

УДК 004.8

## ОБ ОДНОМ МЕТОДЕ ПОЛУЧЕНИЯ ВИЗУАЛЬНОГО ОБЪЯСНЕНИЯ АВТОМАТИЧЕСКИ СГЕНЕРИРОВАННОГО ТЕКСТОВОГО ОПИСАНИЯ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Волков Е.Н. (Государственный университет «Дубна»)

Научный руководитель – кандидат физико-математических наук, доцент Аверкин А.Н.  
(Государственный университет «Дубна»)

**Введение.** Задача автоматического описания изображений медицинского домена является одной из наиболее сложных среди всех задач описания изображений (image captioning, image report generation), что объясняется как различным форматом их представления и многообразием типов, так и сложностью получения наборов данных из-за необходимости привлекать экспертов-клиницистов для их разметки [1]. Учитывая априорную непрозрачность любых программных решений, основанных на использовании нейронных сетей различных архитектур и типов, являющихся моделями «чёрного ящика», появляется необходимость получения объяснения предсказания модели. Использование методов объяснительного искусственного интеллекта (от англ. eXplainable Artificial Intelligence) помогает получить объяснение (визуальное, текстовое) как дополнительный вывод модели. Однако, при такой постановке задачи выводится объяснение целевых классов, что приемлемо для задач классификации, сегментации и детекции, но сомнительно для задач описания изображения. Изменение постановки задачи в сторону получения объяснения в виде визуального вывода маски значимости для каждого из слов описания может дать конечному пользователю большую уверенность в правильности полученного предсказания.

**Основная часть.** Исследование и программная реализация представленной задачи стала возможно благодаря применению комплексного подхода, заключающегося во внимании не только к выбору модели (типа и архитектуры нейронной сети), но и к данным для её обучения.

На первом этапе, после выбора набора данных IU X-Ray [2], представляющем собой пары изображений и текстовых описаний, была проведена оценка его качества этих пар. Отдельные нерелевантные элементы описания были удалены, а изображения подверглись стандартизации размеров и улучшению контрастности.

На втором этапе, были дообучены несколько предобученных моделей BLIP с различным набором параметров [3]. Оценка качества работы моделей производилась с использованием метрик BLEU-avg, CIDEr, ROUGE.

На третьем этапе, в качестве метода визуального объяснения результата генерации текстового описания, был использован Grad-CAM, оценивающий вклад регионов пикселей в конечное предсказание модели [4].

**Выводы.** Результаты реализации использования метода интерпретации Grad-CAM для получения визуального объяснения текстового описания медицинских изображений расширяет возможности использования данной технологии за счёт возможности выделения тепловой картой регионов интереса связанных непосредственно с конкретными словами текстового описания, а не только с искомым классом.

### Список использованных источников:

1. Pang T., Li P., Zhao L. A survey on automatic generation of medical imaging reports based on deep learning //BioMedical Engineering OnLine. – 2023. – vol. 22. – no. 1. – pp. 1-16. DOI: 10.1186/s12938-023-01113-y.
2. Chest X-rays (Indiana University) // Kaggle URL: <https://www.kaggle.com/datasets/raddar/chest-xrays-indiana-university?resource=download> (дата обращения: 09.01.2024).

3. Li J. et al. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation //International Conference on Machine Learning. – PMLR, 2022. – P. 12888-12900.
4. Selvaraju R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization //Proceedings of the IEEE international conference on computer vision. – 2017. – P. 618-626.