МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра физики открытых систем

Сравнительный анализ методов визуализации свёрточных искусственных нейронных сетей в задаче оценки возраста

название темы выпускной квалификационной работы полужирным шрифтом

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

Студента 2 курса 2241 группы
направления 09. 04. 02. «Информационные системы технологии»
код и наименование направления (специальности)
Института физики
наименование факультета, института, колледжа
Хромова Павла Игоревича
фамилия, имя, отчество

Научный руководитель
Профессор кафедры ФОС,
д.ф.-м.н., профессор _____ _____ О.И. Москаленко
должность, уч. степень, уч. звание дата, подпись инициалы, фамилия

Заведующий кафедрой
физики открытых систем,
д.ф.-м.н., профессор _____ _____ А.А. Короновский
должность, уч. степень, уч. звание дата, подпись инициалы, фамилия

Саратов 2026 год

Введение

Искусственные нейронные сети широко применяются для анализа изображений, в том числе в биометрических задачах и оценке возраста по лицу. Однако высокая точность предсказаний не показывает, за счёт каких областей изображения было получено решение. Поэтому для таких моделей требуется интерпретация результата и анализ зон внимания [1].

Целью данной работы является сравнительный анализ методов визуализации внимания свёрточных искусственных нейронных сетей при решении задачи оценки возраста по изображениям лиц, а также выявление их особенностей, преимуществ и ограничений с точки зрения интерпретируемости.

Для достижения цели рассматриваются особенности оценки возраста по лицевым изображениям, методы Grad-CAM, Integrated Gradients, LIME и RISE; для предсказания возраста используется модель MobileNetV2; по выборкам изображений строятся карты значимости и рассчитываются метрики для их сравнения.

Научная новизна работы состоит в разработке единого подхода к сравнительному анализу методов визуализации внимания свёрточной нейронной сети в задаче оценки возраста по лицу, включающего визуальное сопоставление карт значимости, их количественное сравнение и содержательную интерпретацию выделяемых областей лица.

Структура основной части работы состоит из пяти глав. Первая глава – «Предметная область оценки возраста по изображению лица» – раскрывает постановку задачи и ограничения нейросетевой оценки возраста. Вторая глава – «Методы визуализации значимых областей изображения» – посвящена Grad-CAM, Integrated Gradients, LIME и RISE. Третья глава – «Методика экспериментального исследования» – описывает выбор данных, модель, предобработку входных данных и метрики. Четвёртая глава – «Реализация программной части эксперимента» – содержит описание программной реализации. Пятая глава – «Проведение эксперимента и

сравнительный анализ результатов» – представляет предсказания, тепловые карты и сводные метрики.

Основное содержание работы

Оценка возраста по изображению лица рассматривается как регрессионная задача: модель получает лицевое изображение и формирует одно числовое значение возраста. Качество предсказания оценивается абсолютной ошибкой, однако эта ошибка не показывает, какие признаки использовала нейронная сеть.

Визуальная оценка возраста связана с областью глаз, текстурой кожи, морщинами, нижней частью лица и контуром лица; такие признаки отражают изменения кожи, мягких тканей и объёмно-пространственной структуры лица [2, 3]. При этом на изображении присутствуют фон, волосы, очки, головные уборы и другие элементы, которые не являются собственно возрастными признаками, но могут влиять на работу модели из-за побочных корреляций.

В исследовании рассматриваются методы визуального объяснения решений свёрточных искусственных нейронных сетей, позволяющие определить области изображения, наиболее значимые для предсказания возраста. Grad-CAM связывает итоговый результат с активациями выбранного свёрточного слоя и соответствующими градиентами выхода модели [4]. Integrated Gradients рассчитывает атрибуцию входных признаков на основе накопленного значения градиентов [5]. LIME строит локальное приближение поведения модели, используя разбиение изображения на суперпиксели [6]. RISE, в отличие от градиентных методов, применяет набор случайных масок и оценивает вклад участков изображения по изменению ответа модели [7].

В ходе эксперимента была обучена модель MobileNetV2, модифицированная для возрастной регрессии [8]. После обучения для выбранных изображений были вычисляются предсказания возраста, строятся

карты значимости и рассчитываются метрики FaceRatio, BackgroundRatio, Entropy и PointingHit. Значение средней абсолютной ошибки на валидационной выборке составило около 7,1 года.

Перед построением карт значимости модель была применена к двум группам изображений. Первая группа включает четыре изображения UTKFace, вторая – три реальные фотографии. Результаты предсказаний приведены в табл. 1.1 и табл. 1.2:

Таблица 1.1 – Результаты предсказаний возраста для изображений из UTKFace

№	Исходное изображение	Предсказанный возраст	Истинный возраст	Абсолютная ошибка
1		12,8 лет	7 лет	5,8 лет
2		14,4 года	18 лет	3,6 лет
3		35,7 лет	26 лет	9,7 лет
4		35,7 лет	53 года	5,5 лет

Таблица 1.2 – Результаты предсказаний возраста для реальных фотографий

№	Исходное изображение	Предсказанный возраст	Истинный возраст	Абсолютная ошибка
5		17,6 лет	20 лет	2,4 лет
6		29,2 лет	20 лет	9,2 лет
7		21,4 лет	31 год	9,6 лет

На следующем этапе для тех же фотографий строятся визуализации зон внимания. Для изображений UTKFace результаты представлены в табл. 2.1, для реальных фотографий – в табл. 2.2:

Таблица 2.1 – Применение методов визуализации к изображениям из UTKFace

№	Исходное изображение	Grad-CAM	IG	LIME	RISE
1					
2					
3					
4					

Таблица 2.2 – Применение методов визуализации к реальным фотографиям

№	Исходное изображение	Grad-CAM	IG	LIME	RISE
5					
6					
7					

Количественное сравнение карт значимости путём вычисления среднего значения ранее упомянутых метрик выполнено в табл. 3.1 для первой группы фотографий и табл. 3.2 – для второй. Метрика FaceRatio показывает долю значимости внутри области лица, BackgroundRatio – вне лица, Entropy характеризует распределённость карты, а PointingHit показывает, попадает ли максимум значимости в область лица.

Таблица 3.1 — Средние значения метрик для изображений из UTKFace

Метод	FaceRatio	BackgroundRatio	Entropy	PointingHit
Grad-CAM	0,636	0,364	0,972	1,000
Integrated Gradients	0,675	0,325	0,961	0,750
LIME	0,675	0,325	0,956	0,750
RISE	0,564	0,436	0,994	0,000

Таблица 3.2 — Средние значения метрик для реальных фотографий

Метод	FaceRatio	BackgroundRatio	Entropy	PointingHit
Grad-CAM	0,598	0,402	0,970	0,667
Integrated Gradients	0,705	0,295	0,960	1,000
LIME	0,696	0,304	0,960	1,000
RISE	0,532	0,468	0,994	0,333

По данным табл. 3.1 для изображений UTKFace Grad-CAM во всех четырёх случаях обеспечил попадание максимума значимости в область лица, но уступил Integrated Gradients и LIME по отдельным значениям FaceRatio. RISE чаще давал более размытые карты и нулевой PointingHit для фотографий UTKFace.

Для реальных фотографий, по данным табл. 3.2, наибольшая доля значимости внутри лица чаще наблюдалась у Integrated Gradients и LIME. При этом дополнительные элементы изображения, такие как головной убор,

очки и особенности фона, заметнее влияли на распределение значимости, чем на более стандартизированных изображениях UTKFace.

Сравнение методов показывает, что Grad-CAM удобен для общей локализации значимых зон, Integrated Gradients даёт более детальные карты, LIME наглядно связывает результат с крупными сегментами изображения, а RISE не требует доступа к внутренним слоям сети, но формирует более сглаженные и вычислительно затратные объяснения.

Визуальное сопоставление табл. 2.1 и табл. 2.2 показывает, что разные методы могут объяснять одно и то же предсказание по-разному. Поэтому интерпретация не должна восприниматься как единственное доказательство корректности решения модели. Она служит средством проверки: карта значимости позволяет увидеть, согласуется ли поведение сети с предметным ожиданием, согласно которому возраст должен определяться преимущественно по лицевым признакам.

Наличие в таблицах исходных фотографий и тепловых карт важно для проверки результата. Числовая метрика показывает обобщённое свойство карты, но без изображения невозможно понять, какая именно область была выделена. Поэтому визуализации и данные о метриках дополняют друг друга.

Ограниченность контрольной выборки не позволяет распространять полученные численные результаты на все возможные случаи применения методов визуализации. Поэтому значения метрик следует рассматривать как характеристику их поведения в рамках проведённого эксперимента.

Заключение

В работы был выполнен сравнительный анализ методов визуализации значимых областей свёрточной нейронной сети в задаче оценки возраста по изображению лица. Была использована модель MobileNetV2, адаптированная под возрастную регрессию, и построены карты значимости для двух групп фотографий.

Полученные результаты показали, что методы интерпретации дают разные по характеру объяснения одного и того же предсказания. Точность возрастной регрессии и качество визуального объяснения необходимо рассматривать отдельно, поскольку небольшая ошибка предсказания не гарантирует, что модель использует только содержательно релевантные лицевые признаки.

Практическая значимость работы состоит в том, что визуализация внимания может использоваться как дополнительный инструмент проверки нейросетевой модели. Такой анализ помогает выявлять смещение значимости на нерелевантные области изображения.

Список литературы

[1] Arrieta A. B., Diaz-Rodriguez N., Del Ser J. et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI // Information Fusion. 2020. Vol. 58. P. 82-115.

[2] Farage M. A., Miller K. W., Elsner P., Maibach H. I. Characteristics of the Aging Skin // Advances in Wound Care. 2013. Vol. 2, № 1. P. 5-10.

[3] Coleman S. R., Grover R. The anatomy of the aging face: volume loss and changes in 3-dimensional topography // Aesthetic Surgery Journal. 2006. Vol. 26, № 1S. P. S4-S9.

[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 // IEEE International Conference on Computer Vision. 2017. P. 618-626.

[5] Sundararajan M., Taly A., Yan Q. Axiomatic Attribution for Deep Networks // International Conference on Machine Learning. 2017. P. 3319-3328.

[6] Ribeiro M. T., Singh S., Guestrin C. Why Should I Trust You? Explaining the Predictions of Any Classifier // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 1135-1144.

[7] Petsiuk V., Das A., Saenko K. RISE: Randomized Input Sampling for Explanation of Black-box Models // British Machine Vision Conference. 2018.

[8] Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. P. 4510-4520.