**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**КЕМЕРОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**ИНСТИТУТ ЦИФРЫ**

**ОТЧЕТ**

**О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ**

**«Анализ данных»**

**студента 4-го курса, группы ФИТ-204**

**Абрамова Ильи Олеговича**

02.03.02 «Фундаментальная информатика и информационные технологии»

Руководитель:

Кандидат технических наук, доцент

Колесникова Татьяна Геннадьевна

Работа защищена

«\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_»

“\_\_\_\_” \_\_\_\_\_\_\_\_\_\_\_\_\_2024г.

Кемерово 2024

# Cбор данных

Для обучения модели были найдены сторонние (3rd party) данные, содержащие цветные изображения.

Ссылка на датасет:

<https://drive.google.com/drive/folders/1VDdLRZAsGp_jAhjZo7w49yMsRf4qS_8d>

# Описание данных

Датасет содержит изображения без определенной категории, встречаются пейзажи (как городские, так и природные), фото животных, людей, групп людей.

Свойства датасета:

* изображения размером 150х150 пикселов формата «.jpg»;
* 2791 изображение в датасете;
* цветные изображения;

# Исследование данных

Изображения, содержащиеся в датасете, имеют небольшой размер, что положительно скажется на скорости обучения.

Изображения не привязаны к определенной категории, это накладывает определенные трудности в дальнейшем, так как потребуется дообучать модель для некоторых категорий.

Для использования этих данных с GAN, предварительно для каждого изображения будет создаваться парное черно-белое изображение.

В итоге, данные подойдут для обучения модели, однако в будущем может потребоваться дообучение, чтобы достигнуть желаемого результата работы.

# Качество данных

Как говорилось выше, набор содержит почти 2800 пригодных для обучения изображений. Все данные одного разрешения, одного формата. Данные упорядочены и пронумерованы. Нет разделения на train и split.

Ссылка на репозиторий с кодом:

<https://github.com/asskicker54/colorizer>