Analysis of CNN working with logical decision functions in the task of computer tomography images recognition

R.M. Kozinets, V.B. Berikov

November 14, 2019

NSU-2019

Overview

Problem formulation

Task and Data

Current state of the work

Analysis CNN in lung cancer recognition(previous semester)

Problem formulation

Problem

General statement
Assume X set of objects x . Y is set of labels, $X^d \in X$- random sample of size d. $y^* : X \to Y$ goal function known on X^d . Need to build decision function $f(X) \to Y$ minimizing the probability of error.

More accurate formulation

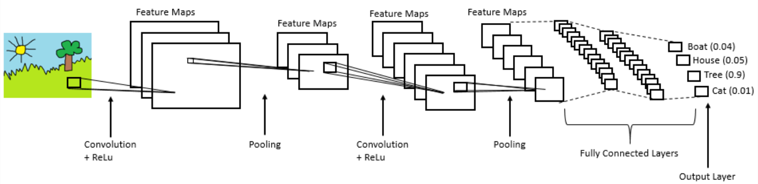
- X is set of objects and Y is set of labels, X corresponds with Y, the task is: Using elements in CNN functional set of tool build feature representation of object? which most optimal in classification task by logical functions.
- Based builded feature representation of object, develop logical predictive model with high interpretive decision making process and high quality predictive ability.

Task and Data

Tasks and Data

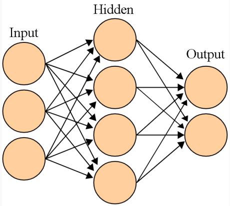
- The task: two types stroke recognition by Computer tomography images.
- Data: CT images in DICOM format of human brain with/(without) pathologies.
- Labels: Ichemic stroke, hemorrhagic stroke, none.
- Expected result: recognition model with CNN module for feature extracting with interpretative decision making process using logical recognition functions and vision based explanation.

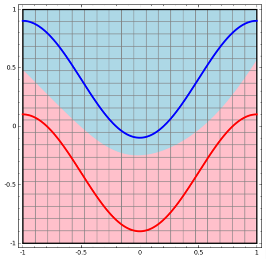
CNN classifier

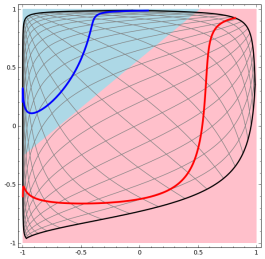


- Feature extractor: CNN block
- Classifer: MLP block with softmax
- CNN learned weights in optimal way for MLP classifier after: $f(x) = argmin(L(MLP(f(x))))$
- MLP with hidden layers transform input feature space to be linearly separable

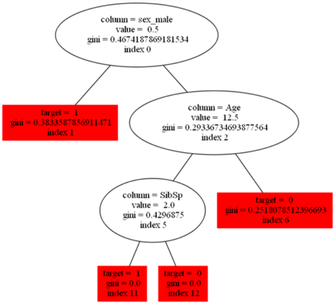
MLP data transformation







Decision Tree



Example of decision making($x_k \to y$):
$(f_1(x_k) < M_i) and (f_2(x_k)) < (M_j) \to y_k$,where f_i - features of x_k, y_k - predicted label

CNN with decision tree as classifier

Take feature vector from last conv. layer of pretrained network and build decision tree on given data

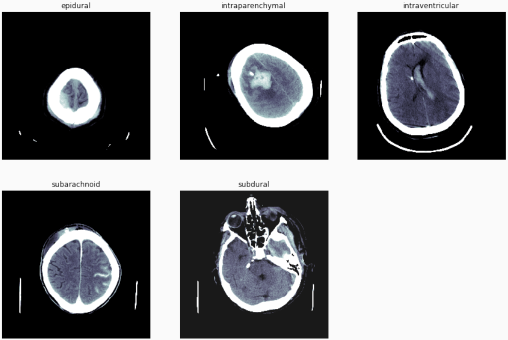
- it give as weakly Analysis of feature importance in decision making process
- problem: this features can be not optimal for decision tree type of model.

Solution - train CNN with decision tree as classifier, if CNN and DT will be train simultaneously, learned features are optimal for DT, and decision tree is optimal for learned features.

- optimal learned features for logical recognition function
- strong CNN features analysis
- main problem - decision tree learning process is not differentiable

Current state of the work

Task and Data

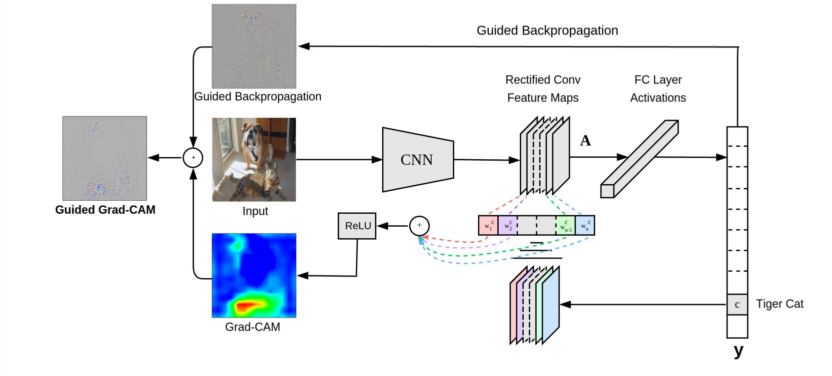


- RSNA Intracranial Hemorrhage Detection (multi label classification)
- Five Hemorrhage subtypes (epidural, intraparenchymall, intraventricular, subarachnoid, subdural)
- 674258 raw Dicom images for train with class imbalance

Intracranial Hemorrhage Detection

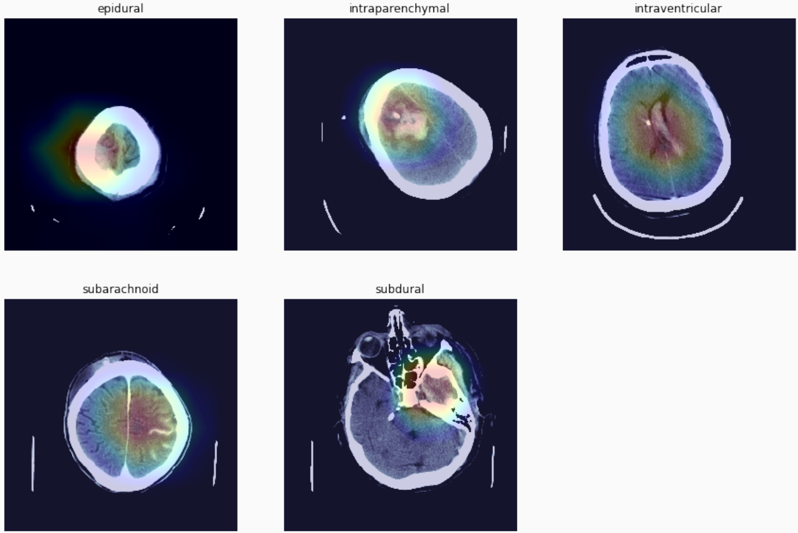
- EfficientNet B2 with one hidden layer MLP, binary cross entropy loss
- Trainig: 3 epoch
- Result
    - Test loss 0.23
    - Test AUC = 0.95
    - Test Acc = 0.97

GradCAM



Guided Backpropagation
Guided Backpropagation
Rectified Conv
Feature Maps
FC Layer
Activations
A
CNN
Guided Grad-CAM
Input
ReLU
Grad-CAM
Tiger Cat
c
y

GradCAM


epidural
intraparenchymal
intraventricular
subarachnoid
subdural

Tree Regularization CNN training

Analysis CNN in lung cancer recognition(previous semester)

Data preposition

- Crop tumours from images with 50x50 resolution
- There are a total of 551065 annotations.
  - 1351 positive(cancer)
  - 549714(non-cancer)
- So there big class imbalance for training
- Augment positive class by rotating images by 90,180 and 270 degrees.

a) b) c) d)

Figure 2. Augmentation: a) - original,b) - 90 degree rotation c) - 180 degree rotation, d) - 270 degree rotation

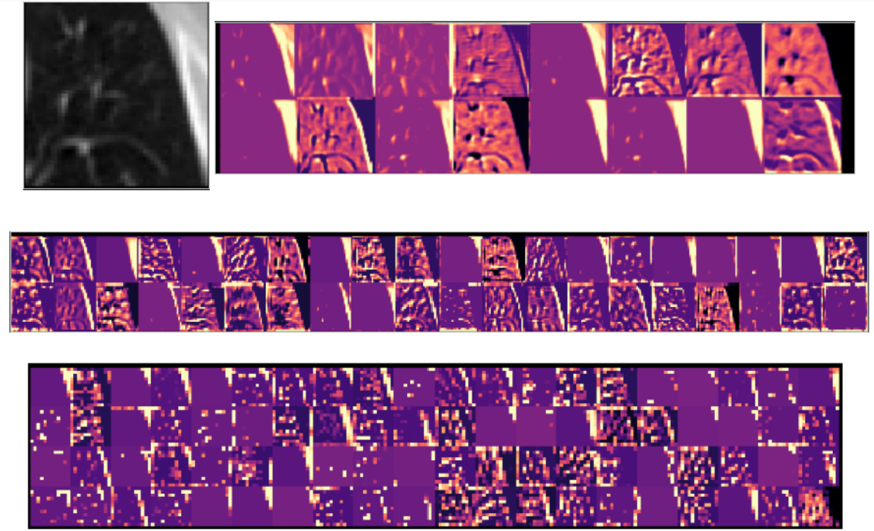
CNN model training

- Train (7722):
    - cancer 3380
    - non-cancer 4342
- Test (2251):
    - cancer 1073
    - non-cancer 1178

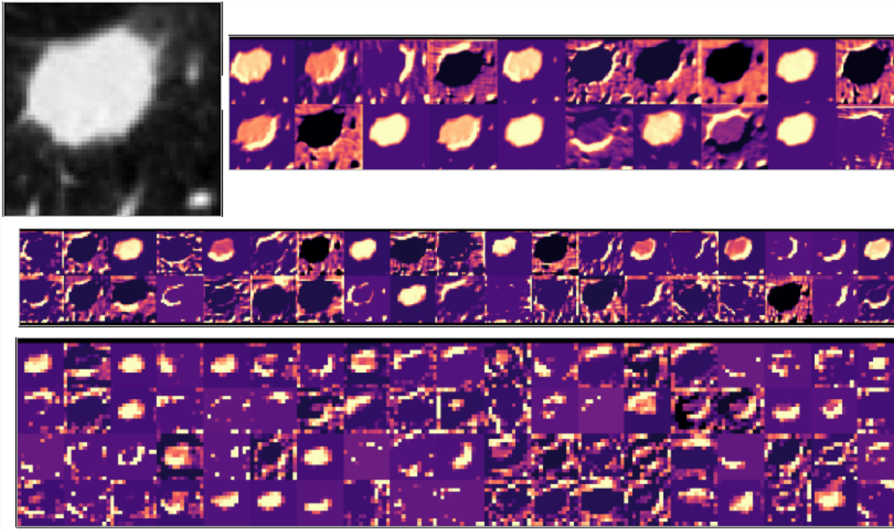
Input layer shape = (batch_size,50,50,1)
Conv2D 20 (5x5) + relu
Maxpooling (2,2)
Conv2D 40 (3x3) + relu
MaxPooling (2,2)
Conv2D 80 (3,3) + relu
Maxpooling (2,2)
Flatten
Dence(180) + relu
Dence(2) + softmax

sample	accuracy	loss
Train	0.982906	0.050968
Test	0.9253665	0.210678

Actuation's visualization: benign tumor



Actuation's visualizing:malignant tumor



Filter Visualization

Assume I - origin image, F_l^c - filter c of layer l. We want to get image I' witch get maximal activation of filter F_l^c. In this way we train our images by this process:

$$I' = I' + \frac{dL(F_l^c, I')}{dI'} rate,$$

where $L(F_l^c, I') = \frac{1}{n*m} \sum_{i,j}^{m,n} A_{i,j}^{l,c}$ We change our image I in gradient of loss of F_l^c

Figure 1: conv layer 1

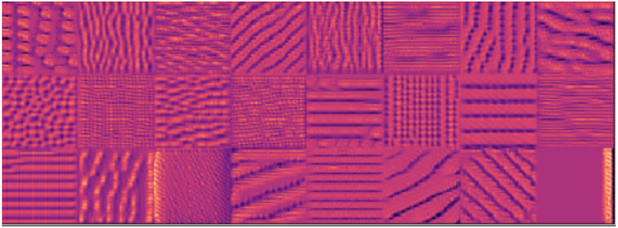


Figure 2: conv layer2

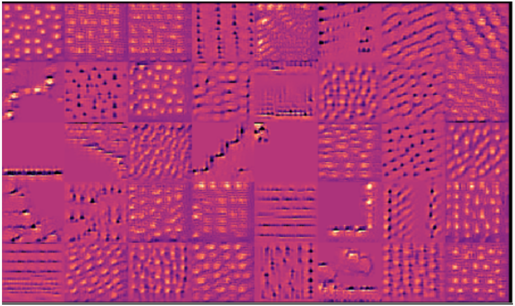


Figure 3: conv layer3

Covariation "Style" matrix

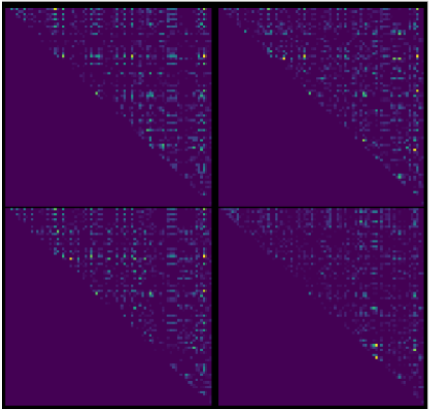
I – input image,$A^{c,l}_{i,j}(I)$ activation in position (i,j) of filter c of layer l for image I. Covariation between two filter defined as:

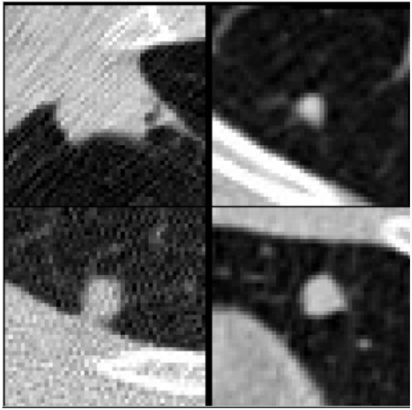
$$cov(f_1, f_2) = \frac{\sum_{i,j}^{n}((A^{f1}_{i,j} - mean(A^{f1}_{i,j}))((A^{f2}_{i,j} - mean(A^{f2}_{i,j})))}{n^2}$$

where n – size of feature map.

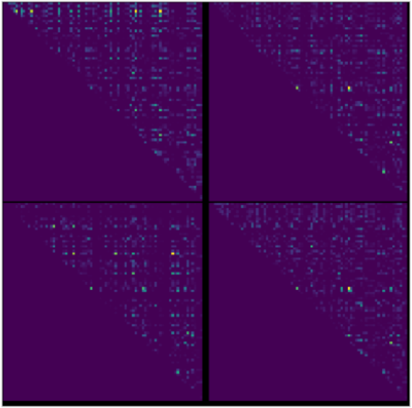
malignant tumour	benign tumour
(28, 40)	<u>(34,56)</u>
(57, 76)	**(56,73)**
<u>(34, 56)</u>	(34,35)
(56, 73)	***(56,76)***
(40, 68)	(53,76)

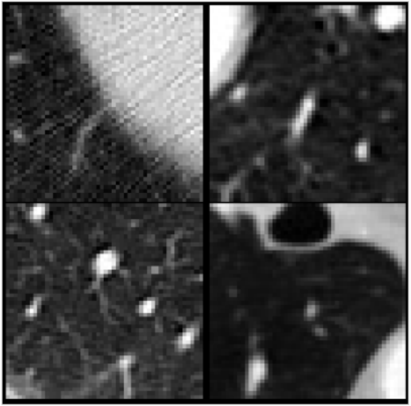
Covariation matrix:malignant tumor



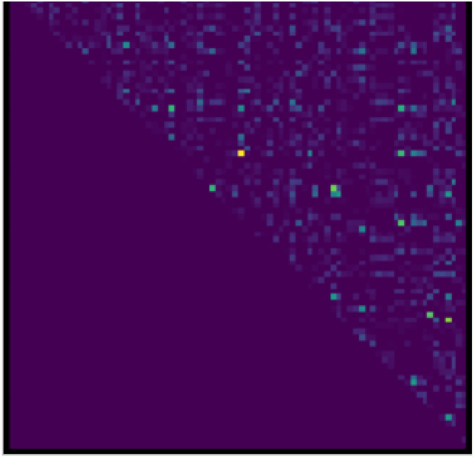


Covariation matrix: benign tumor

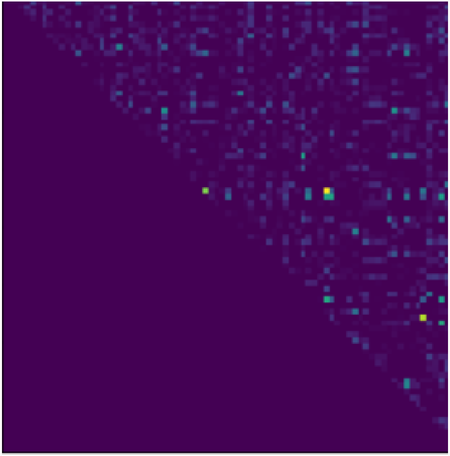




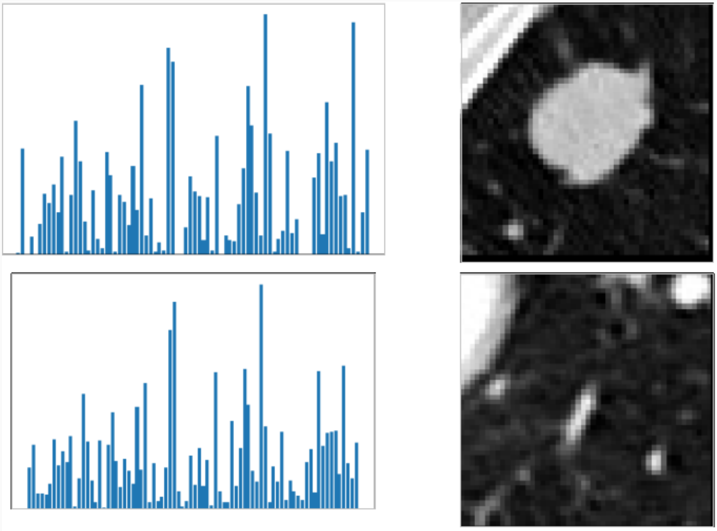
Averaged covariation:malignant tumor



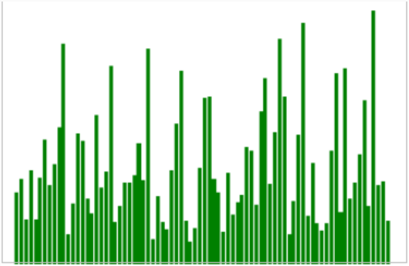
Averaged covariation:benign tumor

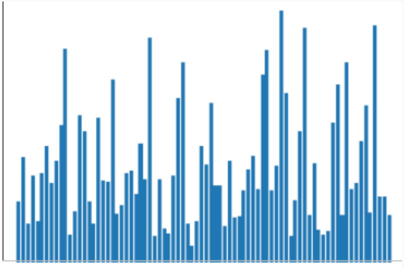


Layer distribution

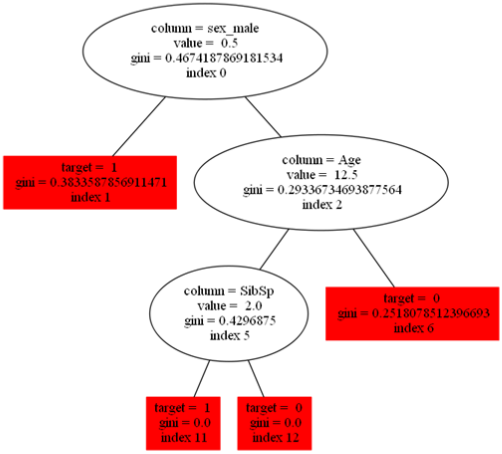


Layer distribution





Decision tree



Decision tree

Method	Depth = 3	Depth = 4	Depth = 5	Depth = 6
DT + max activations	0.7209	0.7526	0.79056	0.81122
DT + average activations	0.75370	0.76472	0.76610	0.78091
DT + full layer	0.88219	0.867034	0.89700	0.887357
DT + maxpooling	0.85291	0.848432	0.88804	0.88012

Conclusion

- Visualization the most understandable form but cannot be explained mathematically
- For effective using a decision tree model we need to use spatial information of features activation with values of activation
- In feature work different methods of analysis and visualizations will be combined and used to build decision tree to explain decision making of CNN.

Thanks for attention