architecture (p. 6, figure 2).

Cascaded architecture

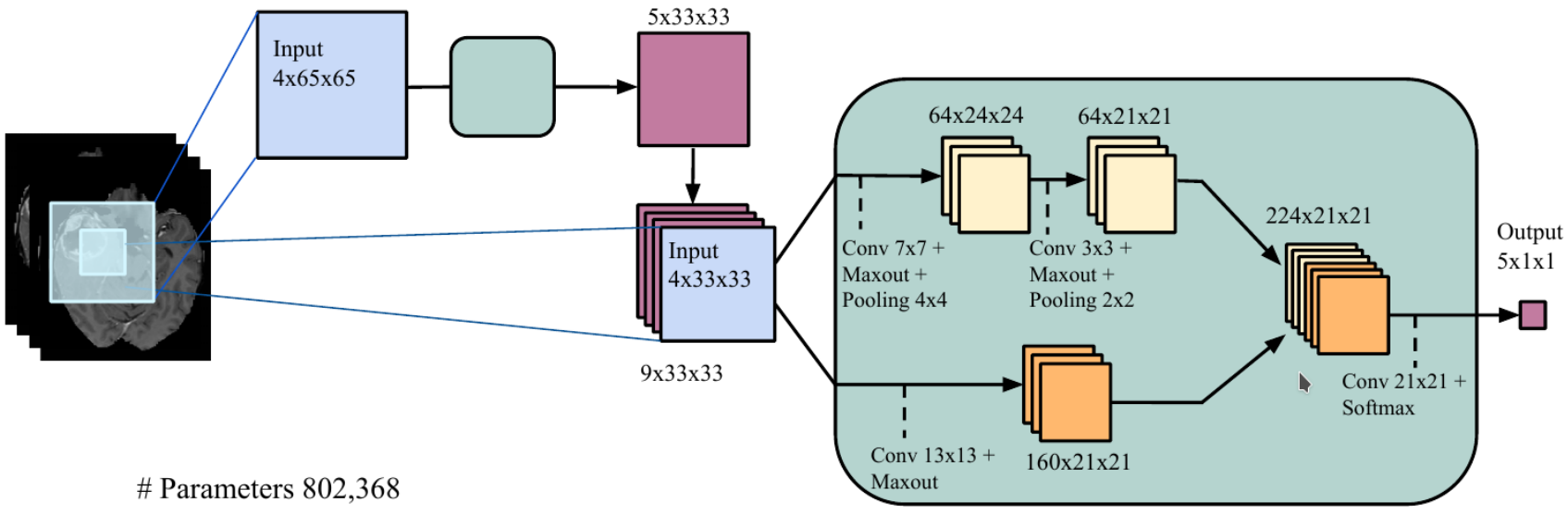

Figure: Cascaded architecture using input concatenation to combine two CNNs (p. 7, figure 3).

Training

Gradient Descent:

- Forward propagation on a mini-batch of patches.
- Compute label probabilities and deviation from ground truth.
- Update the CNNs parameters.

Two-phase training:

- Highly imbalanced data (98% healthy voxels).
- First: Pick patches such that all labels are equiprobable.
- Then: Re-train output layer with original distribution.

Regularization:

- Prevent overfitting by bounding kernel weights and modifying output probabilities.

Two-pathway architecture

- Second phase and joint training of local and global path yields better performance.
- Very fast, about 25s for a whole brain.

Table: Quantitative results of the two-pathway architecture variations on the BRATS 2013 dataset, where the appended * denotes two-phase training (p. 11, table 1).

Rank	Method	Dice	Specifity	Sensitivity
4	TwoPathCNN*	0.85	0.93	0.80
9	LocalPathCNN*	0.85	0.91	0.80
10	AverageCNN*	0.84	0.95	0.77
14	GlobalPathCNN*	0.82	0.93	0.75
14	TwoPathCNN	0.78	0.67	0.96
15	LocalPathCNN	0.77	0.65	0.96

Cascaded architecture

- Fast, about 3 minutes for a whole brain.
- Winner of the challenge takes about 100 minutes.

Table: Quantitative results of the cascaded architecture variations on the BRATS 2013 dataset, where the appended * denotes two-phase training (p. 13, table 2).

Rank	Method	Dice	Specifity	Sensitivity
2	InputCascadeCNN*	0.88	0.89	0.87
4-a	MFCascadeCNN*	0.86	0.92	0.81
4-b	LocalCascadeCNN*	0.88	0.91	0.84

Cascaded architecture

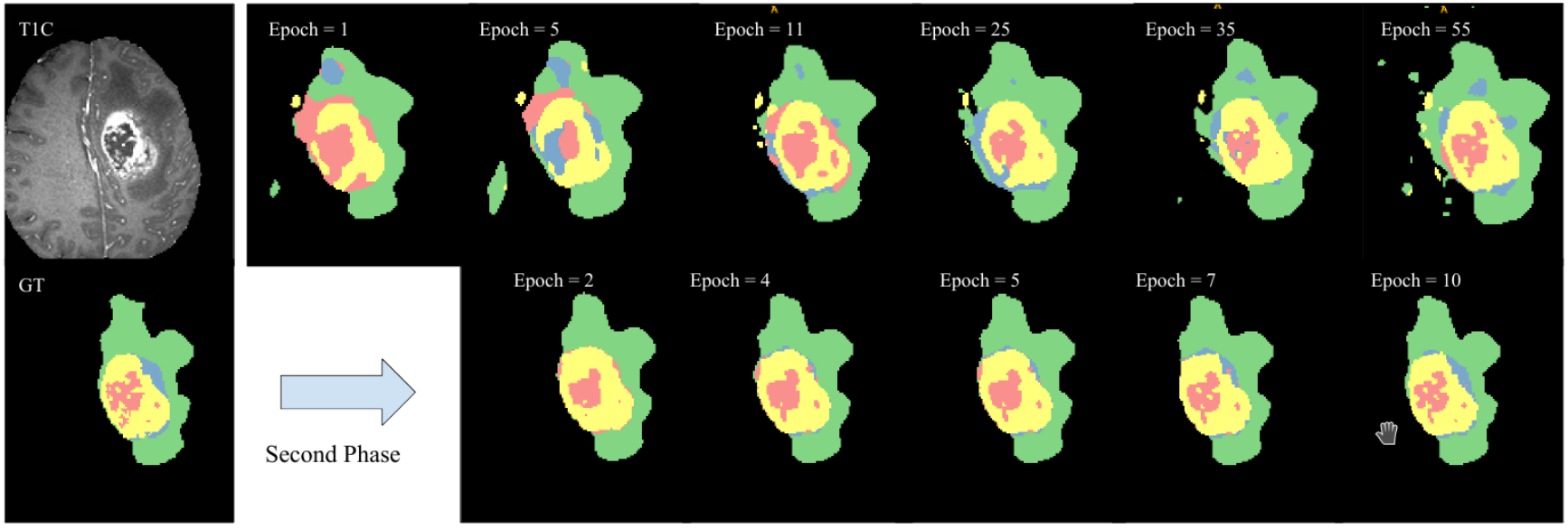

Figure: Progression of learning InputCascadeCNN* (p. 11, figure 6).

Conclusion

- Automatic brain tumor segmentation based on deep learning (Convolutional Neural Networks).

Improve on currently published state-of-the-art methods

- Improvements especially in speed (25s to 3m per brain).

Novel two-pathway architecture

- Fuse local details and global context.
- Model local dependencies of labels.

Thank you for your attention!

Any questions?

Additional information

Imbalanced data

- Assume a predictor trained on 10 malignant and 90 benign tumors. A model could predict "benign" for all samples and still gain a very high accuracy. An unbalanced dataset will bias the prediction model towards the more common class!

Gradient Descent

- Maximize the probability of all labels in the training set or, equivalently, minimize the negative log-probability for the label $\mathbf{Y}$ given the data $\mathbf{X}$:

$$-\log p(\mathbf{Y}|\mathbf{X}) = \sum_{ij} -\log p(Y_{ij}, \mathbf{X}) \tag{1}$$

Additional information

Quantitative Measurements

$$Dice(P,T) = \frac{2|P_1 \cap T_1|}{|P_1| + |T_1|} \tag{2}$$

$$Sensitivity(P,T) = \frac{|P_1 \cap T_1|}{|T_1|} \tag{3}$$

$$Specificity(P,T) = \frac{|P_0 \cap T_0|}{|T_0|} \tag{4}$$

- P - model predictions
- T - ground truth
- Index 1 for positives and index 0 for negatives for the tumor region in question