

## АУГМЕНТАЦИЯ ДАННЫХ ПРИ РЕШЕНИИ ЗАДАЧИ СЕГМЕНТАЦИИ ОБЪЕКТОВ НА МЕДИЦИНСКИХ СНИМКАХ

*Холод Данила Витальевич*

*Магистрант*

*НИЯУ МИФИ, Москва, Россия*

*E-mail: danila.kholod@gmail.com*

*Научный руководитель — Зайцев Константин Сергеевич*

Ультразвуковая диагностика является одним из основных методов выявления узлов щитовидной железы благодаря своей доступности и неинвазивности. Однако интерпретация УЗИ-изображений зависит от опыта врача и сопровождается субъективностью, особенно при определении границ новообразований. Это делает актуальной задачу автоматической сегментации узлов с применением методов глубокого обучения.

Современные модели сегментации демонстрируют высокую точность [1], однако их обучение требует значительных объемов размеченных данных, получение которых в медицинской практике затруднено. Одним из способов компенсации ограниченного датасета является аугментация — расширение обучающей выборки путём применения преобразований к исходным изображениям [2,3]. Вместе с тем чрезмерная аугментация способна исказить распределение данных и ухудшать качество модели, что обуславливает необходимость анализа её оптимальной интенсивности.

В работе исследуется влияние доли аугментированных данных на качество сегментации узлов щитовидной железы. Анализ проводится для различных архитектур энкодеров: ConvNeXt-Tiny, Swin Transformer-Tiny и ViT-Base [4,5,6].

Рассматривается задача бинарной сегментации: задано множество изображений и масок

$$\{(x_i, y_i)\}_{i=1}^N,$$

где  $x_i$  — УЗИ-изображение,  $y_i$  — бинарная маска узла. Аугментация трактуется как отображение

$$\mathcal{A} : x_i \rightarrow \{x_i^{(j)}\}_{j=1}^M.$$

Параметр

$$R = \frac{M}{N}$$

характеризует долю синтетических данных в обучающей выборке. Качество сегментации оценивается с использованием метрик IoU и Dice. Пусть  $P \subset \Omega$  — множество пикселей, отнесённых моделью к области узла, а  $G \subset \Omega$  — множество пикселей истинной разметки, где  $\Omega$  — множество всех пикселей изображения. Тогда

$$IoU(P, G) = \frac{|P \cap G|}{|P \cup G|}, \quad Dice(P, G) = \frac{2|P \cap G|}{|P| + |G|}.$$

Результаты демонстрируют немонотонную зависимость качества от  $R$ : для всех моделей существует оптимальное значение коэффициента аугментации, при котором достигается максимум метрик. При дальнейшем увеличении  $R$  наблюдается деградация качества вследствие смещения распределения обучающих данных. ConvNeXt-Tiny показывает наибольшую устойчивость к росту  $R$ , Swin Transformer-Tiny — промежуточную чувствительность, тогда как ViT-Base наиболее подвержен ухудшению при избыточной аугментации.

### Литература

1. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015. P. 234–241.
2. Shorten C., Khoshgoftaar T. A survey on Image Data Augmentation for Deep Learning // Journal of Big Data. 2019. Vol. 6. Art. 60.
3. Kim M., Bae H.-J. Data Augmentation Techniques for Deep Learning-Based Medical Image Analyses // Journal of the Korean Society of Radiology. 2020. Vol. 81, No. 6. P. 1290–1304.
4. Dosovitskiy A., Beyer L., Kolesnikov A., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // arXiv. 2020. arXiv:2010.11929.
5. Liu Z., Lin Y., Cao Y., et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021. P. 10012–10022.
6. Liu Z., Mao H., Wu C.-Y., Feichtenhofer C., Darrell T., Xie S. A ConvNet for the 2020s // arXiv. 2022. arXiv:2201.03545.