GRAYLIGHT
MEDICAL IMAGING SOFTWARE

Dealing with limited ground truth image data

by Graylight Imaging

graylight-imaging.com

Table of **contents**

What is ground truth and why is it important?

Ground truth is a term used primarily in statistics and Machine Learning that refers to a set of data that allows us to train the algorithm in a supervised way, and to determine the correctness of an algorithm - how accurately it captures the rules of the real world.

What makes ground truth different from a regular data set? The key word is labels (also referred to as annotations in image analysis), that is descriptions prepared by specialists – in the case of radiology, it could be radiologists or medical physicists. It is the annotations that allow us to be sure we are obtaining accuracy against the real world. Too few properly prepared data samples in the ground truth set may render the usefulness, and above all the accuracy, of our algorithm negligible[1]. Therefore, algorithms and applications based on them developed in this way very often fail to obtain the certification necessary for release into clinical practice, remaining at the research-only level, and ground truth itself becomes a critical priority issue

Why is there a problem with **building a good ground truth?**

As mentioned above, while smaller data sets may be sufficient to train an algorithm, high quality ground truth is an essential component for training, validating, and testing an algorithm with clinical application. The literature describes this requirement extensively; one need only refer to the well-known publication by Park and Han[2].

There are numerous challenges in acquiring sets of medical data. Industry literature has already repeatedly raised issues such as proprietary rights and the need to anonymize patient studies. However, this is not the end of the list. It must be said openly that healthcare systems are not technologically prepared to share large amounts of imaging data. These data are very often scattered in the hospital system, preparing them for sharing would require a huge organizational effort, and before that probably a regulatory structuring of the process. This problem is described in more detail in the publication by Willemink et al[3].

All these difficulties relate to the acquisition of the 'raw' data itself. In the case of ground truth, it is additionally necessary to manually or semi-automatically annotate each study, which in practice means the necessity of cooperation with the medical community.
This is a very time and cost consuming process.

---

1 Brodley C., Friedl M. Identifying mislabeled training data. Computer Science; J. Artif. Intell. Res. 1999;11(1):131-167.

2 Park, S. H. & Han, K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 286, 800–809 (2018).

3 Willemink M., Koszek W., Hardell C. et al. Preparing medical imaging data for machine learning. Radiology 295:4–15 (2020)

There is additional difficulty associated with the preparation of training and validation data in the case of algorithms operating on imaging studies performed relatively infrequently. On top of the general problems with access to data, there is the fact that the number of examinations performed, and thus the amount of data available to medical centers in general, is limited. As Kohli, Summers and Geis[1] point out, 'In the U.S., individual healthcare institutions may have 103 up to rarely 107 of an exam type,' and even so, they mostly represent 'common radiology exam types, for example, chest radiographs, unenhanced brain CTs, mammograms, and abdominal CTs'[2].

Anyone who has ever had to prepare a gound truth dataset knows how time-consuming and costly this task is. All the more reason to think about how we can use the right technology to expand the dataset we have and produce a high qualit y set.

In this paper, we describe two of many strategies that can come in handy while dealing with limited ground-truth image data. The examples we will show relate to a brain glioma segmentation algorithm we have been working on. It was on this occasion that we first had to deal with the topic of limited ground truth data.

Primary brain tumors account for about 2% of all malignancies. Malignant gliomas, which include glioblastomas and anaplastic astrocytomas, are the most common primary tumors of the brain. Probability of morbidity rise with age, and it is 50/100 000 at the age of 75 years old.

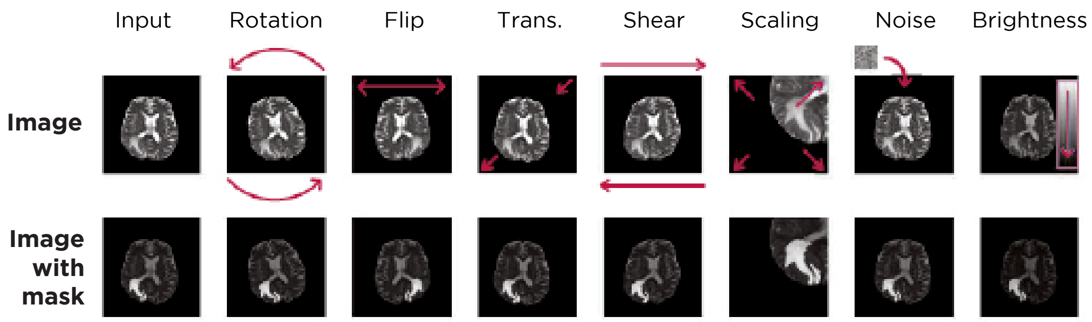
Data augmentation via affine transformation

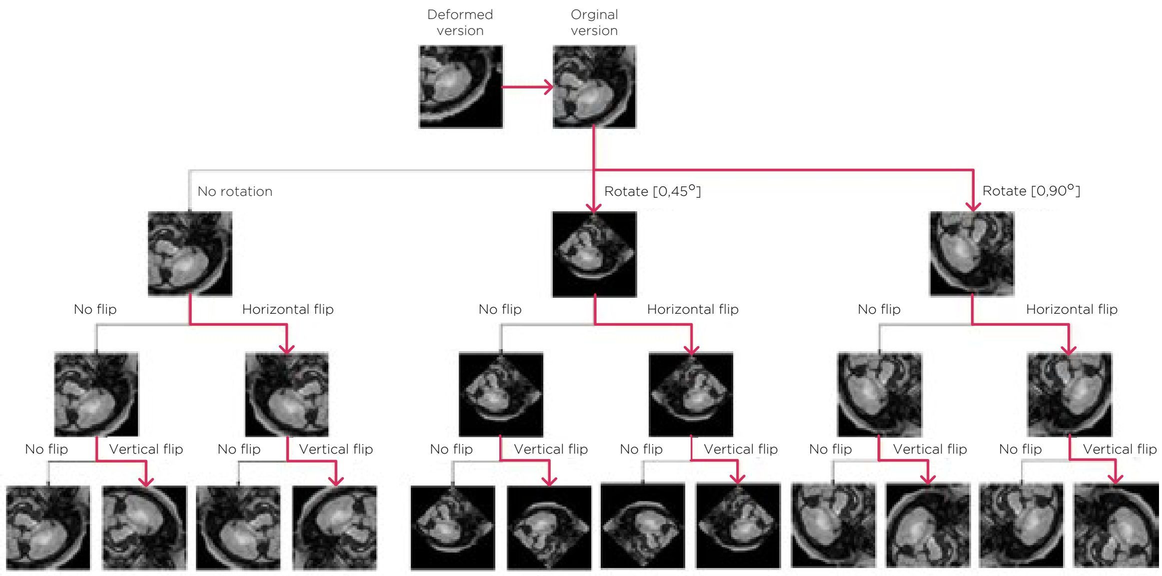
To tackle the problem of limited ground truth datasets, we can synthetically generate artificial examples based on the existing ones in the process of data augmentation.

Affine transformations produce fundamentally correlated images and can easily lead to anatomically incorrect training examples. In the affine approaches, existent image data undergo different operations (rotation, zooming, cropping, flipping, or translations) to increase the number of training examples. Such traditional data augmentation techniques fundamentally produce very correlated images, therefore it can offer very little improvements for the deep-network training process and future generalization over the unseen test data (such examples do not regularize the problem sufficiently). Nevertheless, affine image transformations are trivial to implement (in both 2D and 3D), they are fairly flexible (due to their hyper-parameters), and are widely applied in the literature.

1 Kohli, M., Summers, R., Geis, J. Medical Image Data and Datasets in the Era of Machine Learning-Whitepaper from the 2016 C-MIMI Meeting Dataset Session. J. Digit. Imaging 2017, 30, 392–399

2 Ibidem



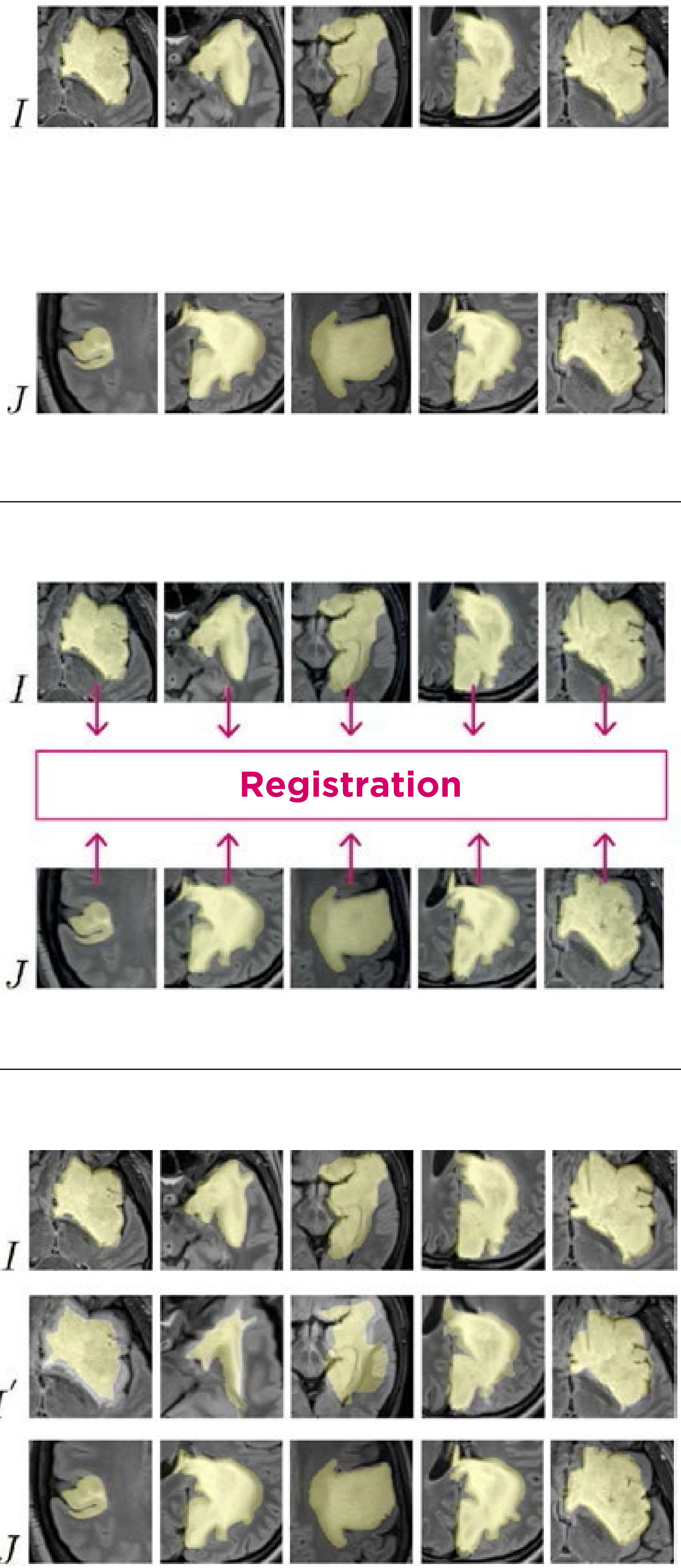


Data augmentation via diffeomorphic image registration

Our data augmentation via diffeomorphic image registration allows us to generate augmented training sets of any size that lead to better generalization of deep networks (here, UNets) in the brain tumor segmentation task. Diffeomophic mappings play an important role in brain imaging, as they are able to preserve topology and generate biologically plausible deformations [1].

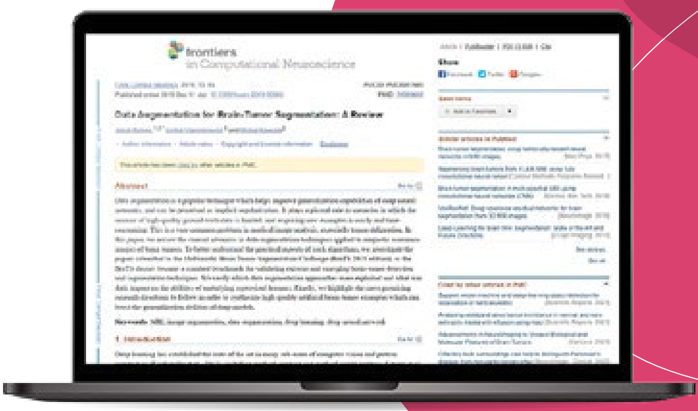
1 Nalepa J., Marcinkiewicz M., Kawulok M. Data Augmentation for Brain-Tumor Segmentation: A Review. Frontiers in Computational Neuroscience. 2019; 13: 83.

Diffeomorphic image registration applied to example brain images allowed for obtaining visually-plausible generated images: source (I), target (J), artificially generated (I') images.



Our review of data augmentation for brain tumor segmentation

We have published an extensive review of the data augmentation techniques targeting brain-tumor segmentation. There you will find a discussion of strategies such as Data Augmentation Using Pixel-Level Image Transformations or Data Augmentation by Generating Artificial Data. We also present a case study of how we tackled the need for data augumentation in practice using a specific example.



We have published an extensive review of the data augmentation techniques targetting brain tumor segmentation.

GO TO THE DOCUMENT

GRAYLIGHT
MEDICAL IMAGING SOFTWARE

About us

By developing advanced machine learning algorithms and AI quick - start modules **we create the possibility to automate and support the diagnostic process.**

At Graylight Imaging we believe that by combining a scientific approach, collaboration with medical experts and clinics around the world and high technological skills we are changing modern medicine.

So, if you have questions about our solutions - let’s talk.

Get in touch

Graylight Imaging
ul. Bojkowska 37a
44-100 Gliwice
Poland

+48 609 995 887
contact@graylight-imaging.com