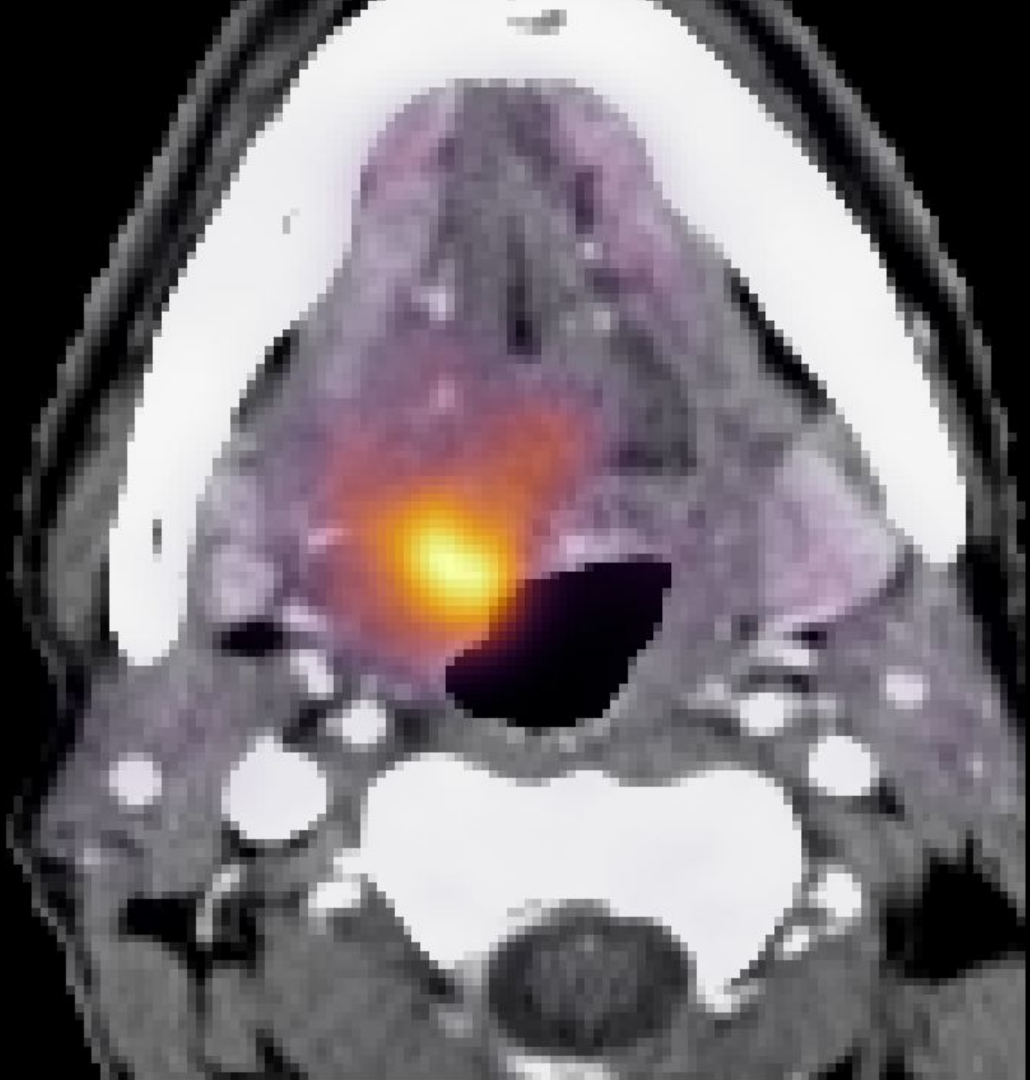
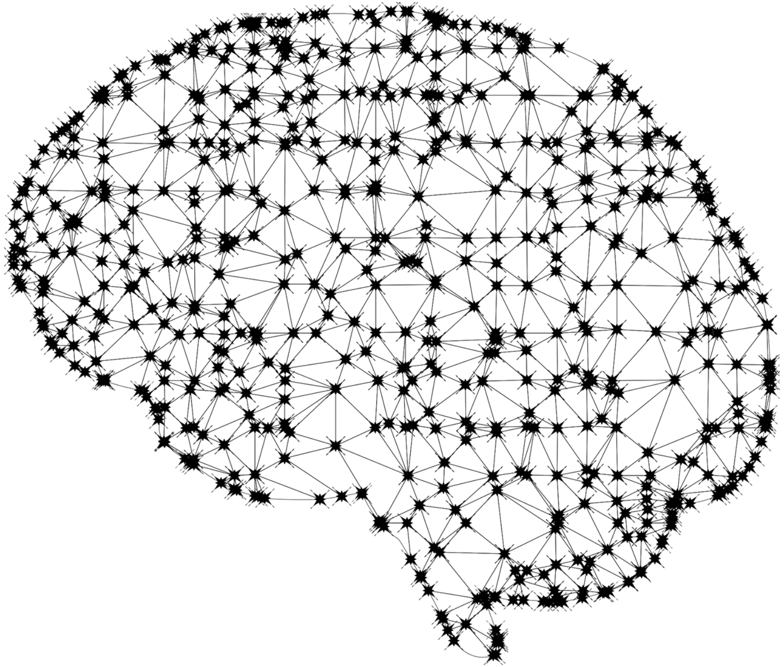


Deep learning for automatic delineation of tumours from PET/CT images



Yngve Mardal Moe

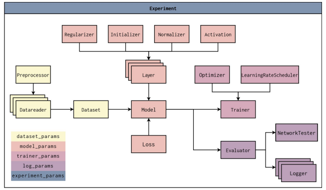
Norwegian University of Life Sciences

5th April 2019

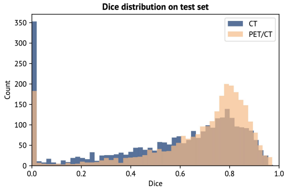
Norges miljø- og biovitenskapelige universitet

This presentation focuses on the work of my master’s thesis

Motivation for automatic tumour delineation



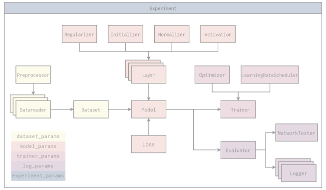
Experimental setup and scientific software



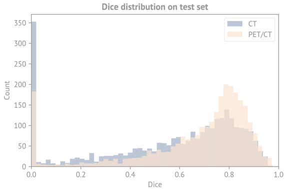
Discussion of the results

This presentation focuses on the work of my master's thesis

Motivation for automatic tumour delineation

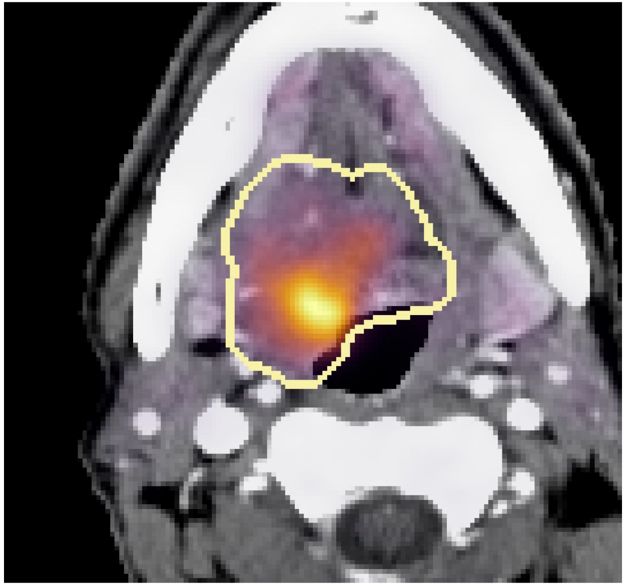


Experimental setup and scientific software



Discussion of the results

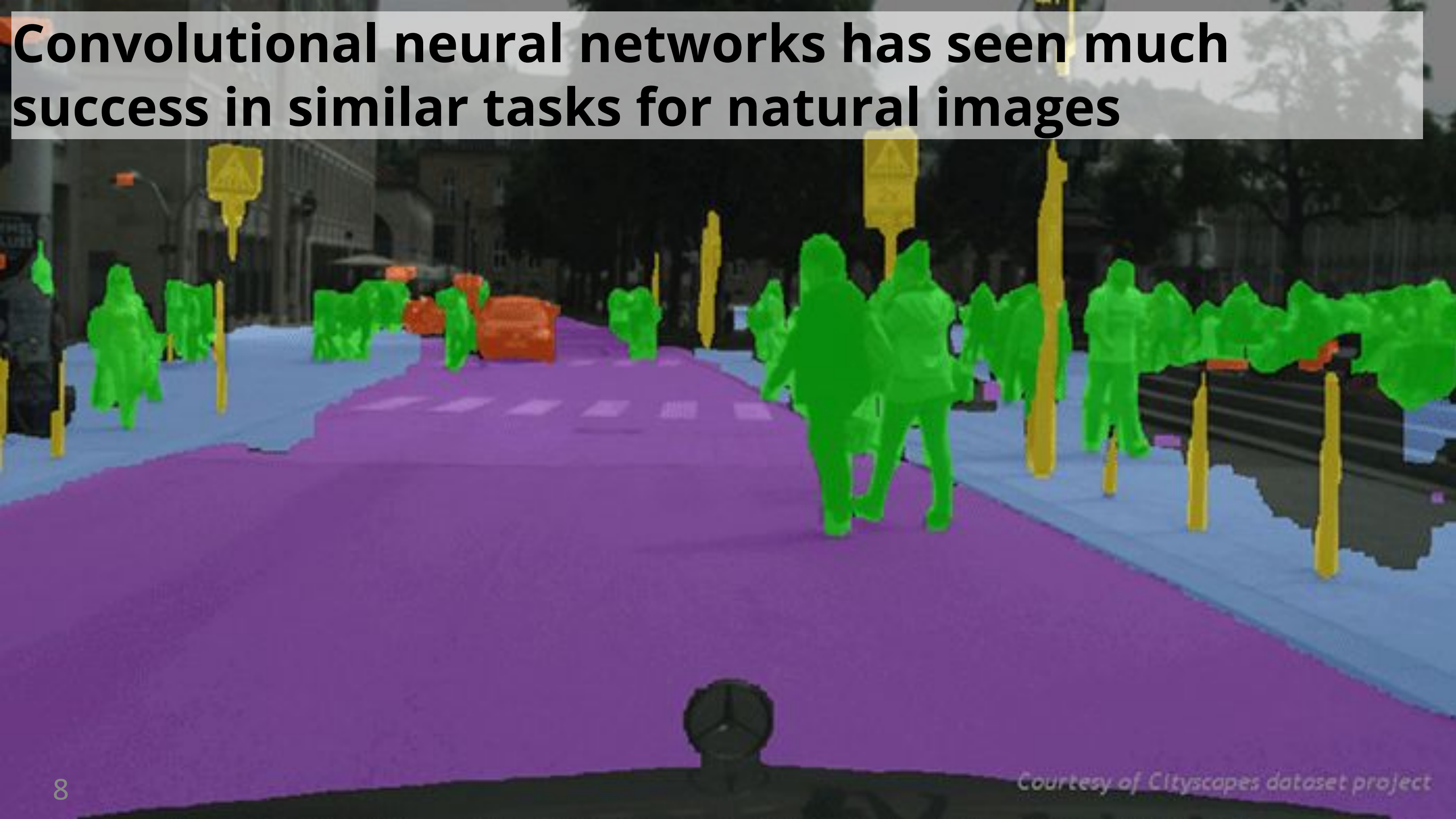
Tumour delineation is an essential and time-consuming part of radiotherapy



Tumour outline

Finding a way to automate this process will reduce the cost and time associated with radiotherapy

Convolutional neural networks has seen much success in similar tasks for natural images



There are, however, few tools that allow for rapid model prototyping as well as ensured reproducibility

☐ Ensured reproducibility

☐ Rapid prototyping

There are, however, few tools that allow for rapid model prototyping as well as ensured reproducibility

TensorFlow

- ☐ Ensured reproducibility
- ☐ Rapid prototyping

PyTorch

- ☐ Ensured reproducibility
- ☐ Rapid prototyping

There are, however, few tools that allow for rapid model prototyping as well as ensured reproducibility

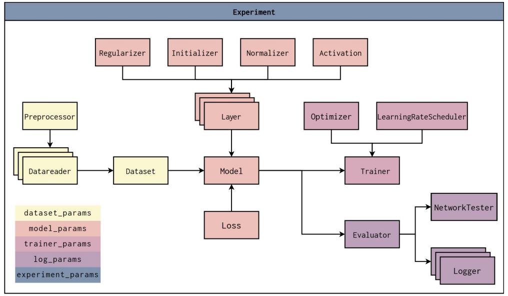
Ensured reproducibility	Rapid prototyping	TensorFlow
Ensured reproducibility	Rapid prototyping	PyTorch
Ensured reproducibility	Rapid prototyping	Keras

One main goal of this thesis was therefore to develop a Python tool that does this

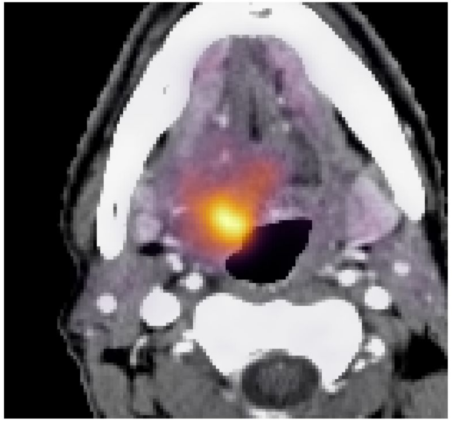
☒ **Ensured reproducibility**

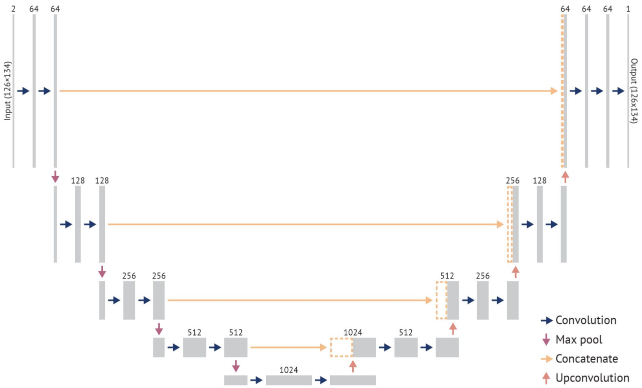
☒ **Rapid prototyping**

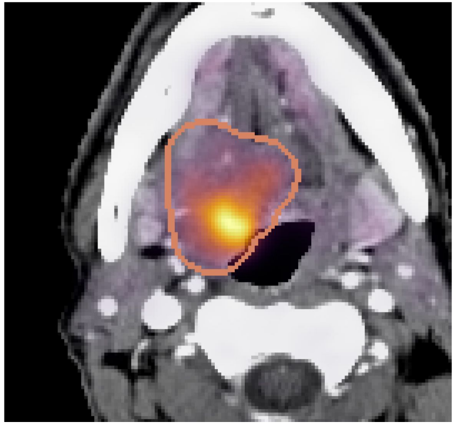
SciNets



The next goal of this thesis was to apply SciNets to assess automatic tumour delineation methods

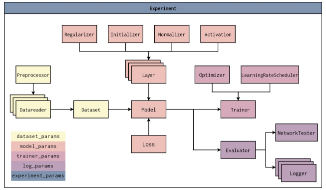




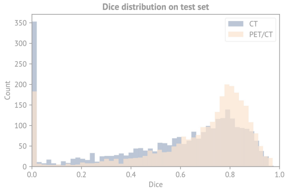


This presentation focuses on the work of my master's thesis

Motivation for automatic tumour delineation

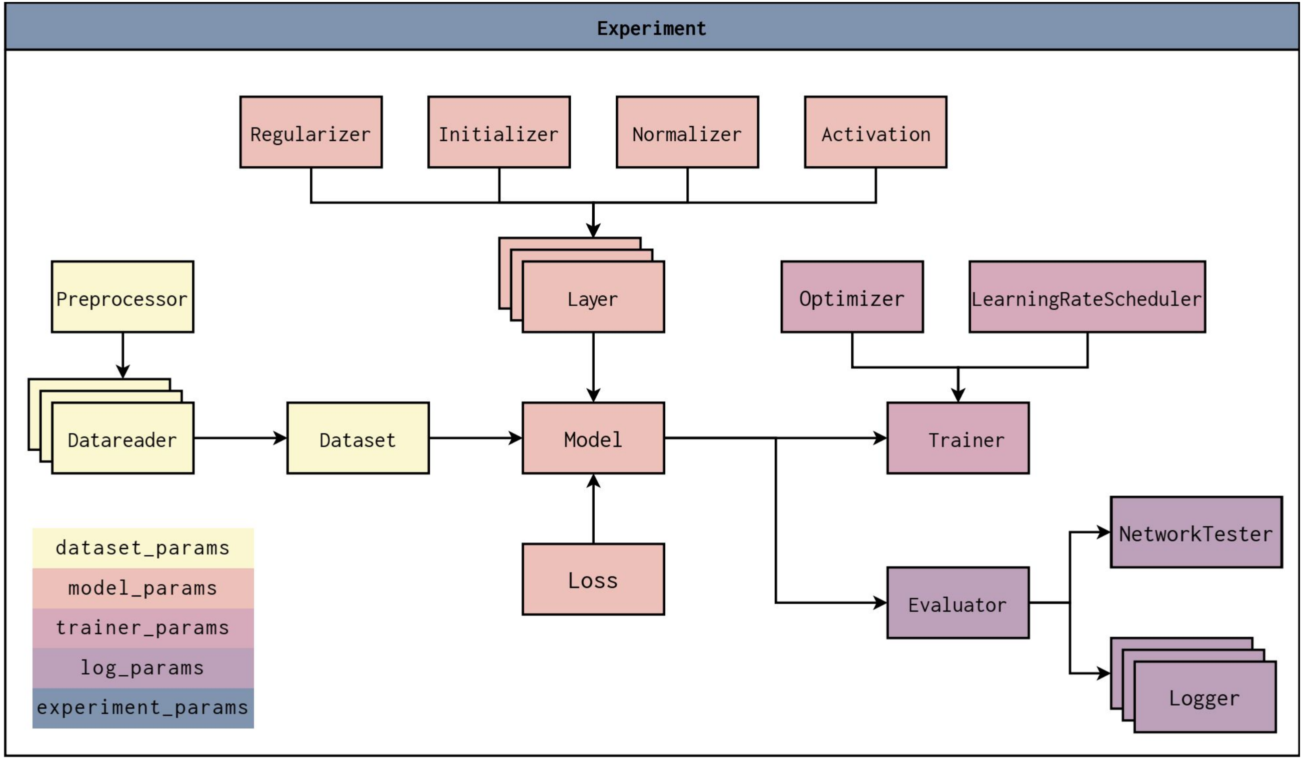


Experimental setup and scientific software

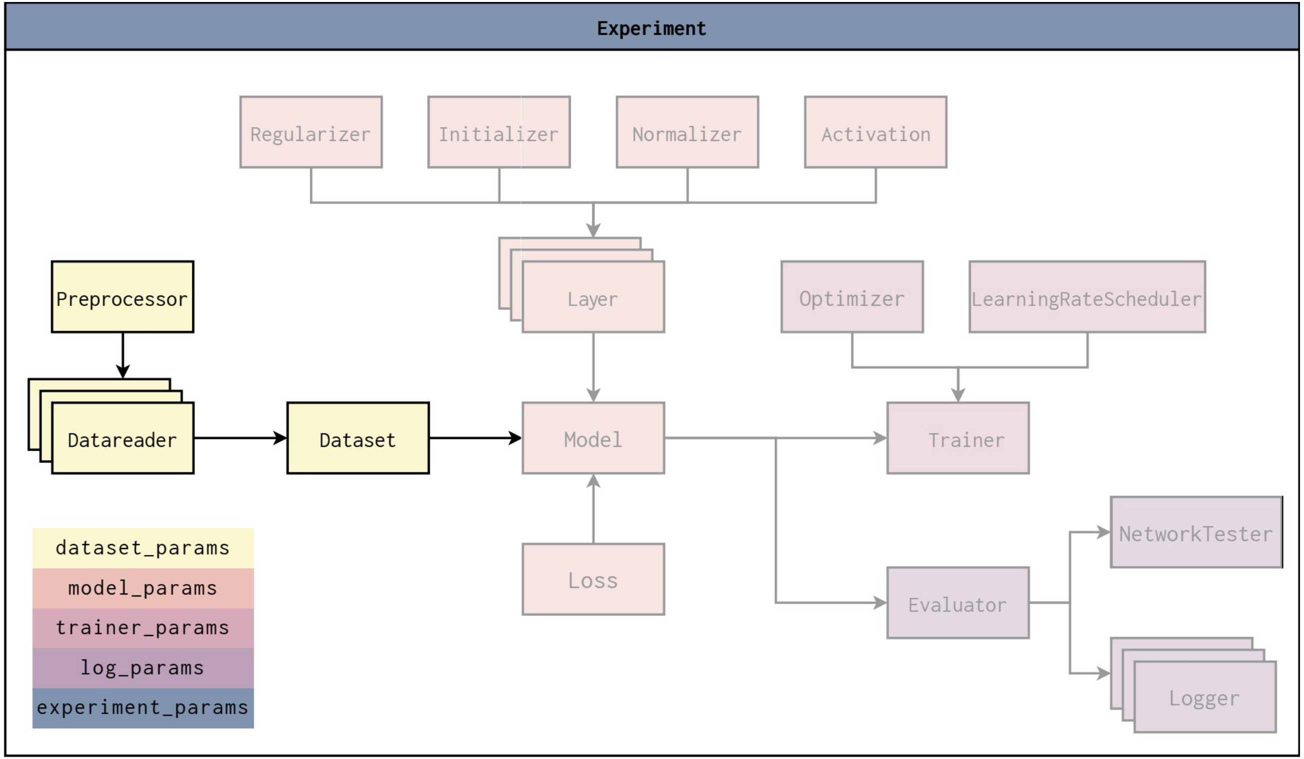


Discussion of the results

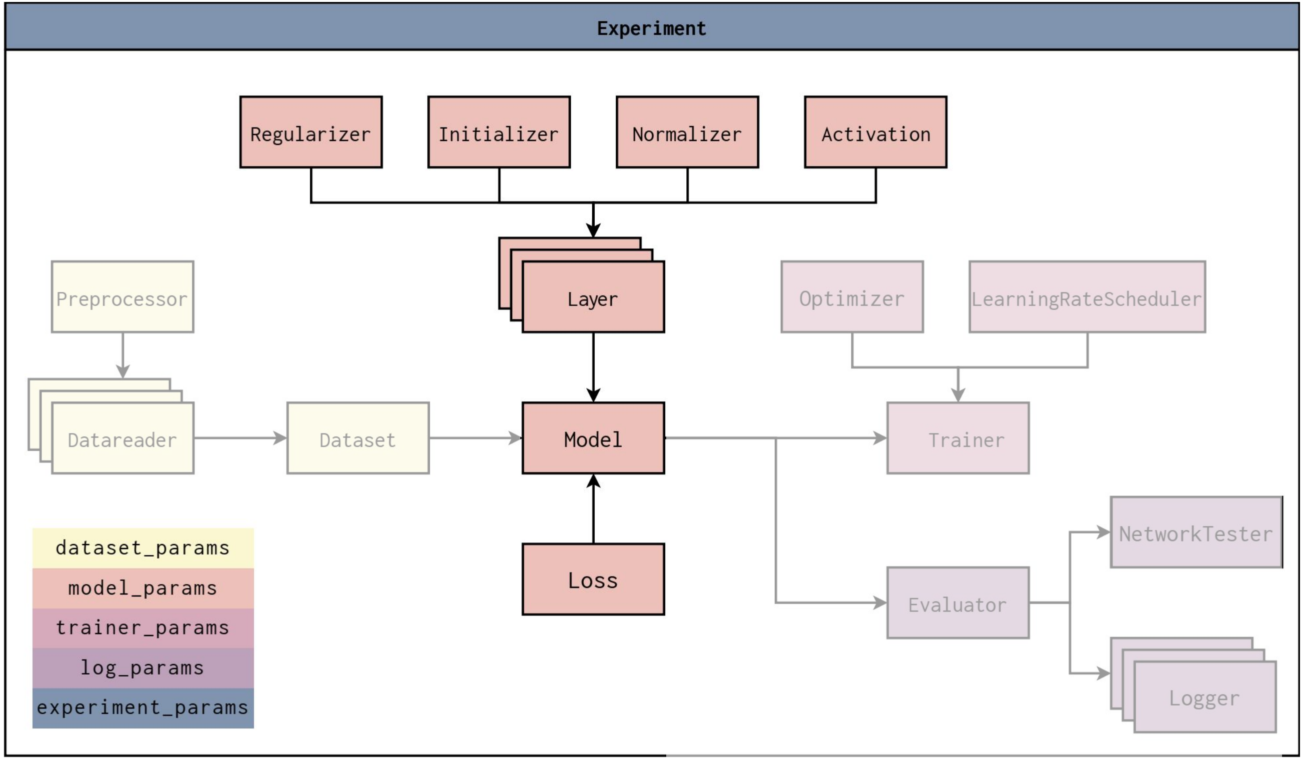
SciNets is an object-oriented tool to seamlessly create and log deep learning experiments



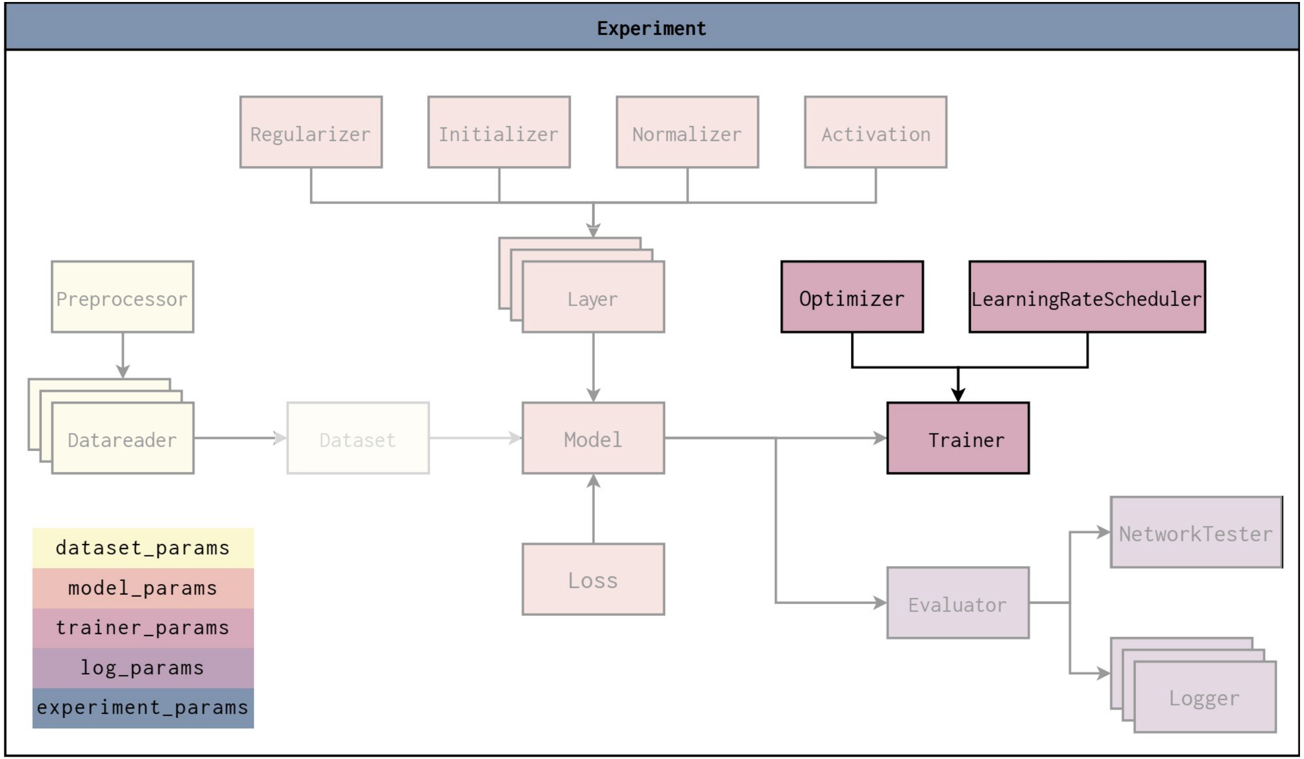
SciNets is an object-oriented tool to seamlessly create and log deep learning experiments



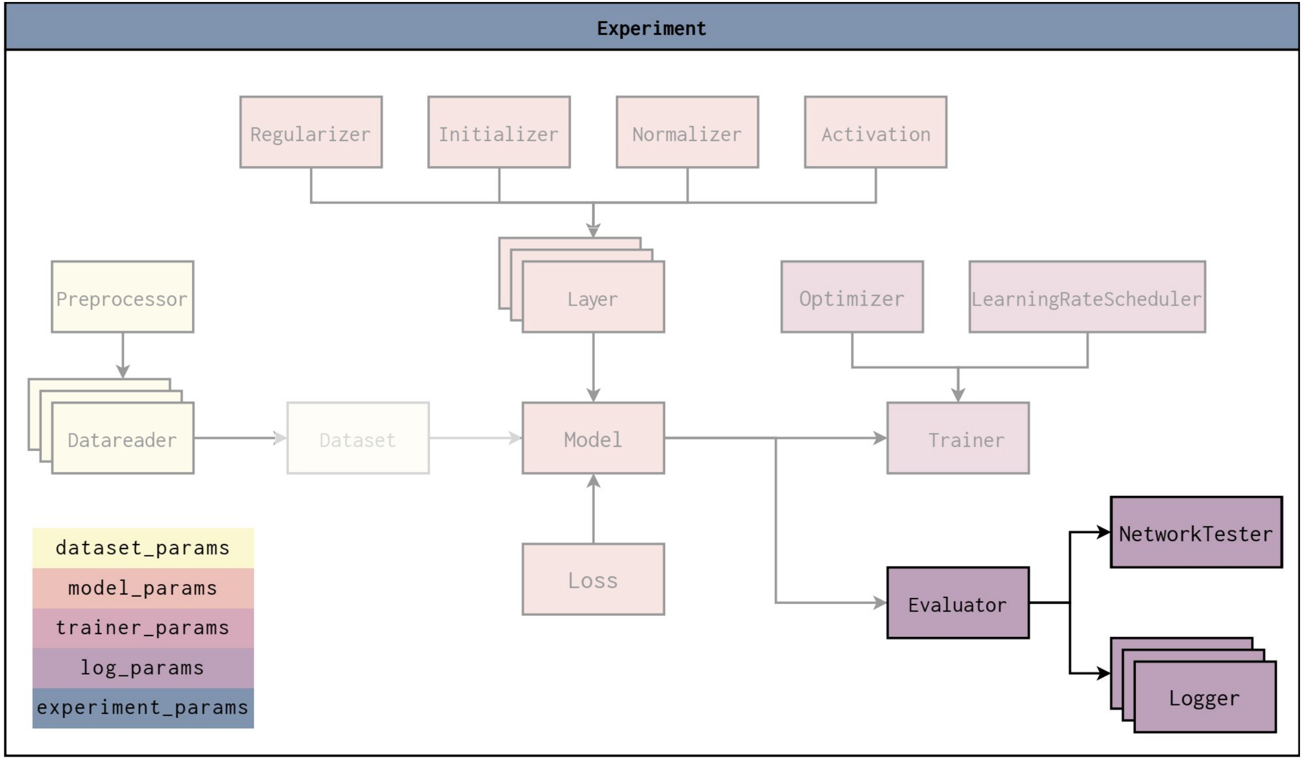
SciNets is an object-oriented tool to seamlessly create and log deep learning experiments



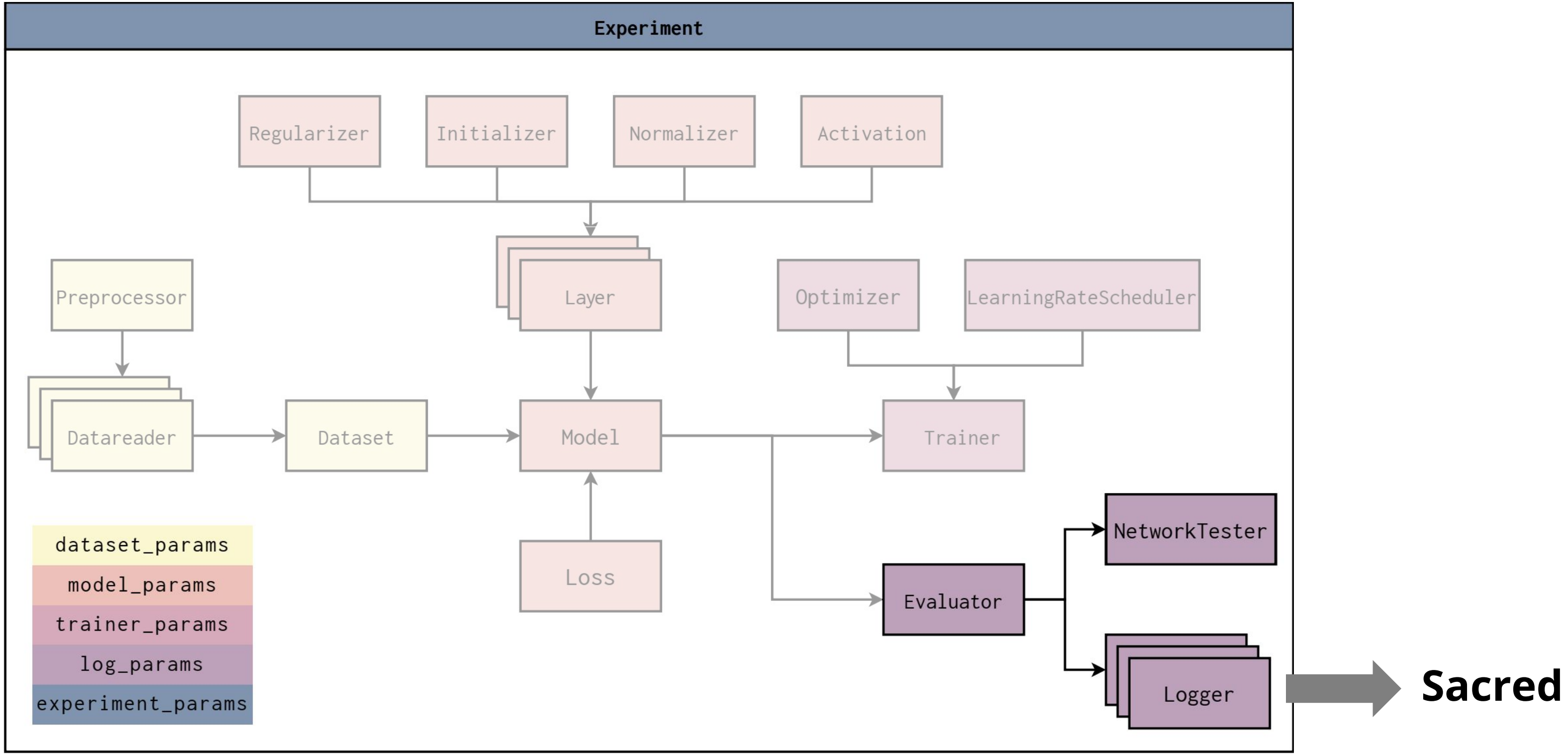
SciNets is an object-oriented tool to seamlessly create and log deep learning experiments



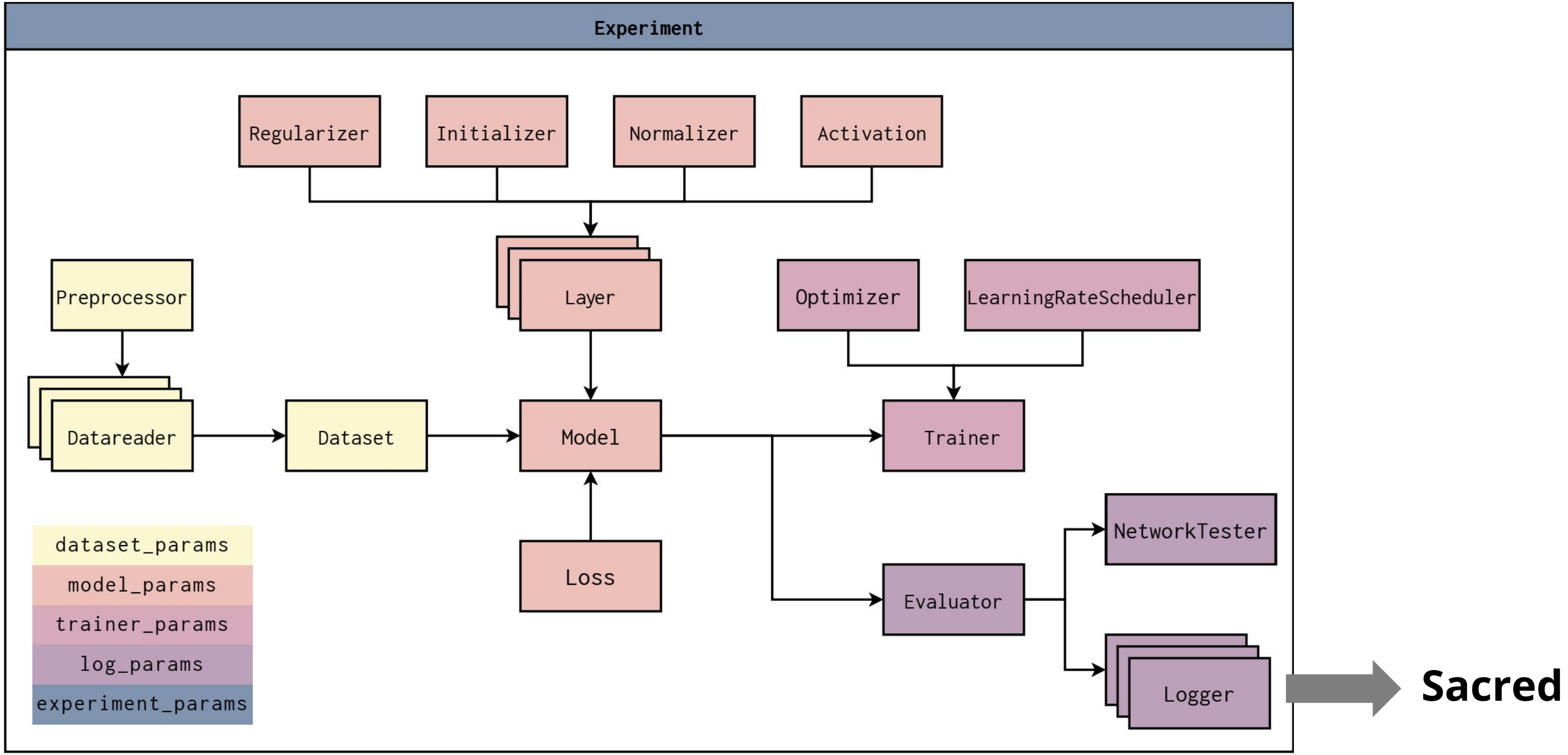
SciNets is an object-oriented tool to seamlessly create and log deep learning experiments



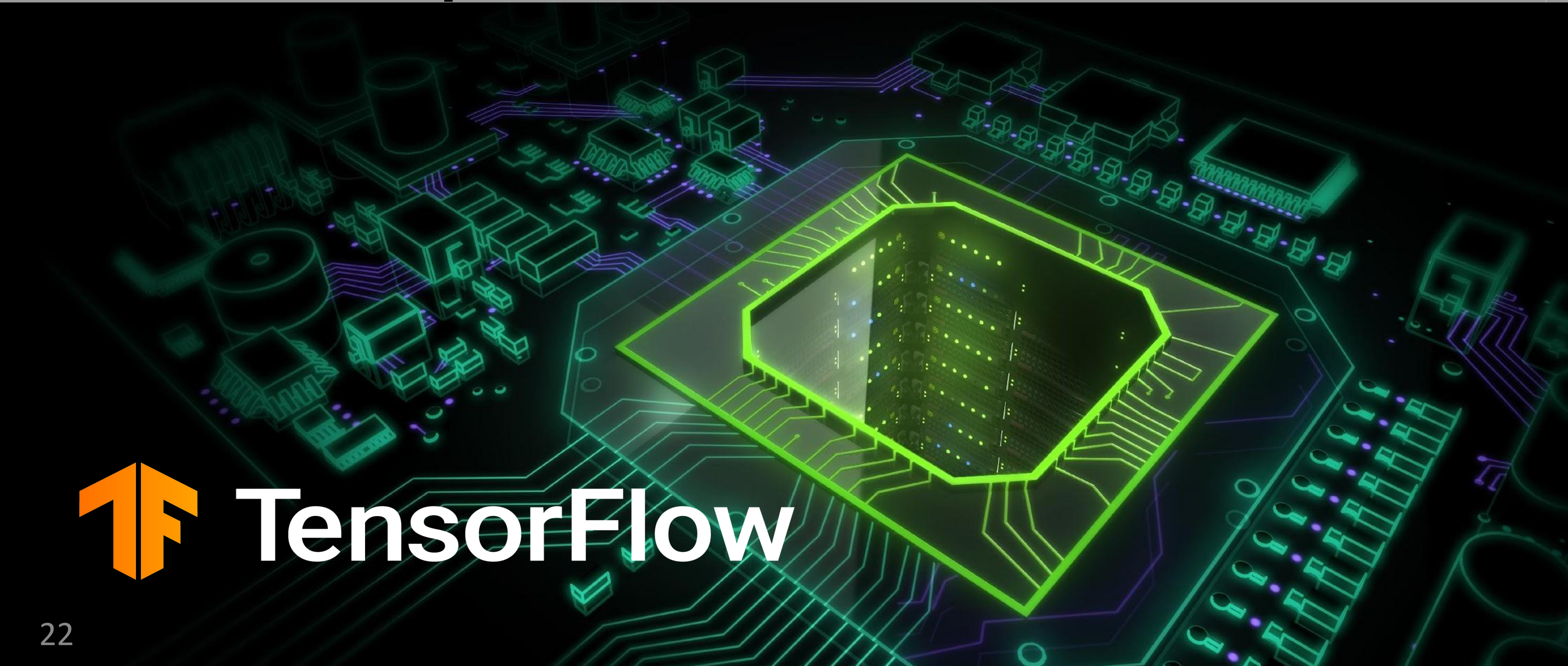
SciNets is an object-oriented tool to seamlessly create and log deep learning experiments



SciNets is an object-oriented tool to seamlessly create and log deep learning experiments

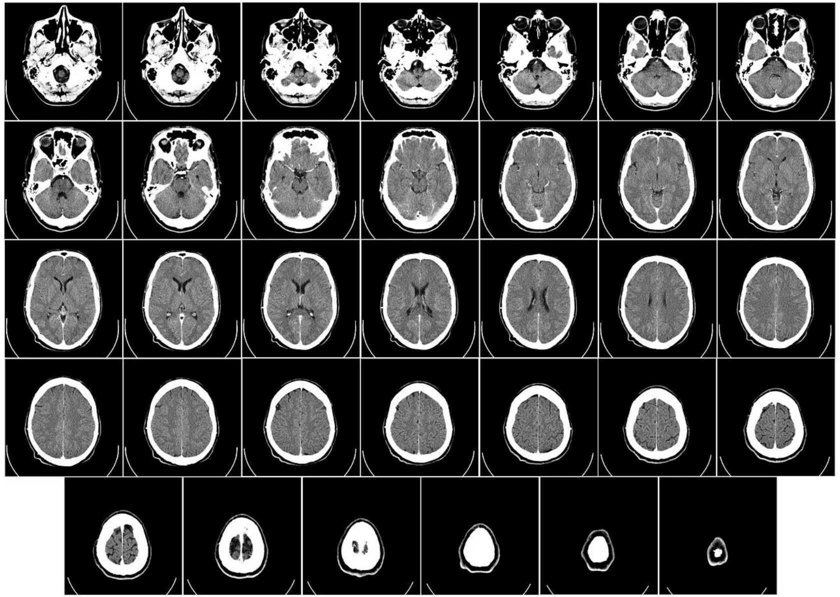


The backend of SciNets is based on TensorFlow 1.1X, allowing for highly performant training using a multi-GPU setup



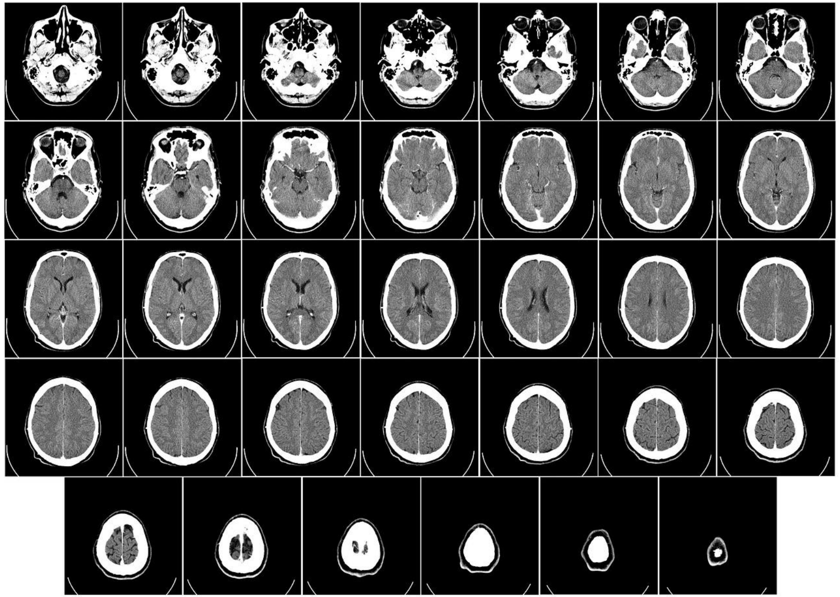
The data was split in three datasets to ensure generalisable results

The three-dimensional medical images were interpreted as a stack of two-dimensional images



Here showing a CT scan of the head

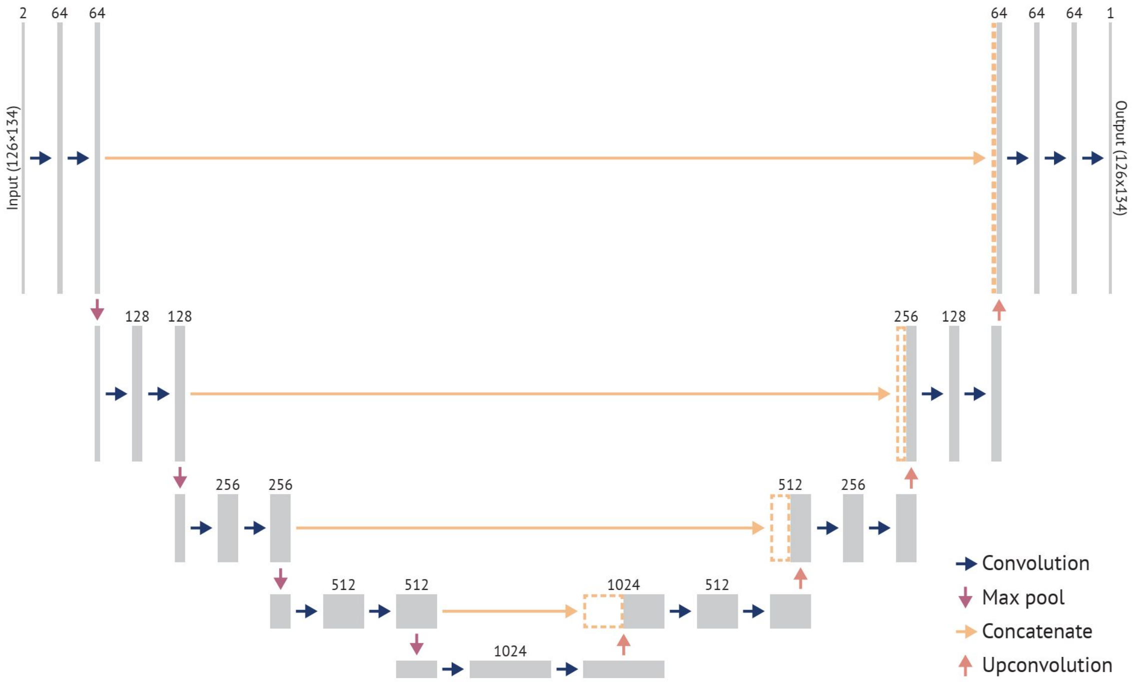
The three-dimensional medical images were interpreted as a stack of two-dimensional images



Here showing a CT scan of the head

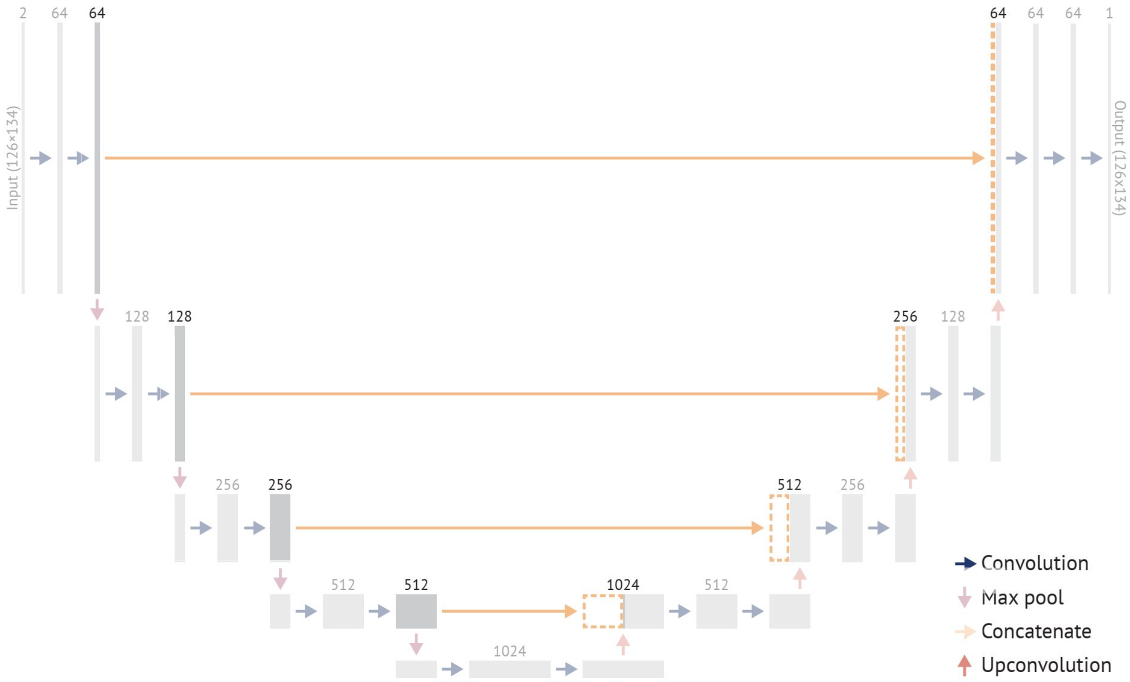
Thus, vertical cues to locate the tumour were ignored

The neural network was built using a U-Net [1] like architecture



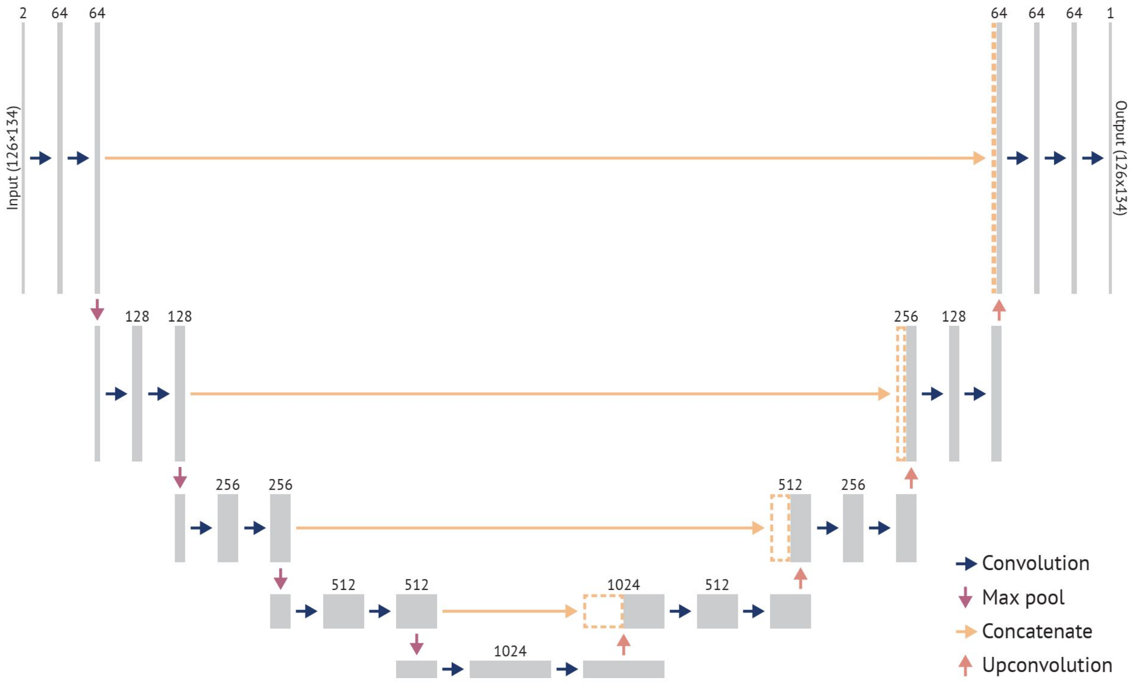
[1] O. Ronneberger, P. Fischer and T. Brox, ‘U-Net: Convolutional networks for biomedical image segmentation’

The long-distance skip connections of U-Net like architectures recovers high-frequency information



[1] O. Ronneberger, P. Fischer and T. Brox, ‘U-Net: Convolutional networks for biomedical image segmentation’

The U-Net architecture has been applied to brain tumour delineation and achieved SOTA results [2]



[1]: H. Dong, et al. ‘Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks’

Combinations of these parameter settings were assessed using a grid-search

Table 1.4: Overview of the hyperparameters used for the U-Net architecture.

Hyperparameter	Value(s)
Learning rate	[0.0001, 0.00001]
Optimiser	Adam
Nonlinearity	ReLU
Normalizer	Batch Normalisation
Initializer	Normally distributed He
Layer type	[Convolutional, Improved ResNet]
Loss	[Cross entropy, F_1, F_2, F_4]
Window centre	[60 HU, 70 HU]
Window width	[100 HU, 200 HU]
Batch size	16
Number of iterations	10000 – 30000
Iterations between checkpoints	2000

We used the Dice coefficient to measure the accuracy of the segmentation algorithms

Another interpretation of the dice score is the harmonic average of the sensitivity and precision

$$F_1 = \frac{2}{\frac{1}{S} + \frac{1}{P}}$$

Sensitivity

Another interpretation of the dice score is the harmonic average of the sensitivity and precision

$$F_1 = \frac{2}{\frac{1}{S} + \frac{1}{P}}$$

Sensitivity

Precision

Another interpretation of the dice score is the harmonic average of the sensitivity and precision

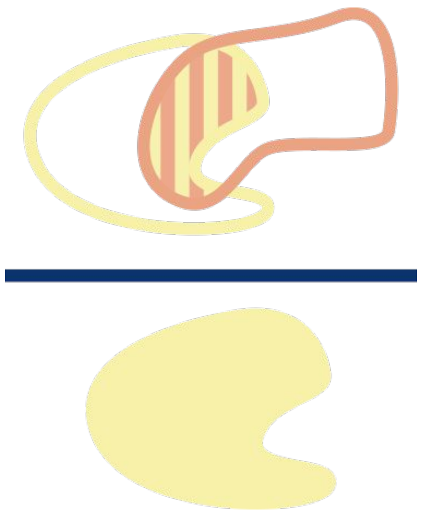
$$F_1 = \frac{2}{\frac{1}{S} + \frac{1}{P}}$$

Sensitivity

Precision

The sensitivity is the fraction of true positives that were found

Sensitivity



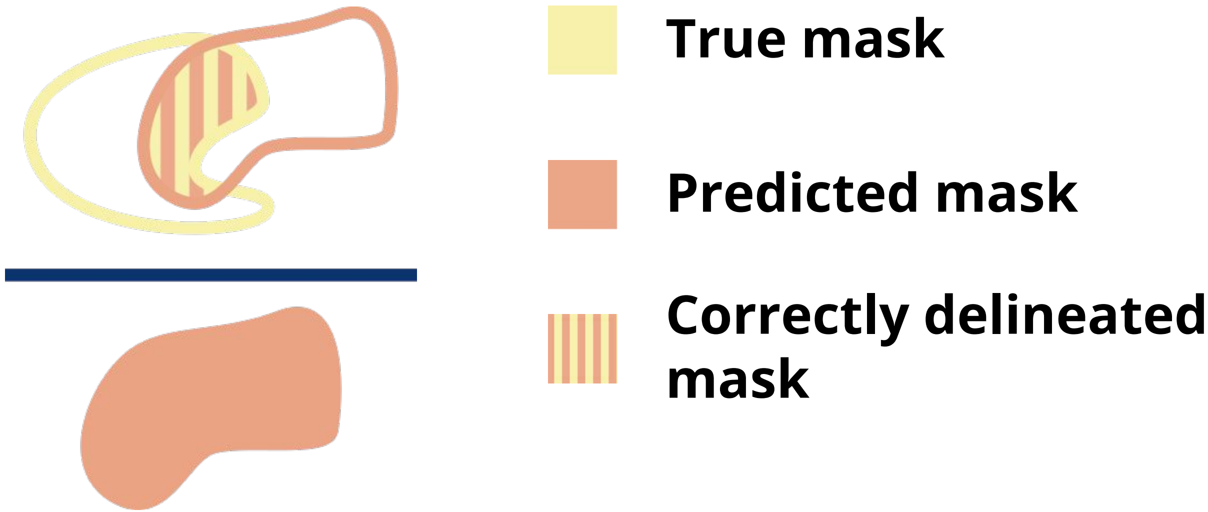
True mask

Predicted mask

Correctly delineated mask

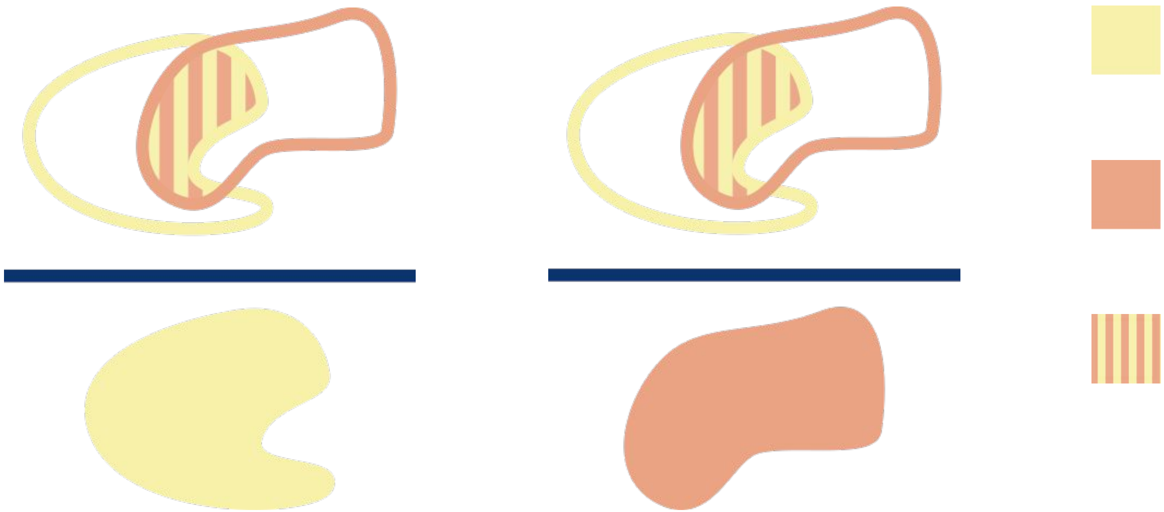
The precision is the probability of a predicted positive being a true positive

Precision



Thus, we have the following description of the sensitivity and precision

Sensitivity **Precision**



True mask

Predicted mask

Correctly delineated mask

We introduced a generalisation of the popular dice loss [2], based on the F_β score

$$F_1 = \frac{2}{\frac{1}{S} + \frac{1}{P}}$$

$$F_\beta = \frac{1 + \beta^2}{\frac{\beta^2}{S} + \frac{1}{P}}$$

[2]: F. Milletari, N. Navab and S.-A. Ahmadi, ‘V-net: Fully convolutional neural networks for volumetric medical image segmentation’

We introduced a generalisation of the popular dice loss, based on the F_β score

$$F_1 = \frac{2}{\frac{1}{S} + \frac{1}{P}}$$

$$F_\beta = \frac{1+\beta^2}{\frac{\beta^2}{S} + \frac{1}{P}}$$

Sensitivity

Precision

We introduced a generalisation of the popular dice loss, based on the F_β score

$$F_1 = \frac{2}{\frac{1}{S} + \frac{1}{P}}$$

$$F_\beta = \frac{1 + \beta^2}{\frac{\beta^2}{S} + \frac{1}{P}}$$

Weighting constant

Sensitivity

Precision

We introduced a generalisation of the popular dice loss, based on the F_β score

$$F_1 = \frac{2}{\frac{1}{S} + \frac{1}{P}}$$

$$F_\beta = \frac{1 + \beta^2}{\frac{\beta^2}{S} + \frac{1}{P}}$$

Weighting constant

Sensitivity

Precision

The β-value specifies the degree in which sensitivity is weighted versus the precision

$$\frac{\partial F_\beta}{\partial S} = \frac{\partial F_\beta}{\partial P}$$

Whenever

$$S = \beta P$$

The β-value specifies the degree in which sensitivity is weighted versus the precision

$$\frac{\partial F_\beta}{\partial S} = \frac{\partial F_\beta}{\partial P}$$

Whenever

Sensitivity $\rightarrow$ $S = \beta P$ $\leftarrow$ **Precision**

The F_β be computed in the following fashion, making it suitable as a loss function

$$F_\beta = \frac{(1+\beta^2)|y\,\hat{y}|}{\beta^2|y| + |\hat{y}|}$$

 # Tumorous pixels

The F_β be computed in the following fashion, making it suitable as a loss function

$$F_\beta = \frac{(1+\beta^2)|y\,\hat{y}|}{\beta^2|y| + |\hat{y}|}$$

Tumorous pixels

Predicted tumorous pixels

The F_β be computed in the following fashion, making it suitable as a loss function

 # Correctly classified pixels

$$F_\beta = \frac{(1+\beta^2)|y\,\hat{y}|}{\beta^2|y| + |\hat{y}|}$$

 # Tumorous pixels

 # Predicted tumorous pixels

The F_β be computed in the following fashion, making it suitable as a loss function

 # Correctly classified pixels

$$F_{\boldsymbol{\beta}} = \frac{(\mathbf{1} + \boldsymbol{\beta}^2)|\boldsymbol{y}\,\widehat{\boldsymbol{y}}|}{\boldsymbol{\beta}^2|\boldsymbol{y}| + |\widehat{\boldsymbol{y}}|}$$

 # Tumorous pixels

 # Predicted tumorous pixels

The F_β be computed in the following fashion, making it suitable as a loss function

 # Correctly classified pixels

$$F_{\beta} = \frac{(1+\beta^2)|y\,\hat{y}|}{\beta^2|y| + |\hat{y}|}$$

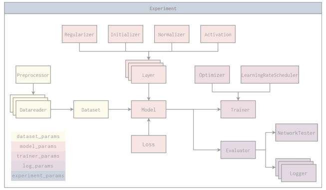
 # Tumorous pixels

 # Predicted tumorous pixels

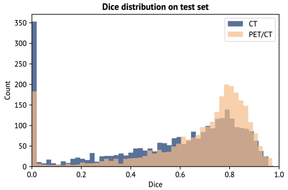
Proof in thesis

This presentation focuses on the work of my master's thesis

Motivation for automatic tumour delineation

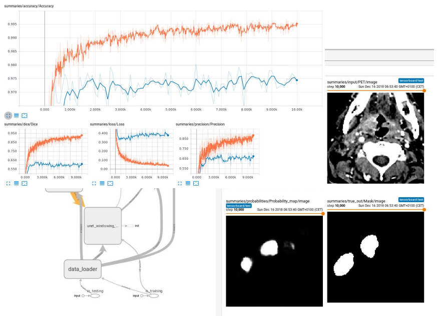


Experimental setup and scientific software

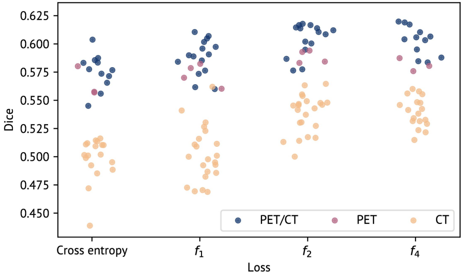


Discussion of the results

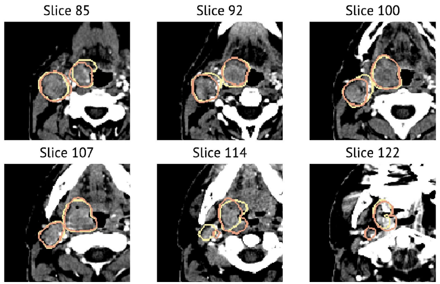
The results were three-fold



Evaluation of the SciNets library

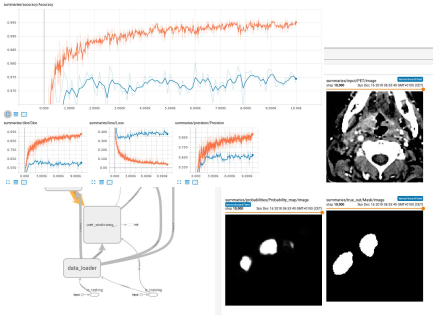


Comparison of different models

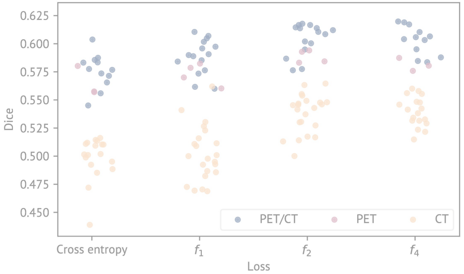


Assessment of the highest performing models

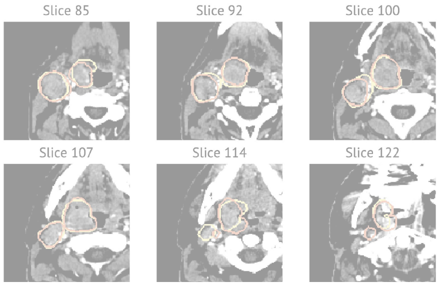
The results were three-fold



Evaluation of the SciNets library

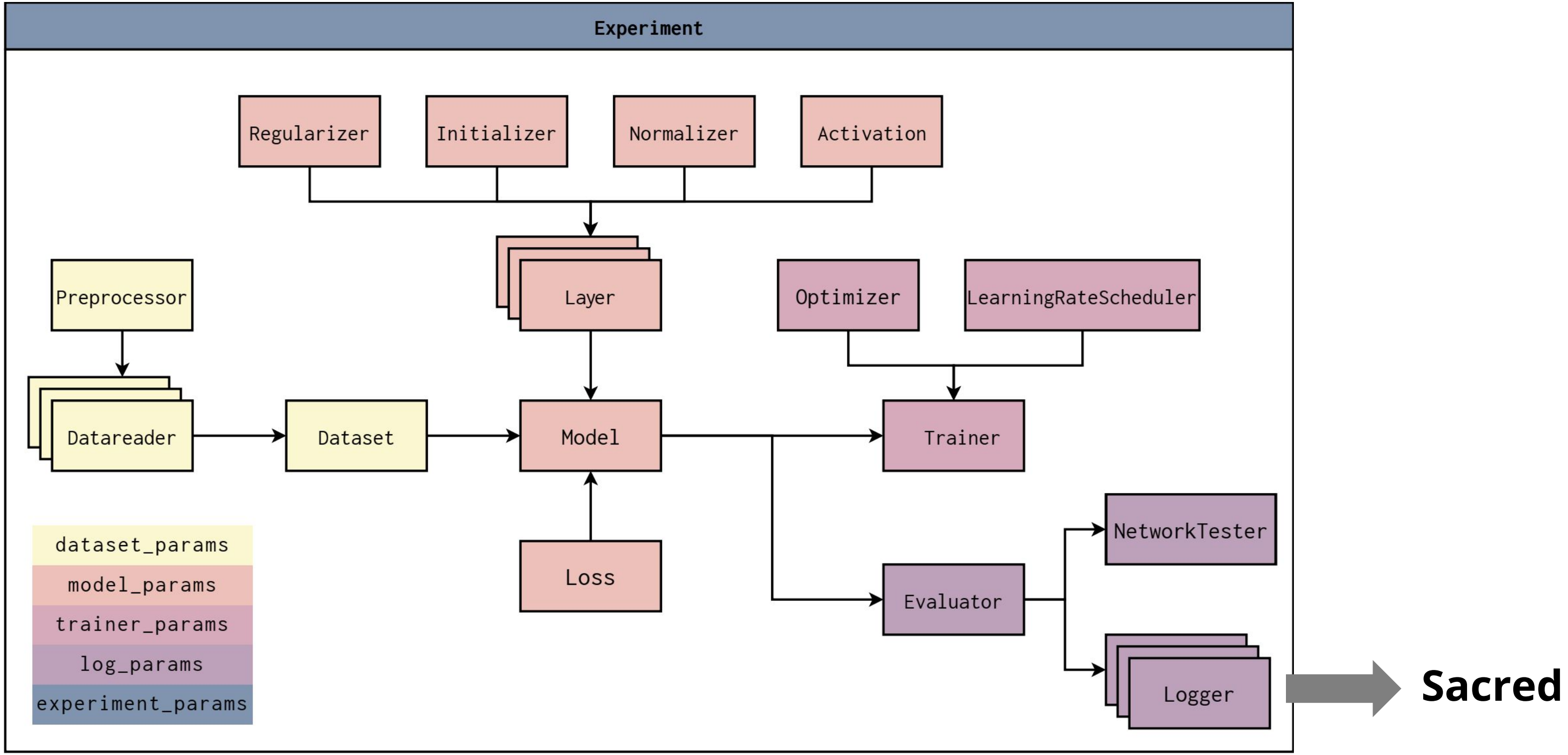


Comparison of different models

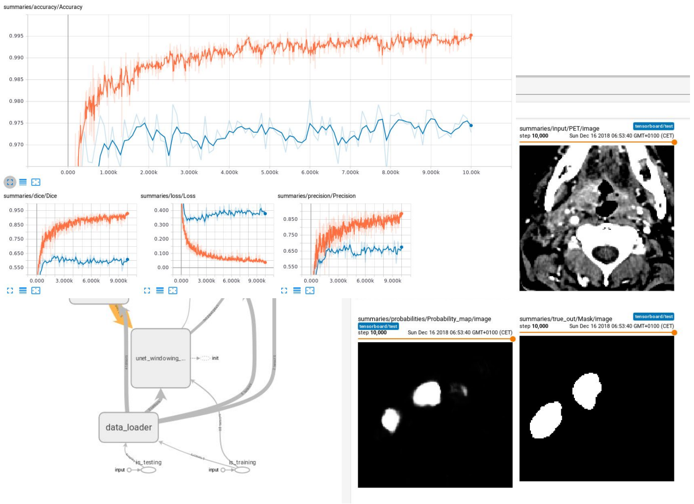


Assessment of the highest performing models

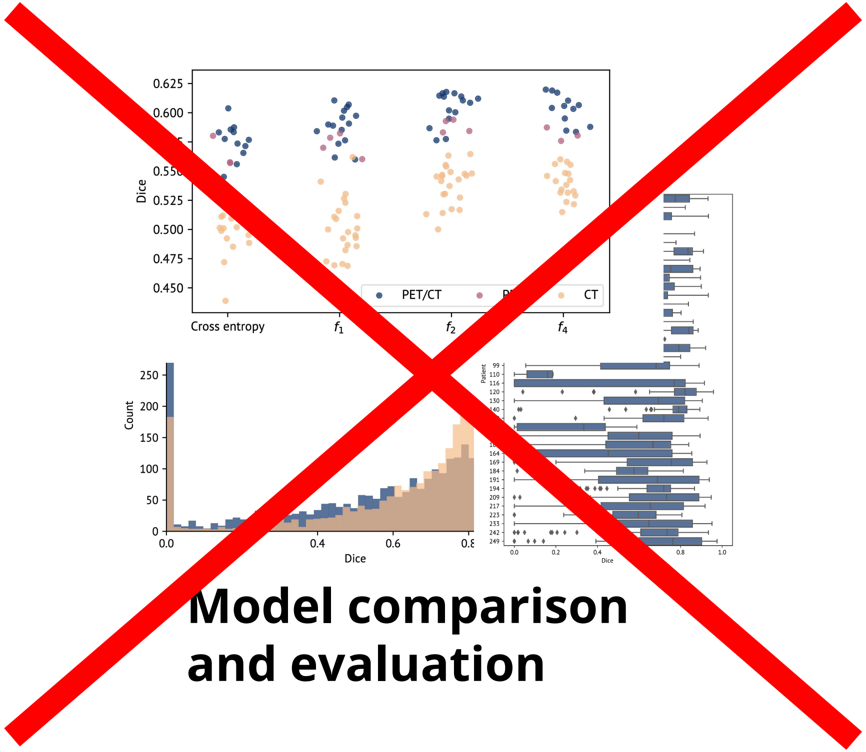
The SciNets library allowed for a vast parameter sweep without focusing on implementation details



The main drawback of the SciNets was that comparing results for several models proved cumbersome



Single model training diagnostics



Model comparison and evaluation

Furthermore, Sacred was not ideal for experiment logging

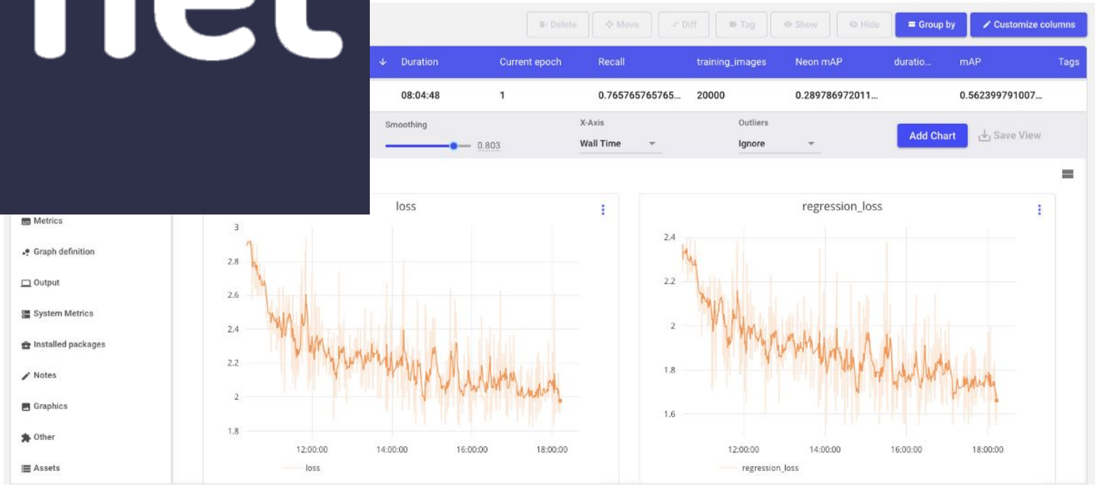
Sacred

> *Every experiment is sacred*
> *Every experiment is great*
> *If an experiment is wasted*
> *God gets quite irate*

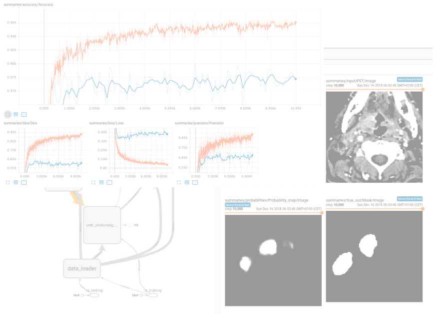
pypi v0.7.4 | python 2.7 | 3.4 | 3.5 | license MIT | docs failing | DOI 10.5281/zenodo.16386

unix build passing | windows build passing | coverage 84% | Scrutinizer 8.54 | Codacity rating

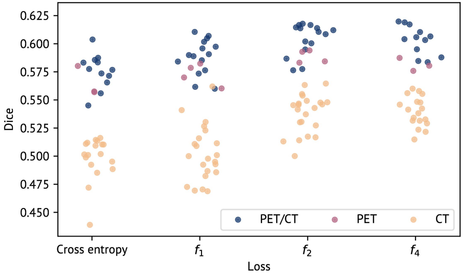
For future work, we, therefore, recommend to use the proprietary system *comet.ml* instead



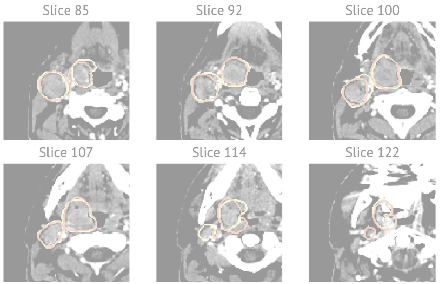
The results were three-fold



Evaluation of the SciNets library



Comparison of different models

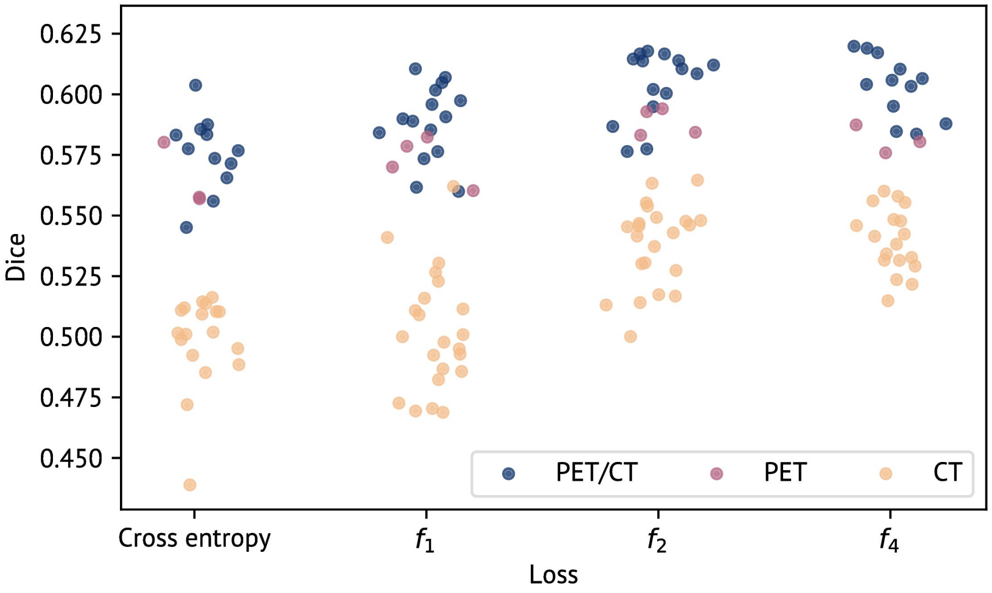


Assessment of the highest performing models

Models with ResNet layers did not train as they were affected by exploding gradients on skip-connections

$$\frac{\partial \check{f}}{\partial w_{skip}} =$$

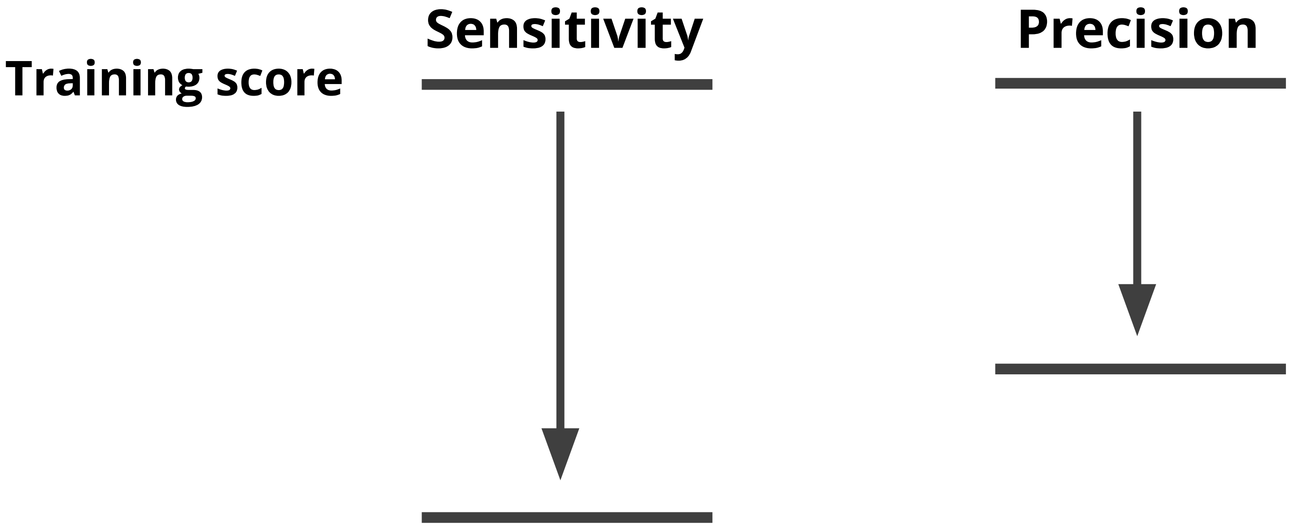
The choice of loss function had a clear effect on model performance



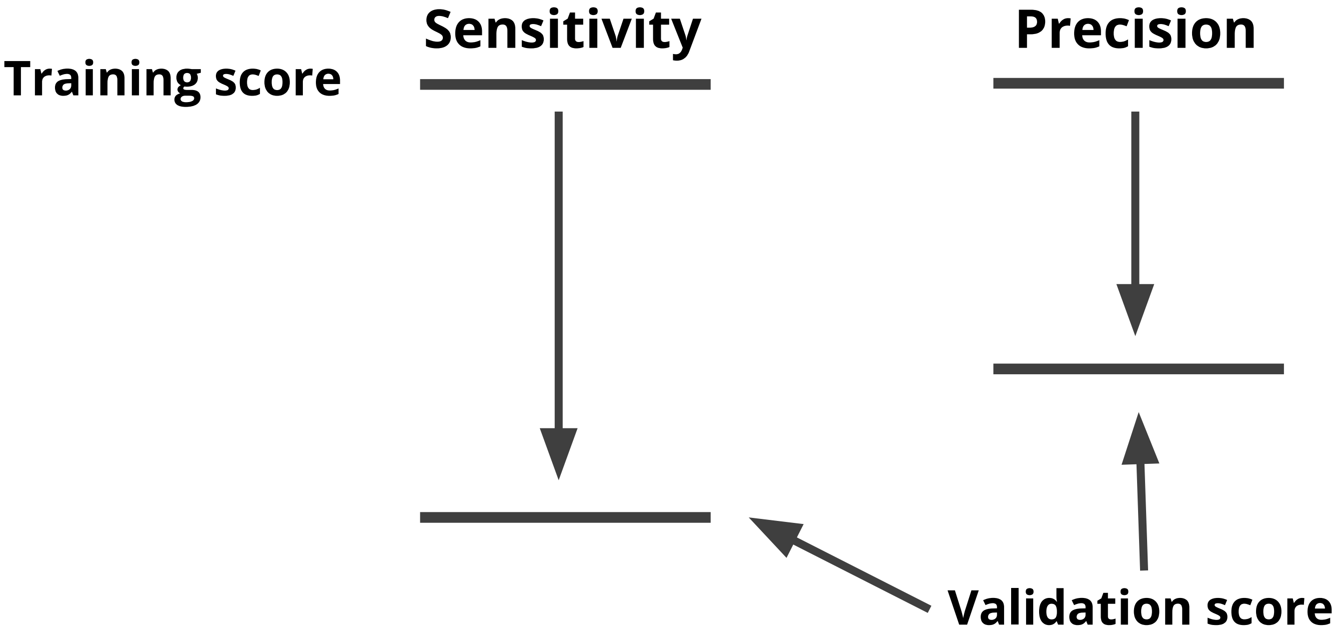
Thus, the generalisation gap is larger for the sensitivity than the precision for models trained with a dice loss

	Sensitivity	Precision
Training score		

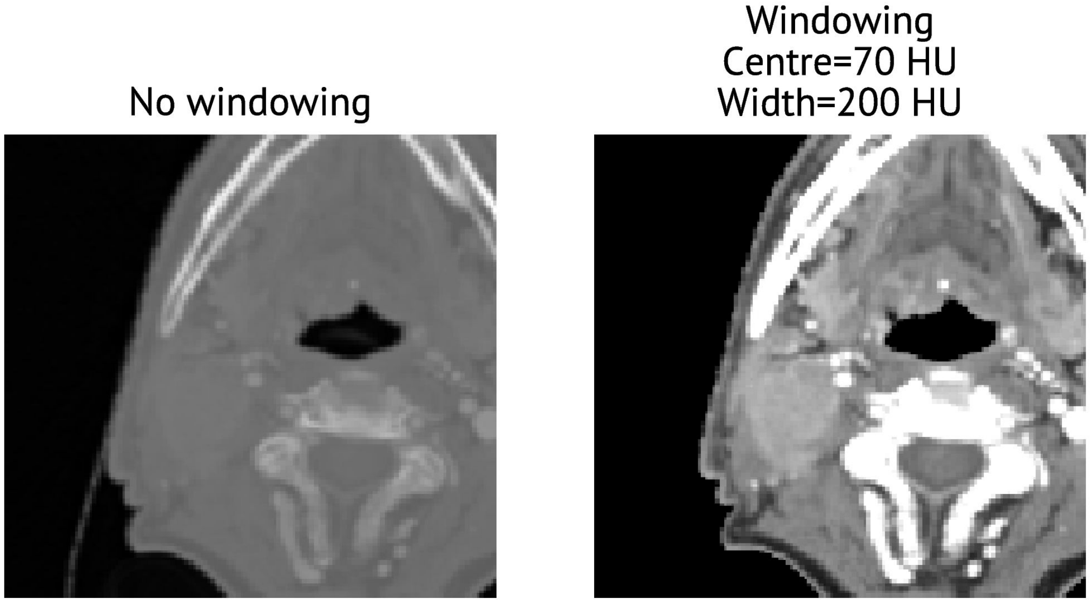
Thus, the generalisation gap is larger for the sensitivity than the precision for models trained with a dice loss



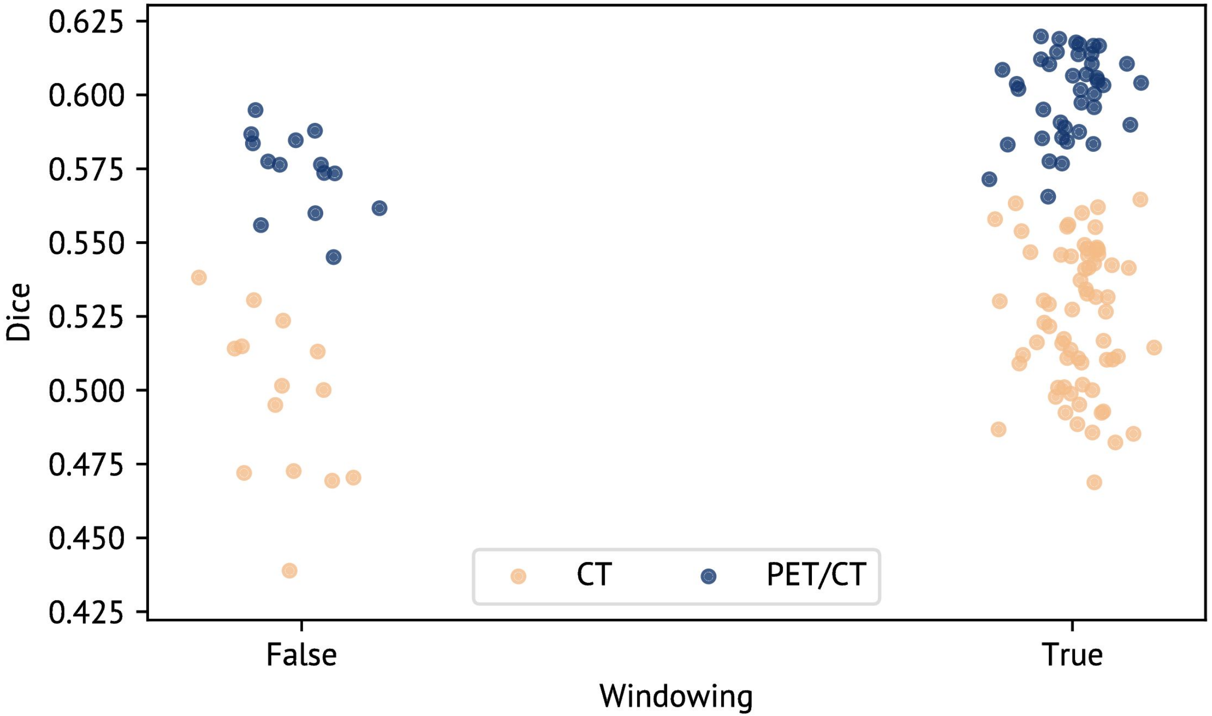
Thus, the generalisation gap is larger for the sensitivity than the precision for models trained with a dice loss



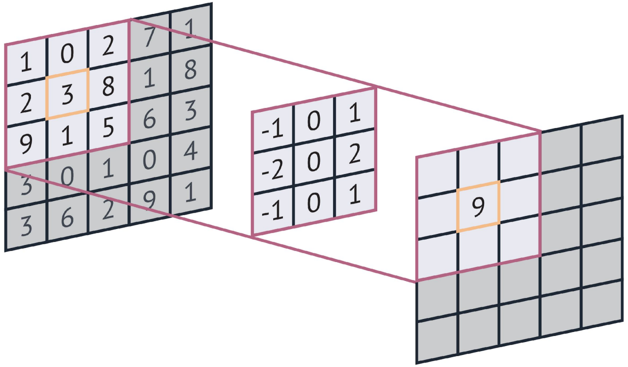
Reducing the dynamic range of the CT images also had a clear effect on model performance



Reducing the dynamic range of the CT images also had a clear effect on model performance

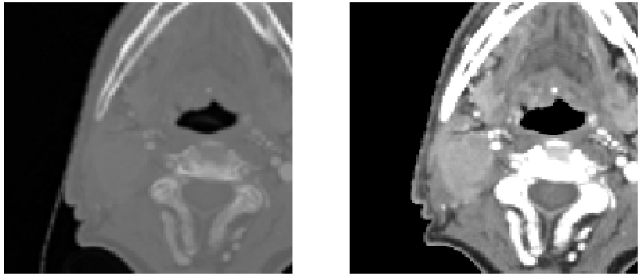


This is intriguing, as a windowing operation can be learned from a single layer neural network.

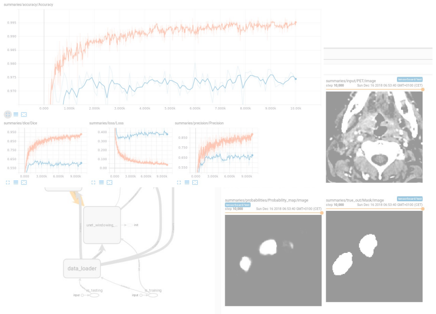


No windowing

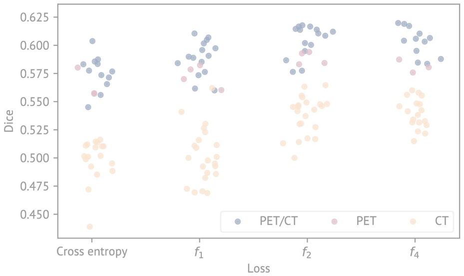
Windowing
Centre=70 HU
Width=200 HU



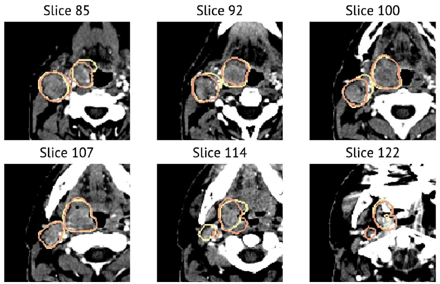
The results were three-fold



Evaluation of the SciNets library



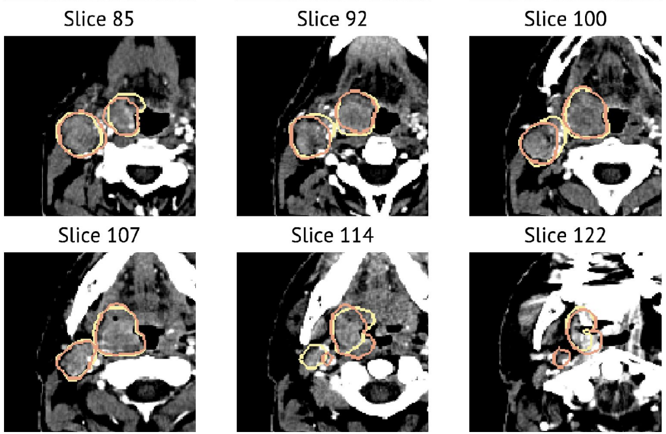
Comparison of different models



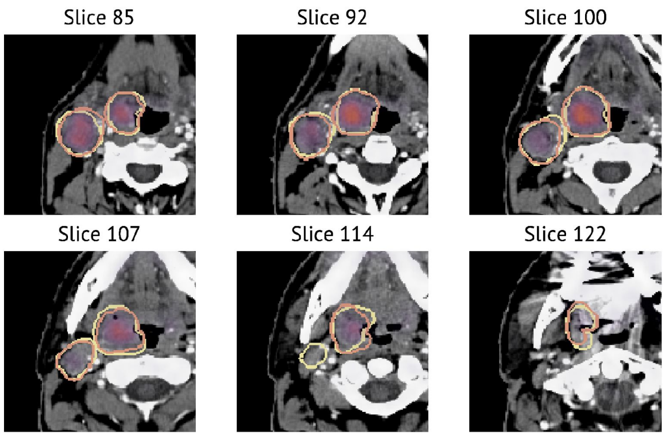
Assessment of the highest performing models

The highest performing CT-only model and PET/CT model were assessed on the test dataset

CT-Only



PET/CT



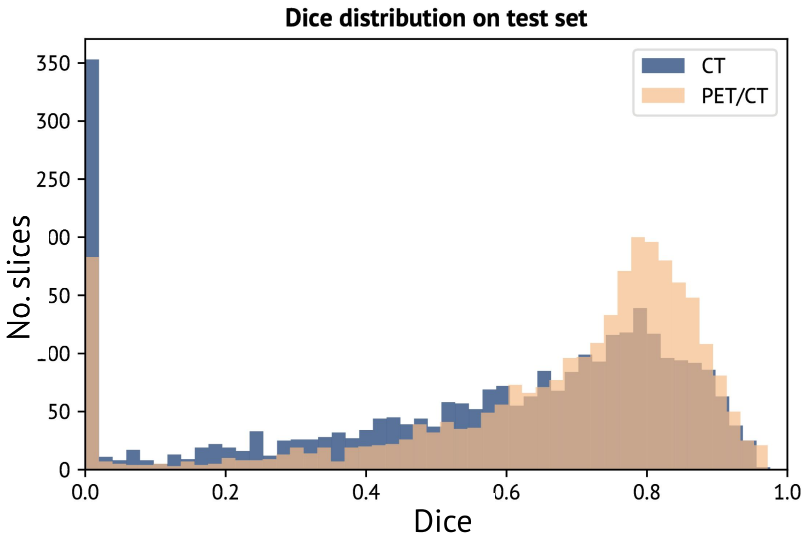
The best performant models achieved an average dice per slice comparable to that between radiologists [3]

Modality	Dice	Radiologist dice[3]
CT-only	0.56±0.29	0.57±0.12
PET-only	0.64±0.24	N/A
PET/CT	0.66±0.24	0.69±0.08

[3]: S. Gudi, et al., 'Inter-observer variability in the delineation of gross tumour volume and specified organs-at-risk during IMRT for head and neck cancers and the impact of FDG-PET/CT on such variability at the primary site'

The best performant models achieved an average dice per slice comparable to that between radiologists [3]

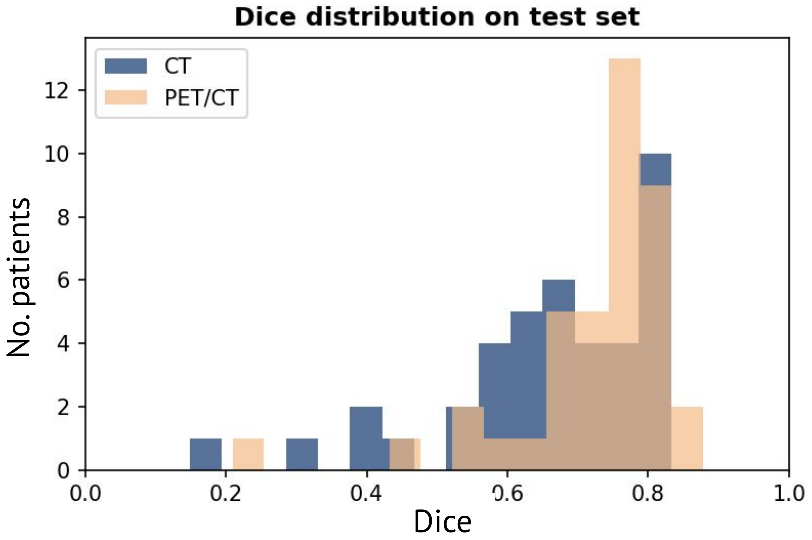
Modality	Dice	Radiologist dice[3]
CT-only	0.56±0.29	0.57±0.12
PET-only	0.64±0.24	N/A
PET/CT	0.66±0.24	0.69±0.08



[3]: S. Gudi, et al., ‘Inter-observer variability in the delineation of gross tumour volume and specified organs-at-risk during IMRT for head and neck cancers and the impact of FDG-PET/CT on such variability at the primary site’

However, the dice-per-patient is much higher than that between radiologists

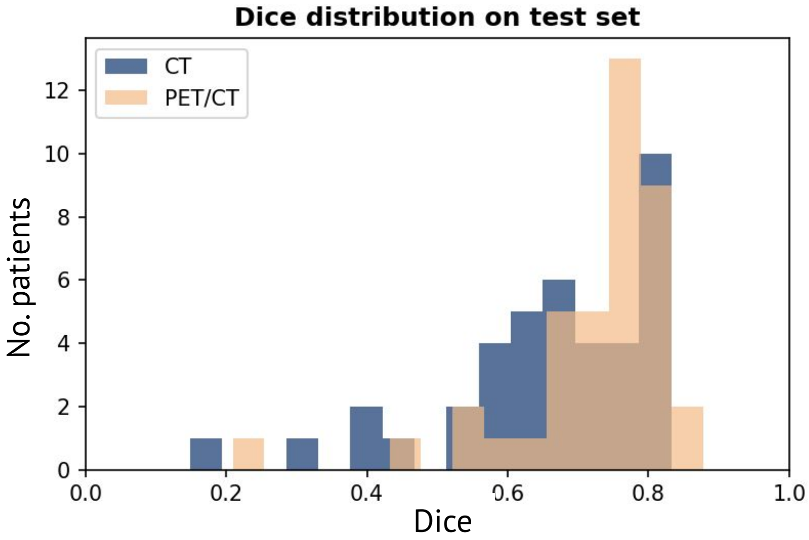
Modality	Dice	Radiologist dice[3]
CT-only	0.66±0.15	0.57±0.12
PET-only	0.71±0.10	N/A
PET/CT	0.73±0.12	0.69±0.08



[3]: S. Gudi, et al., ‘Inter-observer variability in the delineation of gross tumour volume and specified organs-at-risk during IMRT for head and neck cancers and the impact of FDG-PET/CT on such variability at the primary site’

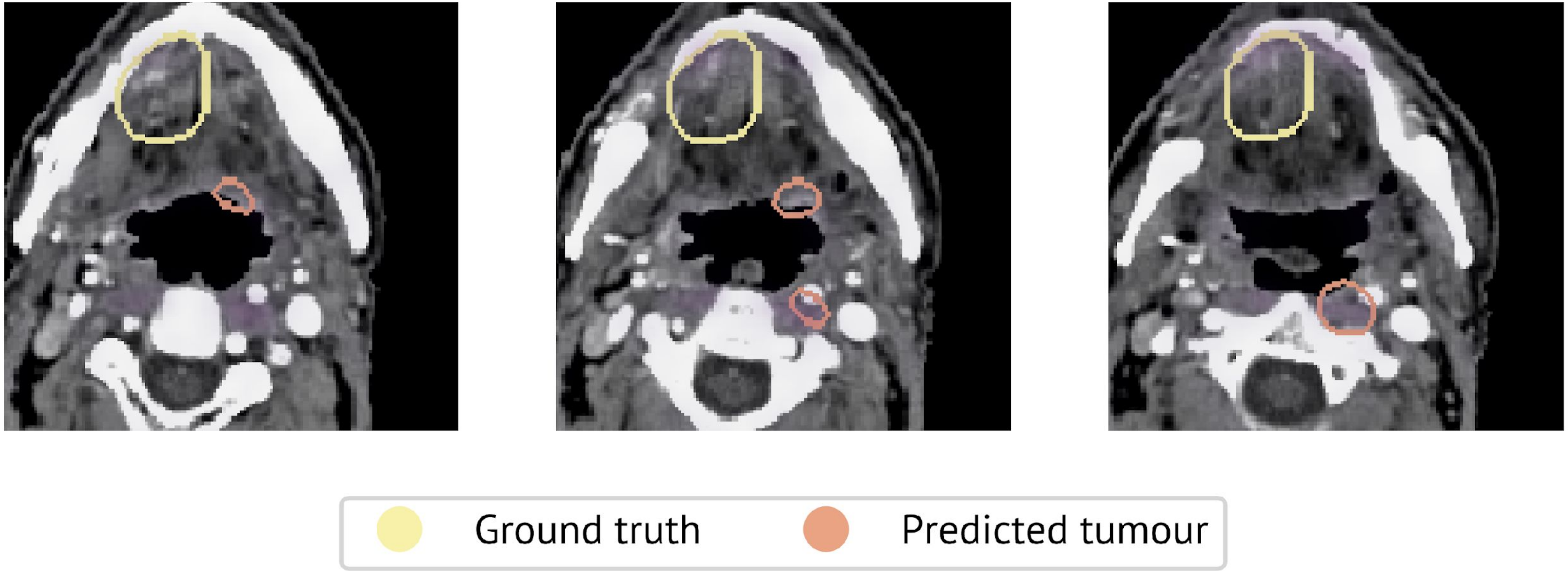
The reason for this is that slices with little to no tumour is weighted less with this metric

Modality	Dice	Radiologist dice[3]
CT-only	0.66±0.15	0.57±0.12
PET-only	0.71±0.10	N/A
PET/CT	0.73±0.12	0.69±0.08

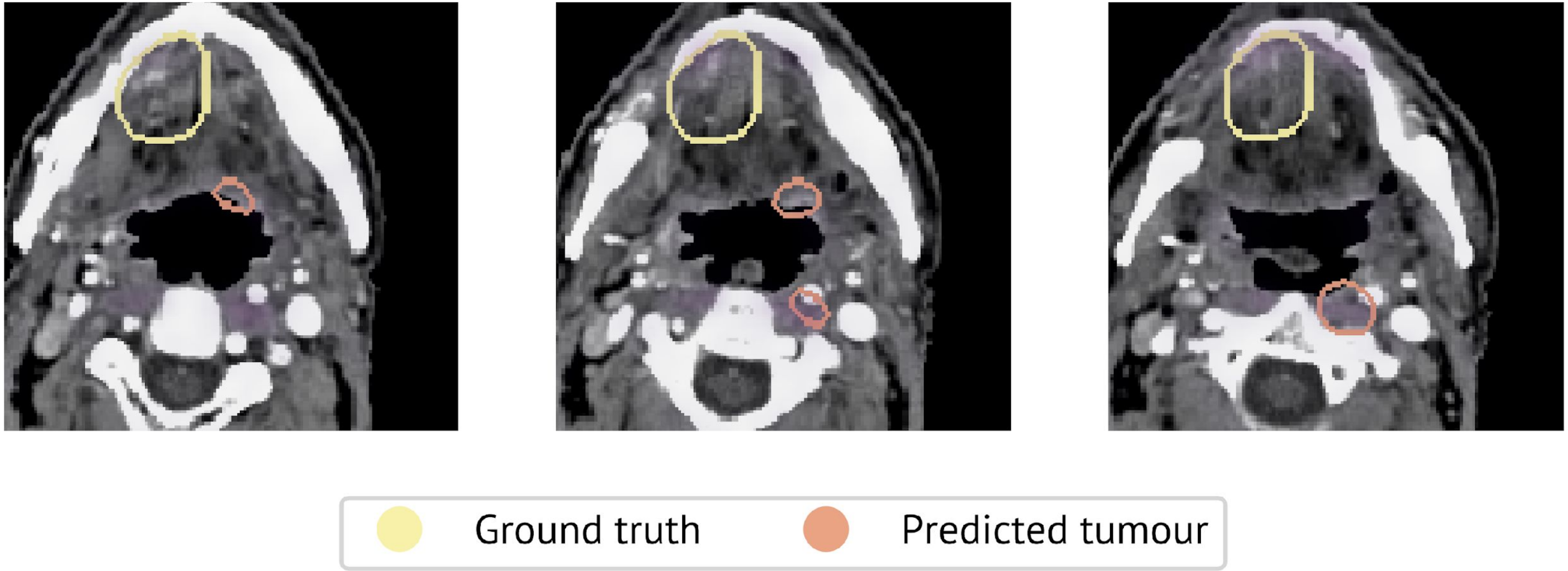


[3]: S. Gudi, et al., 'Inter-observer variability in the delineation of gross tumour volume and specified organs-at-risk during IMRT for head and neck cancers and the impact of FDG-PET/CT on such variability at the primary site'

A main weakness of deep learning is that the algorithms are overconfident on their predictions

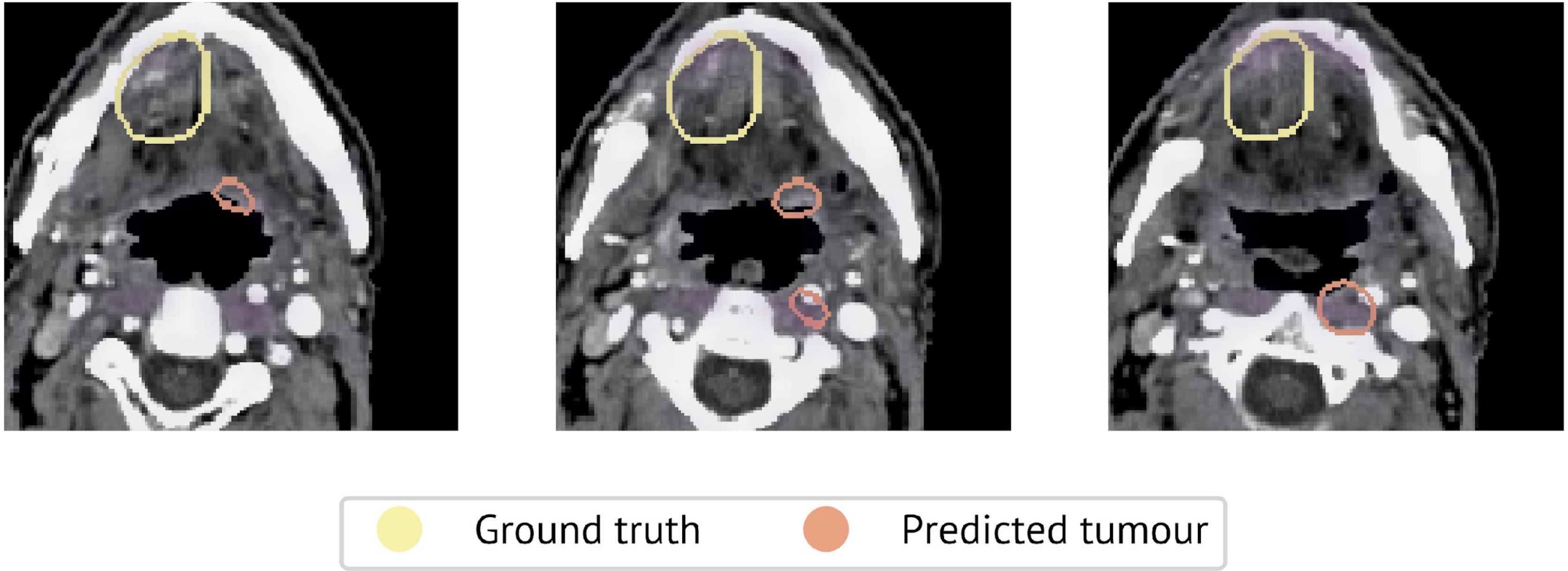


A main weakness of deep learning is that the algorithms are overconfident on their predictions



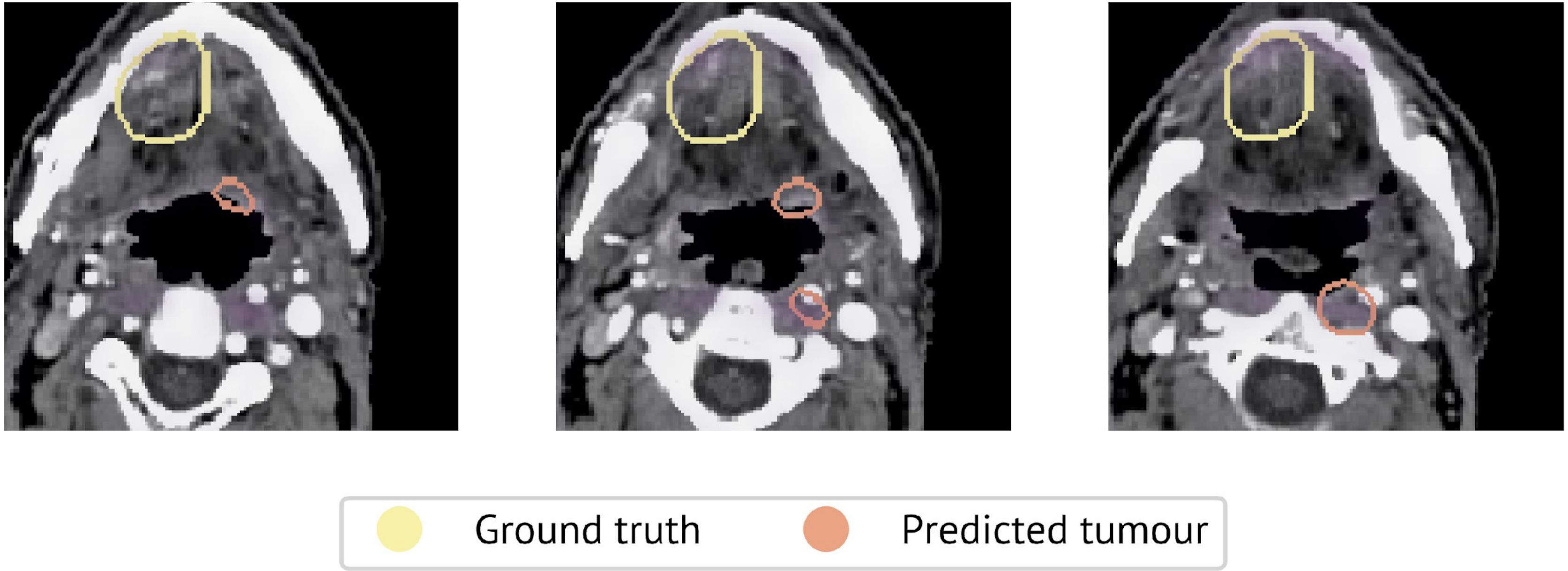
Bayesian deep learning might alleviate this

A main weakness of deep learning is that the algorithms are overconfident on their predictions

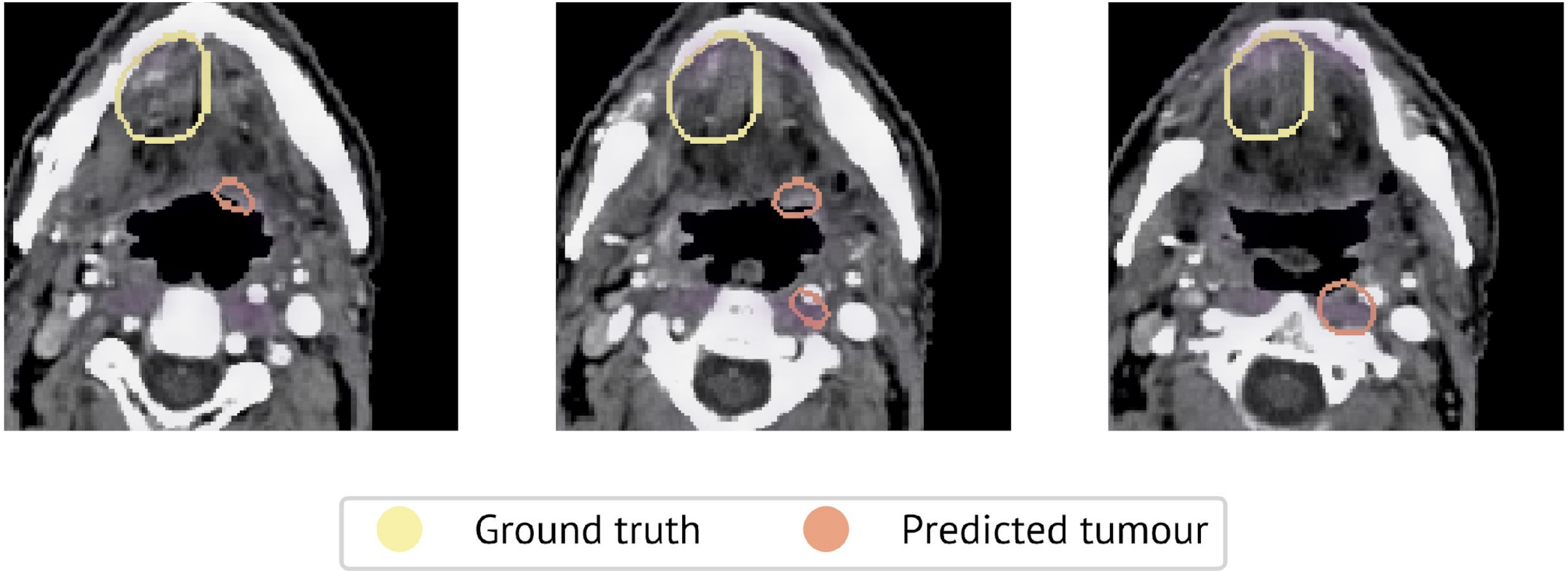


A two-step process in which radiologists are prompted with every proposed tumour might also be used

The mistakes made by the algorithm might be more severe than those made by professionals

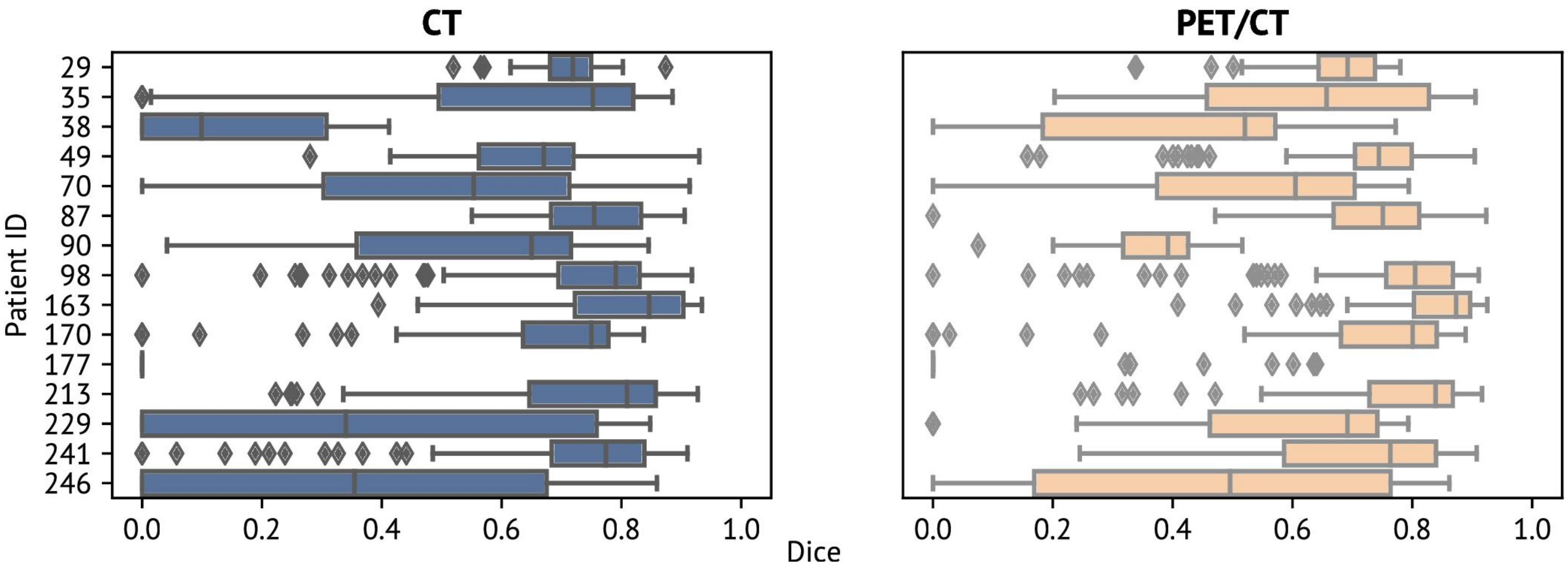


The mistakes made by the algorithm might be more severe than those made by professionals



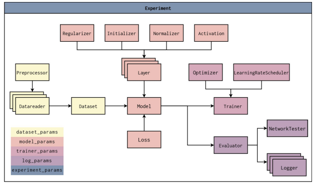
Radiologists should be consulted on this matter

The main difference between the PET/CT model and the CT-only model was their consistency

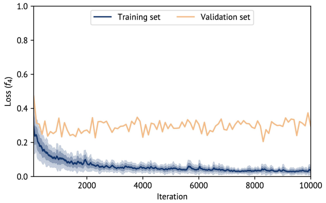


Dice performance on the validation set
(test results omitted for space considerations)

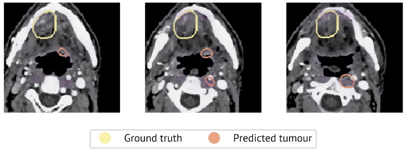
There is still much future work necessary



Extend and improve SciNets



Train models with more <u>and less</u> data

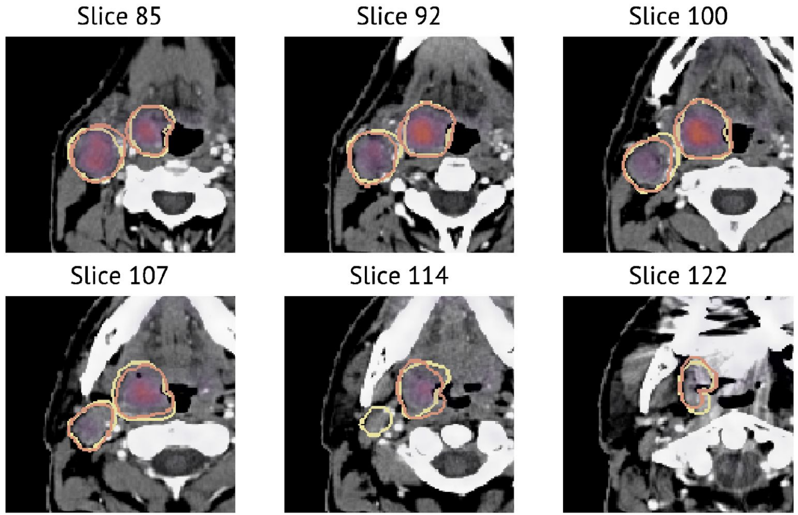


Assess the validity of the segmentation masks

We have shown that deep learning is a promising approach for automatic tumour delineation from PET/CT images

The dice performance is similar to that between clinicians

However, some mistakes made by these algorithms are severe

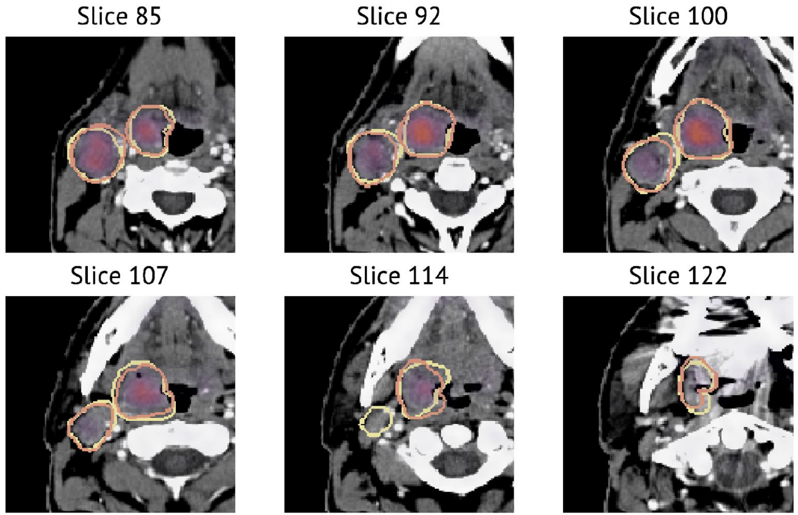


Norges miljø- og biovitenskapelige universitet

We have shown that deep learning is a promising approach for automatic tumour delineation from PET/CT images

The dice performance is similar to that between clinicians

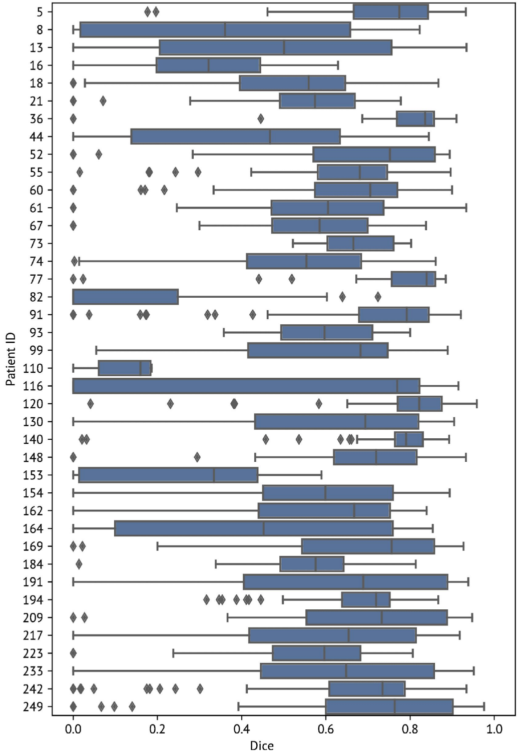
However, some mistakes made by these algorithms are severe



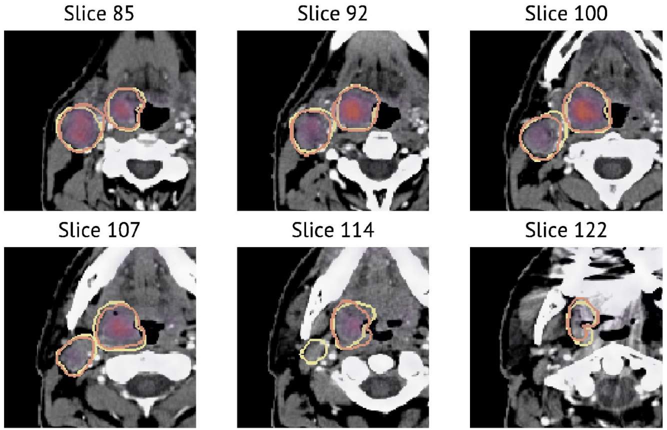
Questions?

Norges miljø- og biovitenskapelige universitet

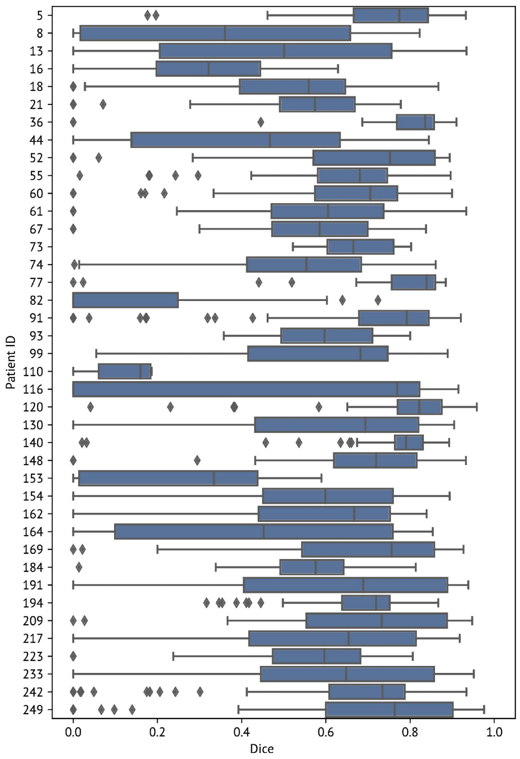
The model evaluation procedure was particularly cumbersome as the model outputs were shuffled

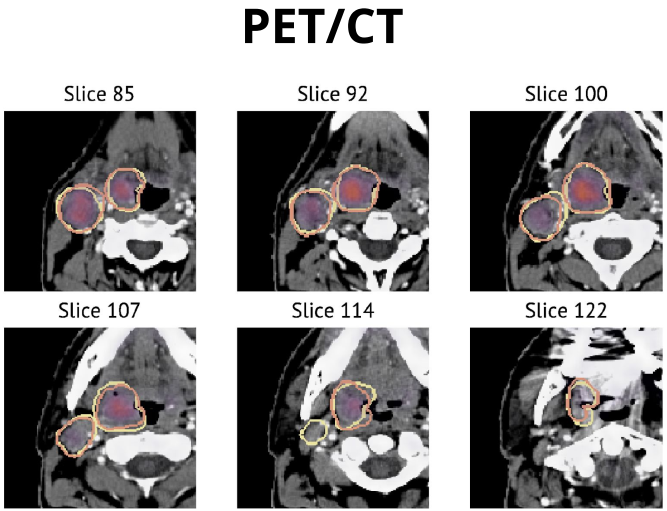


PET/CT



The model evaluation procedure was particularly cumbersome as the model outputs were shuffled





Figures like these were therefore particularly troublesome to make

Proof for this equality

$$\frac{\partial F_\beta}{\partial S} = \frac{\partial F_\beta}{\partial P}$$

Whenever

Sensitivity

$$S = \beta P$$

Precision

The β-value specifies the degree in which sensitivity is weighted versus the precision

$$\frac{\dfrac{\partial F_\beta}{\partial S} = G_\beta(S,P)\,\dfrac{\beta^2}{S^2}}{\dfrac{\partial F_\beta}{\partial S} = G_\beta(S,P)\,\dfrac{1}{P^2}}$$

The β-value specifies the degree in which sensitivity is weighted versus the precision

$$\frac{\partial F_\beta}{\partial S} = G_\beta(S,P)\,\frac{\beta^2}{S^2}$$

$$\frac{\partial F_\beta}{\partial S} = G_\beta(S,P)\,\frac{1}{P^2}$$

$$G_\beta(S,P) = \frac{F_\beta}{\left(\frac{\beta^2}{S} + \frac{1}{P}\right)}$$

We now set these derivatives equal and solve for S with respect to P

$$\frac{\partial F_\beta}{\partial S} = \frac{\partial F_\beta}{\partial S}$$

We now set these equal to each other and solve for S with respect to P

Substituting for the derivatives yields:

$$\boldsymbol{G_{\beta}(S,P)\frac{\beta^2}{S^2} = G_{\beta}(S,P)\frac{1}{P^2}}$$

We now set these equal to each other and solve for S with respect to P

The G_β are on both sides of the equality sign:

$$G_\beta(S,P)\,\frac{\beta^2}{S^2} = G_\beta(S,P)\,\frac{1}{P^2}$$

We now set these equal to each other and solve for S with respect to P

Computing reciprocals yields:

$$\frac{S^2}{\beta^2} = P^2$$

We now set these equal to each other and solve for S with respect to P

Thus:

$$S = \beta P$$