



САМАРСКИЙ УНИВЕРСИТЕТ  
SAMARA UNIVERSITY

# Применение алгоритмов машинного обучения для задач захвата движения человека на видеоизображении

Студент группы Б409-0103020,  
Сорока Александр Александрович

Научный руководитель:  
Савельев Дмитрий Андреевич, к.ф.-м.н., доцент  
кафедры технической кибернетики

Самара 2024



Цель данной работы – исследовать применимость модели машинного обучения, способную с высокой точностью определять ключевые точки человеческого тела на изображениях и видео.

Задачи, подлежащие решению в работе:

- проведение аналитического обзора научной литературы по тематике решения задачи захвата движений, включая описание методов с маркерами и без маркеров;
- исследование алгоритмов машинного обучения и их сравнительный анализ для применения в задачах захвата движения;
- реализация алгоритма для определения двумерных ключевых точек на основе выбранных архитектур моделей, оптимизаторов и методов обучения
- проведение аналитического обзора научной литературы по применению методов преобразования двумерных ключевых точек в трёхмерные координаты
- проведение экспериментов и анализ полученных результатов для моделей с различными архитектурами, оценка точности и полноты моделей на основе ключевых метрик





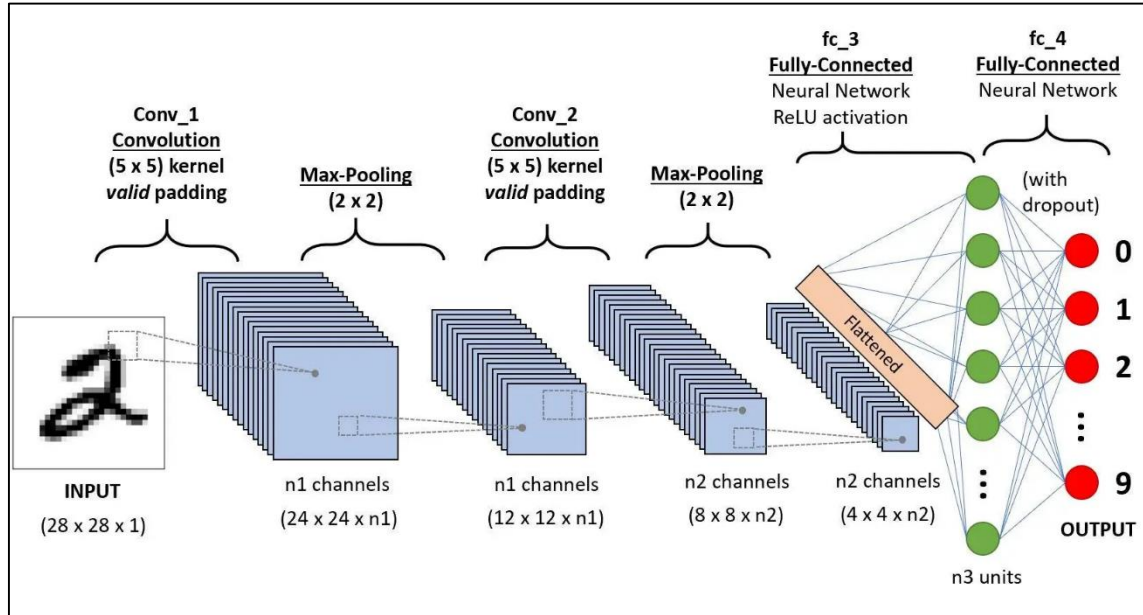
Технология Motion Capture (маркерная) [1]

Пусть  $I$  обозначает кадр из видеопотока, а  $P = (p_1, p_2, \dots, p_n)$  – набор ключевых точек, которые необходимо определить. Каждая точка  $p_i$  описывается своими координатами на изображении  $(x_i, y_i)$ .

- Задача алгоритма машинного обучения – максимизировать вероятность правильного определения координат на основе обучающих данных с аннотированными изображениями.
- Процесс обучения модели заключается в минимизации функции потерь, оценивающей разницу между предсказанными и истинными координатами точек.

1. Common Problems in Motion Capture // Axis : [сайт]. – 2024.





Пример архитектуры CNN [2]

Backbone	$AP$	$AP_{50}$	$AP_{75}$	$AP_M$	$AP_L$
ResNet-50	70,4	88,6	78,3	67,1	77,2
Res2Net-50	71,5	89,0	79,3	68,2	78,4
ResNet-101	71,4	89,3	79,3	68,1	78,1
Res2Net-101	72,2	89,4	79,8	68,9	79,2
Res2Net- vlb-50	72,2	89,5	79,7	58,5	79,4
Res2Net- vlb-101	73,0	89,5	80,3	69,5	80,0

Эффективность оценки ключевых точек на валидационном наборе COCO [3]

Softmax функция для классификации объектов

$$p(c|r) = \frac{e^{s_c}}{\sum_{c'=1}^C e^{s_{c'}}}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

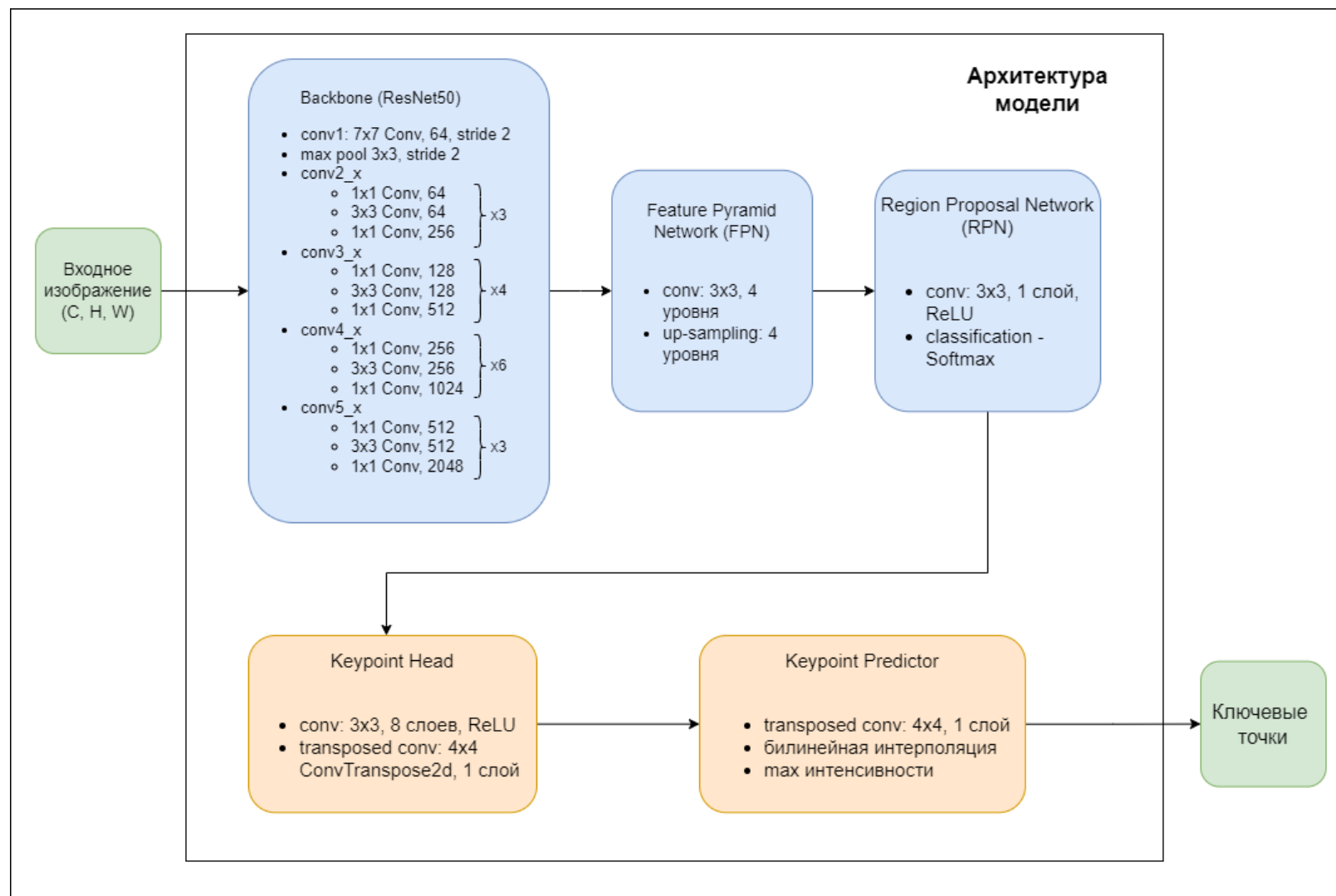
$$AR = \frac{1}{N} \sum_{i=1}^N Recall(i)$$

$$AP = \sum_{i=1}^{N-1} (Recall(i+1) - Recall(i)) \cdot Precision(i)$$

2. Романов, А. А. Сверточные нейронные сети // Научные исследования: ключевые проблемы III тысячелетия : сборник научных трудов по материалам XXI Международной научно-практической конференции, Москва, 09–10 января 2018 года.

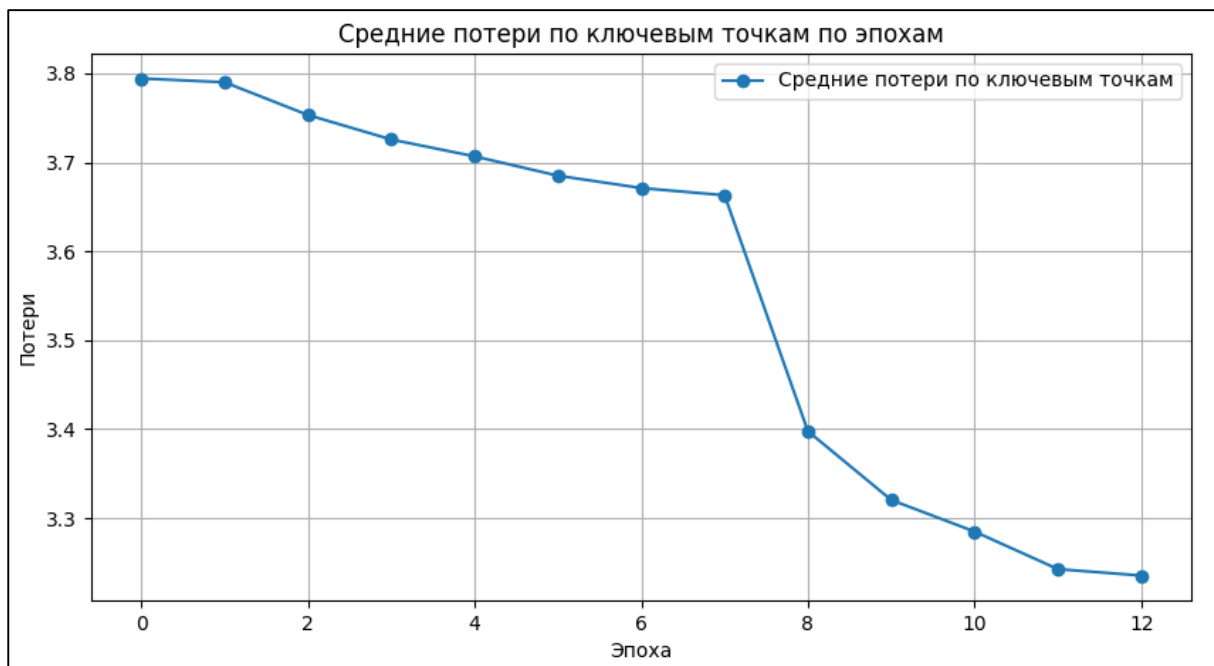
3. Res2net: A new multi-scale backbone architecture / S. Gao [et al.] // IEEE transactions on pattern analysis and machine intelligence. – 2019. – Vol. 43, Number 2. – P. 652-662.



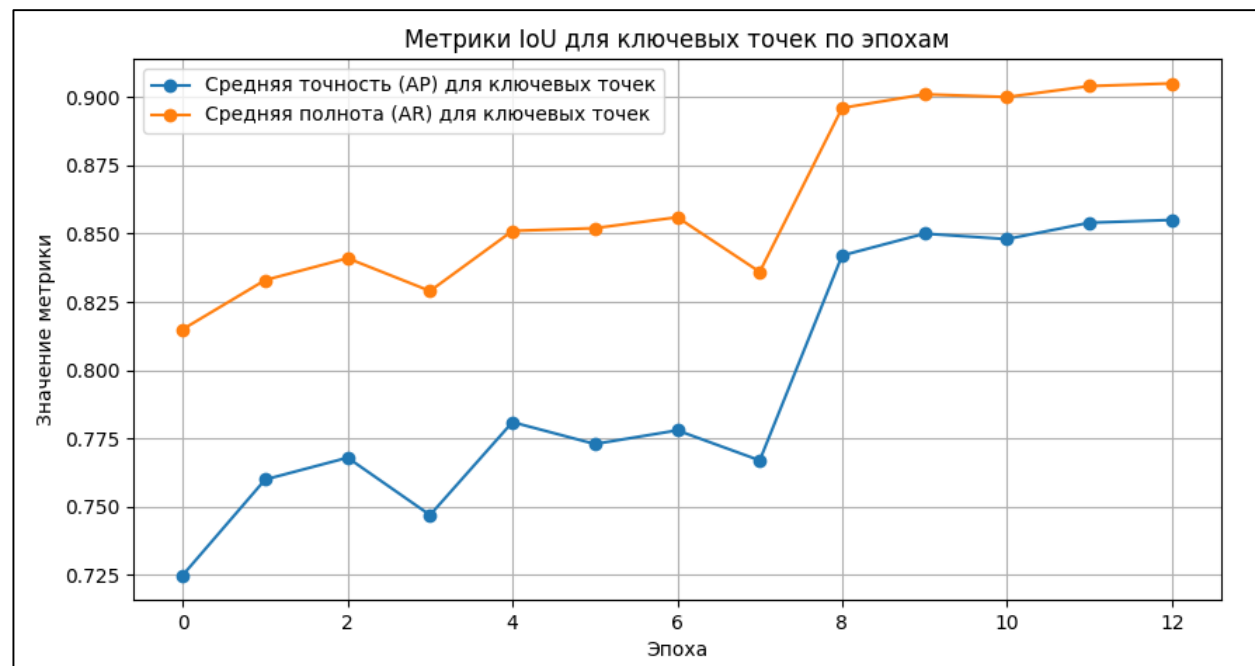




## Метрики обучения собственной модели на ResNet50



Изменение средней потери по ключевым точкам по эпохам обучения



Изменения средней точности и средней полноты для ключевых точек по мере обучения модели

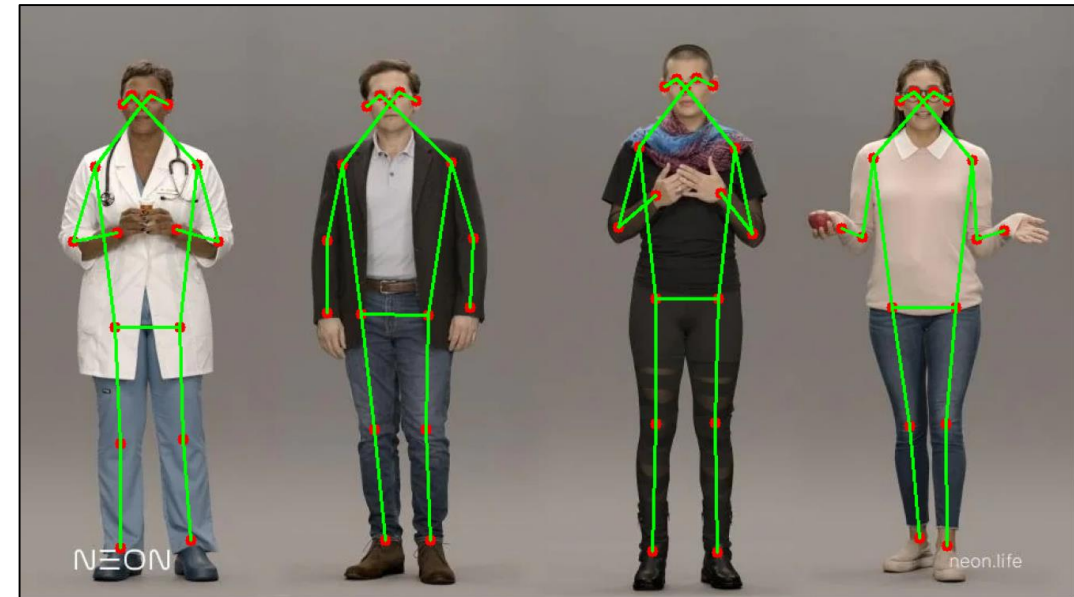






Backbone	$AP_{50}$	$AP_{75}$	$AR_{50}$	$AR_{75}$	Время обучения, ч
ResNet18	82,2	65,1	88,7	72,4	21
ResNet50	85,5	69,7	90,5	76,0	35
ResNet101	84,2	67,0	89,6	74,0	38

Результаты, полученные после обучения моделей на полном датасете COCO с различными архитектурами

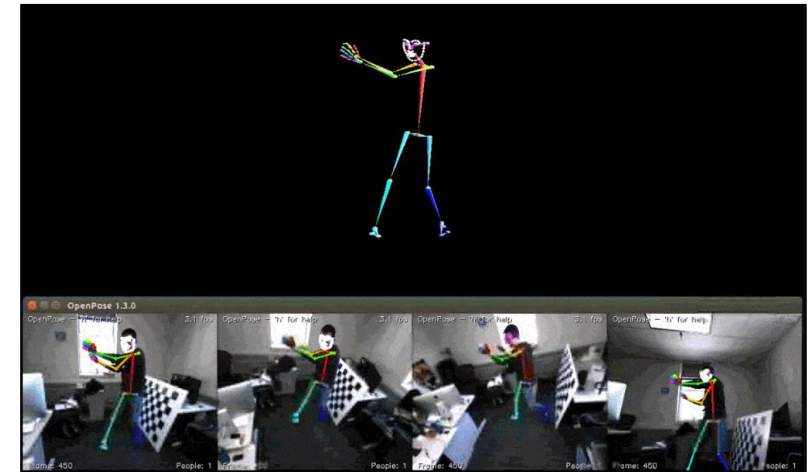


Результат работы собственной модели для определения двумерных ключевых точек с Backbone ResNet50

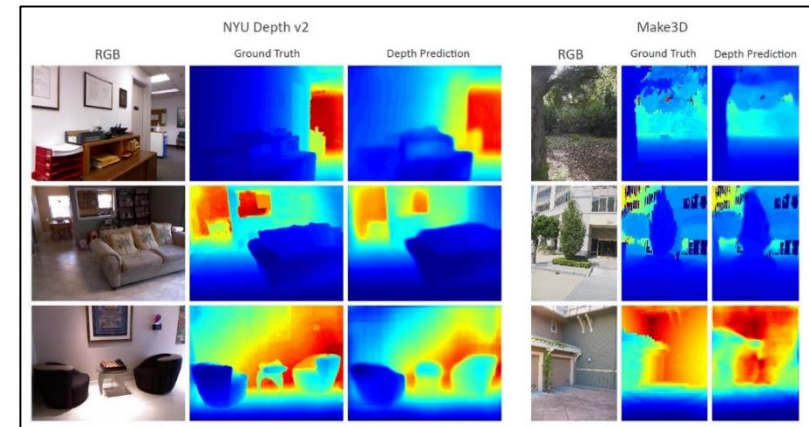




- трёхмерная реконструкция по нескольким изображениям;
- одиночное изображение с использованием глубины;
- использование датасета, обогащенного данными о глубине изображения.



Трёхмерная реконструкция по нескольким изображениям [4]



4. OpenPose Advanced Doc - 3-D Reconstruction Module and Demo // GitHub : [сайт]. – 2022.

Карта глубины, полученная с помощью глубоких нейронных сетей



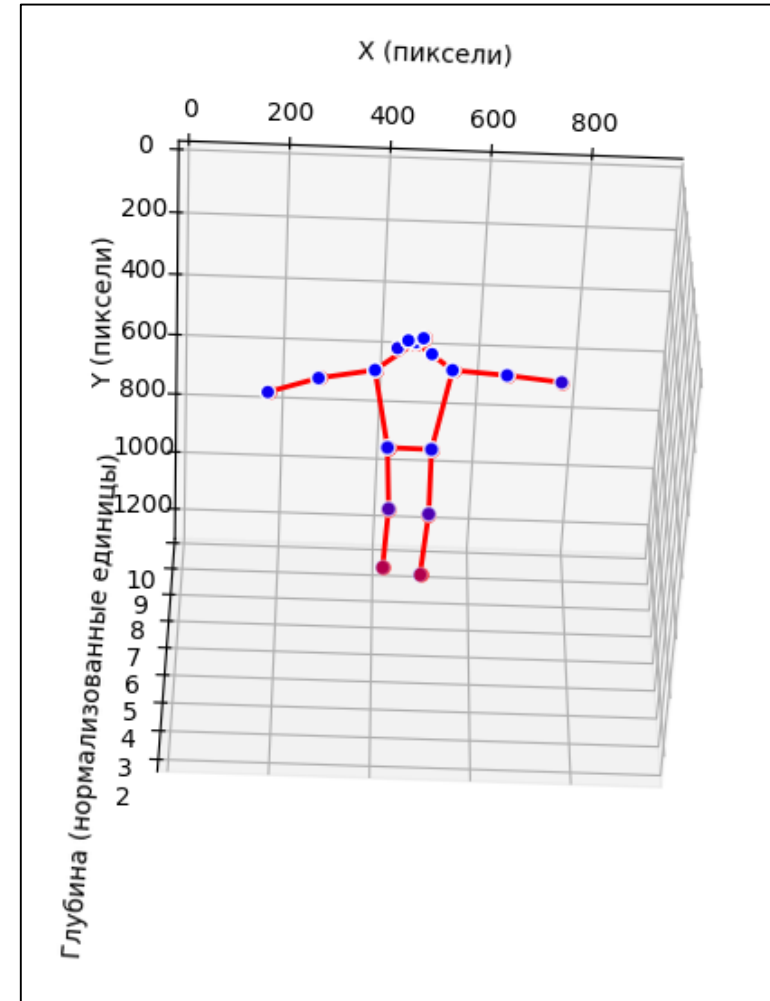




## Получение карт глубины и построение трехмерного скелета



Карта глубины, полученная с помощью MiDaS



Результат совместной работы собственной модели определения двумерных ключевых точек с моделью MiDaS



В ходе работы был проведен аналитический обзор научной литературы по методам захвата движений с маркерами и без маркеров, исследованы и сравнены алгоритмы машинного обучения, а также реализован алгоритм для определения двумерных ключевых точек с использованием различных архитектур моделей. Кроме того, были изучены методы преобразования двумерных ключевых точек в трёхмерные координаты.

Эксперименты с архитектурами ResNet18, ResNet50 и ResNet101 показали, что ResNet50 продемонстрировала наилучшие результаты, достигая  $AP_{50} = 85,5$ ,  $AP_{75} = 69,7$ ,  $AR_{50} = 90,5$ ,  $AR_{75} = 76,0$ .

Также, была произведена апробация научного доклада по теме данной работы в рамках Молодежной научной конференции университета.



**САМАРСКИЙ** УНИВЕРСИТЕТ  
SAMARA UNIVERSITY

**БЛАГОДАРЮ  
ЗА ВНИМАНИЕ**

[a.a.soroka05@gmail.com](mailto:a.a.soroka05@gmail.com)