Brain Tumor Image Retrieval via Multitask Learning

Maxim Pisov*[1,2], Gleb Makarchuk⋆[3], Valery Kostjuchenko[4], Alexandra Dalechina[4], Andrey Golanov[5], and Mikhail Belyaev[3,1]

[1] Kharkevich Institute for Information Transmission Problems, Moscow, Russia
[2] Moscow Institute of Physics and Technology, Moscow, Russia
[3] Skolkovo Institute of Science and Technology, Moscow, Russia
[4] Moscow Gamma-Knife Center, Moscow, Russia
[5] Burdenko Neurosurgery Institute, Moscow, Russia
m.belyaev@skoltech.ru

Abstract. Building an effective retrieval system for brain tumors is an open problem. Every year around 80,000 new cases of primary tumors and more than 200,000 cases of metastatic brain tumors are recorded in the world. The image retrieval system for tumors will help doctors to predict the development of diseases and treatment planning based on analysis of similar cases. Usually, modern image retrieval systems are built by using convolutional neural networks, which are trained on a relevant classification problem. For tumors, however, the similarity is measured simultaneously with multiple criteria: by type, localization, shape, etc. In our work, we propose a convolutional network architecture for Multitask Learning on brain MRI images with heterogeneous labels and implement a corresponding image retrieval system. We demonstrate that it allows us to build representations that contain more relevant information about tumors than classification-based approaches. We also show that it is convenient to perform the search of similar tumors in the learned space.

Keywords: image retrieval, multitask learning, CNN, MRI

1 Introduction

In connection with the success of the development of deep neural networks, the performance of machine learning algorithms in image analysis tasks has increased [3]. This progress had a positive impact on the analysis of medical images, where convolutional networks are now used to automate a variety of time-consuming tasks [7]. For example, in [5] the authors developed a CNN-based approach to whole slide histology images classification. Another important topic is 3D MRI images analysis. In [6] a 3D-convolutional network is proposed to address the problem of human brain MRI images segmentation. A more comprehensive analysis of deep learning applications to medical imaging classification and segmentation can be found in [10].

⋆ Equal contribution

Deep convolutional networks are so effective due to, in part, their ability to automatically build an informative low dimensional representation of images space. An approach to building a retrieval system motivated by this feature is to train CNN on some machine learning problem and afterward to use extracted features as a measure of similarity. For example, in [1] authors train a Siamese-based CNN with a constrastive loss from comparing images of eyes for the presence of retinopathy. In [9] high-dimensional representations of prostate MRI images are built with the help of CNNs and then hashing forests are used to reduce the dimensionality of representations.

In our work, we focus on building a retrieval system for brain tumors. There are multiple ways in how the tumors might differ: by type, shape or localization in the brain. In addition to the class label of every tumor, we know it's contours and localization in the brain. The novelty of our work is that we utilize all these labels by training the network via multitask learning, simultaneously solving different machine learning problems. It allows us to build representations that contain more relevant information about the tumors.

2 Method

The main idea behind the method is to construct a function that maps the bounding box containing a tumor to a fixed-dimensional space, and then to treat the l_1 or l_2 distance in the resulting space as a measure of similarity between the tumors.

To do so, we will assume that each brain MR image x has a corresponding heterogeneous set of metadata: a segmentation mask y^s and various labels for each tumor $y^p_t, t \in T, p \in P$ such as tumor type, localization e.t.c. Here T is the set of tumors in the brain x and P is the set of labels available for a given tumor.

We build the mapping by training a neural network to solve multiple segmentation and classification tasks simultaneously. The network's architecture (Fig. 1) consists of two parts: a backbone CNN used to automatically extract features, and multiple heads each aimed at solving a particular task.

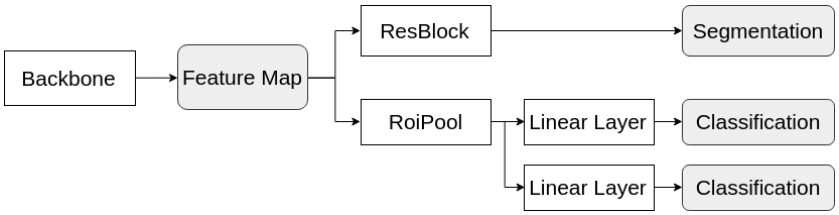

Fig. 1: Network architecture

Since we want the resulting tumor representation to be as informative as possible, the head architectures are designed to be very simple, to "motivate" the backbone to generate a comprehensive feature map.

2.1 Backbone

Fig. 2 shows our backbone architecture: it processes the input image at two different scales which is mainly inspired by state of the art neural networks for segmentation such as DeepMedic [6] and U-Net [8].

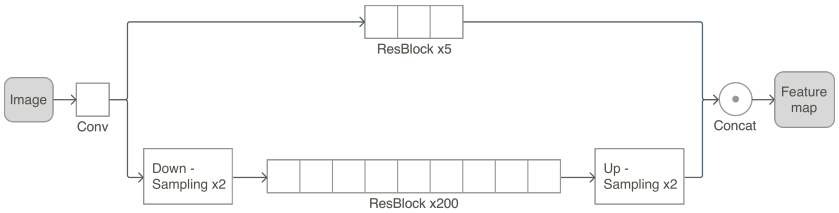

Fig. 2: Architecture of the backbone

2.2 Network's heads

Because the segmentation map is a global characteristic of the entire image, for the segmentation task we apply the head, which consists of a single ResBlock [4], to the entire feature map.

The tumor type and localization, however, is a property of a given image area containing a tumor, thus is a local characteristic. That's why for classification tasks we use a variation of the RoiPool block described in [2], followed by a single linear layer. In our implementation of RoiPool we take a spatial slice corresponding to the tumor's bounding box and apply global max pooling to it (Fig. 3). As a consequence, the output dimensionality is equal to the number of channels in the feature map generated by the backbone.

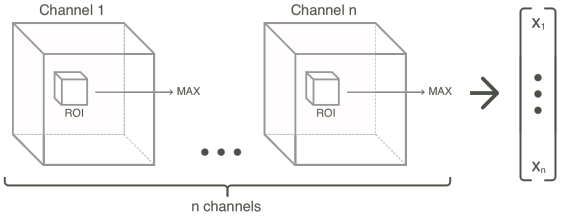

Fig. 3: Schematic representation of RoiPool

Note that RoiPool is essential for our method because our goal is to obtain fixed-size representations for tumors of different volumes.

2.3 Optimization criterion

For each head, we use cross-entropy as the loss function (binary or multiclass, depending on the task being solved). The final loss being optimized is a weighted sum:

$$L_{total} = \lambda_s L(\tilde{y}^s, y^s) + \sum_{p \in P} \lambda_p \sum_{t \in T} L(\tilde{y}_t^p, y_t^p) \quad (1)$$

where $L(\cdot, \cdot)$ is either binary or multiclass cross entropy, y^s is the true segmentation mask, y_t^p is the label for the tumor t and task p, $\tilde{y}^s, \tilde{y}_t^p$ are the corresponding predictions and λ_s, λ_p - the corresponding weights.

Note that in case of missing metadata y_t^p for a given tumor and task, the corresponding term $L(\tilde{y}_t^p, y_t^p)$ can be simply omitted from the final loss.

3 Experiments

3.1 Setup

We train the network for 120 epochs with 200 batches (of size 2) per epoch. Because of technical limitations, instead of feeding entire images into the network, we use patches of size $120 \times 120 \times 120$. The patches are picked at random with the sole condition that each selected patch contains a tumor.

In our experiments, we used stochastic gradient descent with Nesterov momentum and a decreasing learning rate: the initial learning rate - 0.1 was decreased by a factor of 10 each time the train loss plateaued.The weights in the final loss were chosen during the baseline selection stage based on the model's performance: we ended up using $\lambda_s = 1$ for segmentation and $\lambda_p = 10^{-3}$ for classification.

3.2 Data

The data for our experiments was provided by a radiosurgical center that conducts operations with Leksell Gamma Knife, treating about 400 patients every year. The dataset consist of 989 MRI images in the T1c modality with spatial pixel size $0.94 \times 0.94 \times 1mm^3$ and typical image shapes of $200 \times 230 \times 170$, see details in Tab. 1. To avoid overfitting, we use only one MRI per patient.

For each image a set of heterogeneous metadata is available: a binary mask of cancerous tissues and a set of labels (in rare cases some of them are missing) including lesion type (metastasis, meningioma, schwannoma), anatomical area such as *Region frontalis* or *Cerebellum* (in total we have 11 classes) and localization (left/right hemisphere).

	Metastasis	Meningioma	Schwannoma
# of images	399	339	251
# of tumors	1636	414	258

Table 1: Number of images and tumors in the dataset

3.3 Validation

We had a goal of building a tumor retrieval system that searches for tumors that are similar simultaneously by different criteria: by size, type, localization, etc. Having multiple criteria of similarity complicates the validation process. To perform validation, we use two approaches.

First, we check that tumor representation built by CNN contains relevant information about tumors. To do that, we train a classifier/regressor that maps tumor representations to some characteristic of the tumor-like type or size. We perform this evaluation on the test part of the dataset that was seen neither by the network nor by the additional models during training.

Second, we validate that l_1/l_2 distance between two representations is a reasonable way to measure the similarity of two tumors. We emulate the usage of the retrieval system by analyzing how characteristics of the few nearest neighbors match. To do that, we assess the probability that the tumor in the search output would have the same type/localization class and evaluate the average change of tumors' size in the search output. The search performance is evaluated only for the test part of the dataset, but we retrieve images both from the train and test.

4 Results

4.1 Representation analysis

We start by comparing the representations based on feature maps with different numbers of channels. We use logistic regression to predict the type of tumor, anatomical area, left/right location and linear regression for predicting linear size ($volume^{1/3}$) of the tumor. Tab. 2 shows that representations of higher dimensionality expectantly contain more information about some relevant characteristics. We couldn't go beyond 128 due to technical limitations. As a tradeoff between accuracy and training time, in the rest of our experiments we use 64 channels.

Next, we analyze the feasibility of using several heads in the network's architecture. We trained several networks excluding some problems from the loss function, see Tab. 3. The results show that logistic/linear regression performs better on representations obtained by training on all available heterogeneous metadata. This proves that the network is able to aggregate the additional information contained in tumor labels.

Dimensionality of tumor representations	Mean absolute error of linear size regression	Accuracy of tumor type classification	Accuracy of anatomical area classification	Accuracy of left/right location classification
128	**1.29**	0.92	0.61	0.8
64	1.59	**0.93**	**0.63**	**0.83**
32	1.71	**0.93**	0.58	0.73

Table 2: The quality of relevant characteristics predictions depending on tumor representation dimensionality.

Loss from tumor segmentation problem	Loss from classification problems	Mean absolute error of linear size regression	Accuracy of tumor type classification	Accuracy of anatomical area classification	Accuracy of left/right location classification
Yes	Tumor type, anatomical area, left/right location	**1.36**	**0.93**	0.58	**0.81**
Yes	Tumor type, anatomical area	1.38	0.91	0.56	0.59
No	Tumor type, anatomical area, left/right location	1.94	0.91	**0.66**	0.56
Yes	-	1.52	0.86	0.38	0.56
N/A	N/A	6.12	0.71	0.33	0.51

Table 3: The quality of prediction of relevant characteristics from the representation. First two columns describe the configuration of the CNN's heads and the overall loss function. The last 4 columns contain the quality of predictions. The last row shows the performance of a dummy model.

4.2 Retrieval evaluation

We started with the visual analysis decreasing the dimensionality of the representation to 2D using PCA. Figures 4, 5 demonstrate that the representations cluster by type and size.

Fig. 4: Tumor representations after applying *PCA*: the change of color from blue to red corresponds to change in linear size from smaller to bigger.

Fig. 5: Tumor representations after applying *PCA*: red, green and blue colors correspond to metastasis, meningioma and schwannoma tumors respectively.

Then we estimated the similarity between the reference tumor and the top results retrieved by the system averaged by 2,3 and 5 samples, see Tab. 4.

# of nearest neighbors	Relative difference of linear sizes			Similarity of tumor types			Similarity of anatomical areas		
distance function	l1	l2	cos	l1	l2	cos	l1	l2	cos
2	**0.193**	0.201	0.248	**0.906**	0.893	0.905	0.345	**0.354**	0.347
3	0.195	0.204	0.248	0.893	0.888	0.9	0.335	0.344	0.345
5	0.206	0.215	0.257	0.886	0.888	0.894	0.333	0.334	0.333
10	0.217	0.229	0.281	0.881	0.881	0.893	0.328	0.324	0.328
20	0.232	0.245	0.3	0.875	0.872	0.887	0.309	0.304	0.314
randomly picked tumor	1.019			0.563			0.114		

Table 4: Characteristics of nearest tumors by $l_1/l_2/cos$ distance. We use **accuracy** as a measure of labels similarity and the **mean absolute error** (divided by the reference linear size) as a measure of linear size similarity.

Finally we analyze the robustness of our system: in order to perform the search, a bounding box of the tumor is required, in a real world setting the provided bounding box might not be very accurate. To test the ability of the system to cope with such inaccuracies, we added significant distortions to the bounding boxes. Tab. 5 shows that though performance declines when the boxes are distorted, the system is still able to find relevant tumors.

# of nearest neighbors by l1 distance	Relative difference of linear sizes		Similarity of tumor types		Similarity of anatomical areas	
	n/d	w/d	n/d	w/d	n/d	w/d
2	**0.217**	0.237	**0.881**	0.875	**0.311**	0.304
3	0.227	0.244	0.874	0.864	0.312	0.291
5	0.238	0.265	0.87	0.86	0.289	0.286

Table 5: The performance of the search system after bounding box distortions. n/d - accurate bounding boxes, w/d - bounding boxes with distortion.

Fig. 6 shows two tumors proposed by the system for a randomly picked reference schwannoma: note the similarities such as localization and shape (the resulting tumors are from different patients).

5 Conclusion

We developed a retrieval system that searches for similar brain tumors. The system is based on multitask learning and takes into account heterogeneous metadata available for each dataset entry.

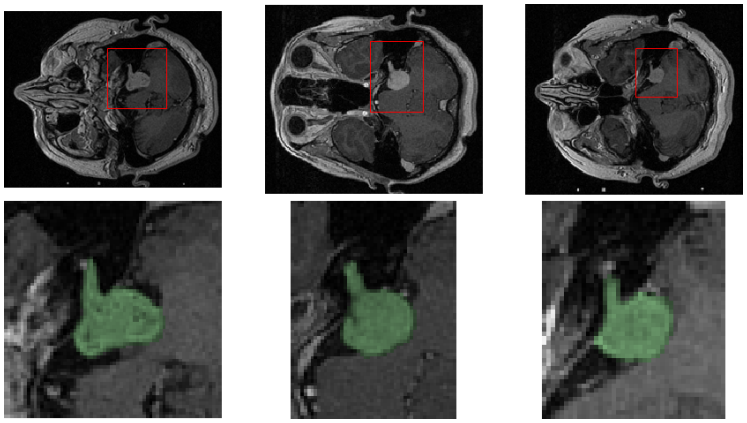

Fig. 6: Retrieval example: a reference schwannoma tumor (left) and two tumors retrieved by the system.

We demonstrated, that our method finds relevant tumors by analyzing various characteristics of lesions in the search output. Also we showed that the system can be used in the real-world setting by showing that it finds relevant tumors even after significant bounding box distortion.

References

1. Chung, Y.A., Weng, W.H.: Learning deep representations of medical images using siamese cnns with application to content-based image retrieval. arXiv preprint arXiv:1711.08490 (2017)
2. Girshick, R.: Fast r-cnn. In: Computer Vision (ICCV), 2015 IEEE International Conference on. pp. 1440–1448. IEEE (2015)
3. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al.: Recent advances in convolutional neural networks. Pattern Recognition (2017)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
5. Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H.: Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2424–2433 (2016)
6. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. Medical image analysis 36, 61–78 (2017)
7. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Laak, J.A.V.D., Ginneken, B.V., Snchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis 42, 6088 (2017)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)

9. Shah, A., Conjeti, S., Navab, N., Katouzian, A.: Deeply learnt hashing forests for content based image retrieval in prostate mr images. Medical Imaging 2016: Image Processing (2016)
10. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging 35(5), 12851298 (2016)