|  |  |
| --- | --- |
| **СОГЛАСОВАНО**  \_\_\_\_\_\_\_\_\_\_\_\_\_ С.М. Липкин  «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2022 г. | **УТВЕРЖДАЮ**  \_\_\_\_\_\_\_\_\_\_\_\_\_ Н.В. Старостин  «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2022 г. |

**Отчет-аннотация**

**«Формирование верификационного базиса»**

**НИР «Разработка программного обеспечения для анализа данных атомной силовой микроскопии»**

**№ 02068143.00237-06**

1. **Введение**

В рамках данного документа сформированы предложения по формированию тренировочных и тестовых данных для нейронной сети.

1. **Генерация обучающей выборки**

Для обучения нейронной сети необходимо подготовить обучающую выборку, состоящую из 10000 (±2000) изображений в матричном варианте.

Формирование этих данных происходит с помощью скрипта “fake\_dataset\_utils.py”, который имеет следующие аргументы:

* Ширина картинки;
* Высота картинки;
* Количество экземпляров.

На выходе скрипта получается папка с 1 текстовым файлом, содержащим в первых 3 строках информацию о количестве данных (N), ширине(W) и высоте(H) каждой картинки. Ниже записана матрица высотой N\*H и шириной W.

Так как реальные данные от заказчика пришли достаточно поздно, то для имитации зародыша было принято решение генерировать треугольник из 1 размером 50-70 пикселей в ширину и 6-8 в высоту и помещать его в разные части картинки.

1. **Генерация верификационного базиса**

Для проверки корректности работы нейронной сети был также сгенерирован верификационный базис, состоящий из 500 картинок размером 256\*20. Данный базис также состоит из сгенерированных треугольников из 1, помещенных в разные части картинки.