

ОБЪЯСНЕНИЕ СЕГМЕНТАЦИИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ GRAD-CAM: ИССЛЕДОВАНИЕ НА НАБОРЕ ДАННЫХ DRIVE

Амируш Анис (КФУ)

Научный руководитель – кандидат физико-математических наук, доцент Арабов М.К. (КФУ)

Введение. Сегментация медицинских изображений играет ключевую роль в анализе изображений глазного дна для выявления заболеваний, таких как диабетическая ретинопатия. Несмотря на то, что модели глубокого обучения достигают высокой точности при сегментации медицинских изображений, их «черный ящик» вызывает опасения по поводу доверия, интерпретируемости и надежности. Методы объяснимого ИИ (XAI) помогают преодолеть этот разрыв, предоставляя информацию о решениях модели. В данной статье мы рассматриваем применение одного XAI-алгоритма для объяснения результатов сегментации модели глубокого обучения на наборе данных DRIVE.

Основная часть. Набор данных Digital Retinal Images for Vessel Extraction (DRIVE) является широко используемым эталоном для сегментации сосудов сетчатки [1]. Он состоит из 40 изображений глазного дна, разделенных на обучающую и тестовую выборки, с вручную размеченными экспертами эталонными сегментациями. Учитывая его клиническую значимость, точная и интерпретируемая сегментация сосудов сетчатки имеет решающее значение.

В данном исследовании используется модель U-Net — популярная архитектура сверточных нейронных сетей (CNN), применяемая для сегментации медицинских изображений. Модель обучается на задаче сегментации кровеносных сосудов сетчатки по изображениям глазного дна [2]. Однако, несмотря на высокую точность сегментации, U-Net не предоставляет встроенных объяснений своих предсказаний [3].

Для повышения интерпретируемости мы применяем алгоритм объяснимого ИИ Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM создает визуальные объяснения, выделяя области входного изображения, которые вносят наибольший вклад в предсказания модели. Метод вычисляет градиенты целевого класса по отношению к картам признаков последнего сверточного слоя, усредняет их для получения весов значимости, а затем использует эти веса для генерации тепловой карты, подчеркивающей наиболее значимые области изображения [4]. Применяя Grad-CAM к модели U-Net, мы можем визуализировать, на какие области модель обращает внимание при предсказании структуры сосудов. Это помогает выявлять возможные предвзятости, ошибки сегментации или участки, требующие доработки модели.

Мы применяем Grad-CAM к нескольким тестовым изображениям из набора данных DRIVE. Полученные тепловые карты показывают, что модель преимущественно фокусируется на участках с высокой контрастностью сосудов, но испытывает трудности с более тонкими сосудистыми структурами. Объяснения также выявляют потенциальные ошибки классификации, например, ложноположительные предсказания в областях с текстурой, похожей на кровеносные сосуды. Тепловые карты Grad-CAM хорошо совпадают с основными сосудистыми структурами, но могут быть менее информативными для тонких сосудов [5]. Иногда модель ошибочно выделяет несосудистые области из-за сходства текстуры и цвета [6].

Выводы. Методы объяснимого ИИ, такие как Grad-CAM, играют важную роль в анализе медицинских изображений, обеспечивая прозрачность и доверие к моделям глубокого обучения. Применяя Grad-CAM к сегментации изображений глазного дна моделью U-Net на наборе данных DRIVE, мы получаем ценные сведения о процессе принятия решений моделью, что в конечном итоге способствует созданию более надежных

и интерпретируемых систем ИИ для диагностики заболеваний сетчатки.

В будущих исследованиях можно рассмотреть применение других методов ХАИ, таких как SHAP или LIME, для дальнейшего повышения интерпретируемости и устойчивости моделей сегментации медицинских изображений.

Список использованных источников:

1. Goodfellow I., Bengio Y., Courville A. Deep Learning. – MIT Press, 2016. – P. 209–211.
2. Amirouche A., Arabov M.K. Comparative analysis of FCN and U-Net for retinal blood vessels segmentation: A performance evaluation // 2024 International Russian Automation Conference (RusAutoCon), Sochi, Russian Federation. – 2024. – P. 89–94. – <https://doi.org/10.1109/RusAutoCon61949.2024.10694438>.
3. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // arXiv preprint arXiv:1505.04597. – 2015.
4. Selvaraju R.R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization // Proceedings of the IEEE International Conference on Computer Vision. – 2017. – P. 618–626.
5. Smilkov D., Thorat N., Kim B., Viégas F., Wattenberg M. SmoothGrad: removing noise by adding noise // arXiv preprint arXiv:1706.03825. – 2017.
6. Staal J., Abramoff M.D., Niemeijer M., Viergever M.A., van Ginneken B. Ridge-based vessel segmentation in color images of the retina // IEEE Transactions on Medical Imaging. – 2004. – № 23(4). – P. 501–509.