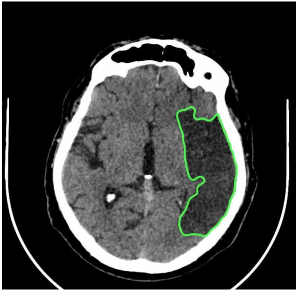
Training of deep neural networks with incomplete training information on the example of recognition of tomographic images

Kalmutskiy Kirill

Dcecember 2020

Formulation of the problem



A large dataset of tomographic images of the brain is given. Each image is a section of the brain. Manual marking of such a dataset is very expensive, since only a highly qualified specialist can do this.

In this regard, it is necessary to create a model that will predict areas of stroke and select them on tomographic images of the brain. The model will be used as a "decision support system."

Complexity of the problem and scientific novelty

Complexity

- Very little labeled data available: only 34 images, which is clearly not enough for training classical deep convolutional neural networks;
- Since the model will be used as a decision support system, it is necessary not only to properly segment the image, but also to have high recognition accuracy;
- It is necessary to choose the right balance between errors of the first and second kind;
- The model must be interpretable, otherwise doctors will not understand its decisions and will not trust it.

Scientific novelty

- All current solutions require a huge amount of training data;
- The interpretability of neural networks is poorly understood.

Initial data

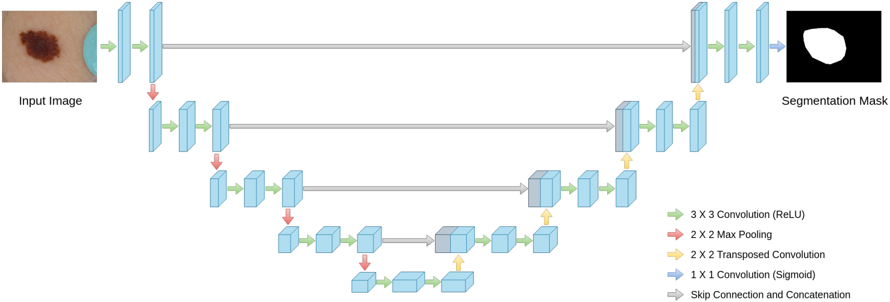
- Images in DICOM format, labels in JPG format;
- Only 34 labeled images of brain slices;
- More than 30 GB of unlabeled images.

Segmentation task

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects).
More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

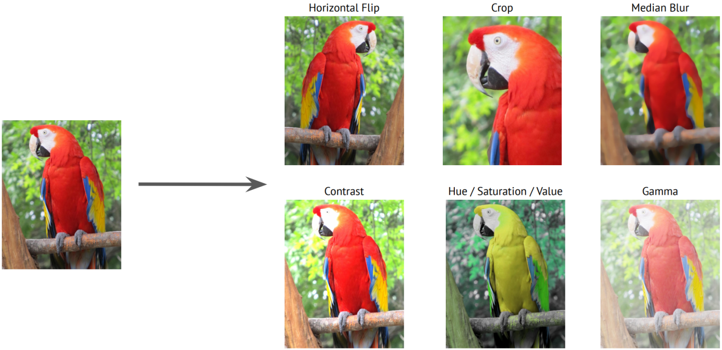
- **Automatic segmentation**: bad accuracy, doesn't work in medical image recognition tasks;
- **Semi-automatic segmentation**: requires the participation of a specialist during segmentation process, poorly suited for creating a decision-making system;
- **Trainable segmentation**: the most promising approach.

Trainable segmentation



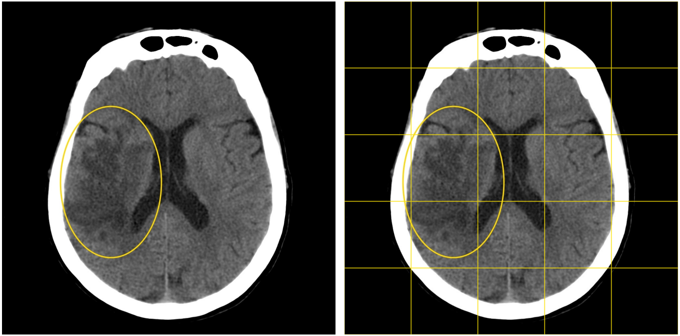
- The model allows you to achieve high metrics;
- A lot of data is needed for training;
- Transfer learning doesn't work well for medical image recognition tasks;
- Inability to interpret model decisions.

Augmentations



The only augmentations that have improved the quality are horizontal and vertical flips, as well as a small zoom.

Cutting the image into sub images



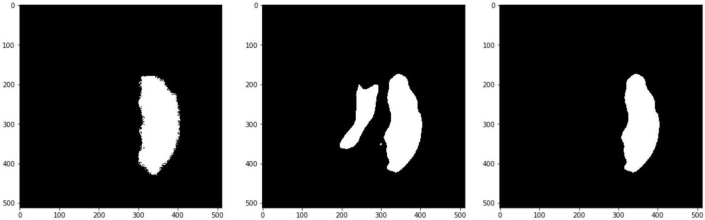
- Learning on crops smaller than the original image;
- To predict, we cut the picture into crop again, predict for them and collect the image back;
- We increase the dataset 16 times using a 4x4 grid.

Network Architecture

To improve the accuracy of stroke diagnosis, the model includes two parts: a segmentator and a classifier. Thus, segmentation is performed only if the classifier reveals a stroke in the image.

- **Segmentator**: an architecture including an encoder and a decoder with skip-connections and higher dilation parameter;
- **Classifier**: a linear layer at the output of the encoder;
- **Loss function**: sum of the weighted cross-entropy of the classifier and the weighted cross-entropy of the segmentator.

Model Prediction Processing



Left image - ground truth, middle image - all components, right image - only the maximum connected component

- **Using TTA**: augmentations are applied not only during training, but also during prediction, then the predictions for the images are averaged;
- To increase the quality of segmentation, the model **leaves only the maximum connected component** in the resulting segmented image;
- In addition to this, **the remaining component is removed if it is less than a certain size** (currently less than 1200 pixels).

Decision trees based on similarity

Decision trees based on similarity is a binary decision tree that uses a measure of object similarity to key points instead of predicates. Cluster centers can be used as key points, and the distance to these points - as a measure of similarity. **Algorithm**:

- Cluster images;
- Find key points (typical images);
- Find the best metric (L1, L2 Hamming etc);
- For each object, calculate the distance to key points, thus moving to a new feature space

This algorithm has excellent interpretability as it looks for the most similar labeled examples when predicting. This classifier is trained on the features of the encoder (like the linear layer in the described neural network architecture).

Results of Experiments

Table: Validation and Test Results

Metric	Validation	Test
Accuracy	0.883	0.867
IoU	0.693	0.621

For the experiment, 2 samples were formed:

- **Validation**: 8 (4+4) images (128 sub images);
- **Test**: 6 images (92 sub images).

Further work

- Improving similarity-based decision tree (by using model-based gradient boosting for example);
- Advanced preprocessing (flattening the histogram, using Sobel filters, highlighting the brain outline and placing it in a minimal rectangle);
- Further improvements to the model architecture.