

Определение пола, возраста и этноса человека по фотографии и генерация лиц по заданным атрибутам с применением нейронных сетей и методов компьютерного зрения

Дипломный проект по профессии «Data Scientist»

Каторгин Иван
Группа DSU-PROD-67



Содержание

1

Постановка задачи
для глубокого
обучения

2

Анализ исходных
данных,
преобразования и
очистка

3

Определение пола,
возраста и этноса
человека по
фотографии

4

Генерация лиц по
заданным
атрибутам

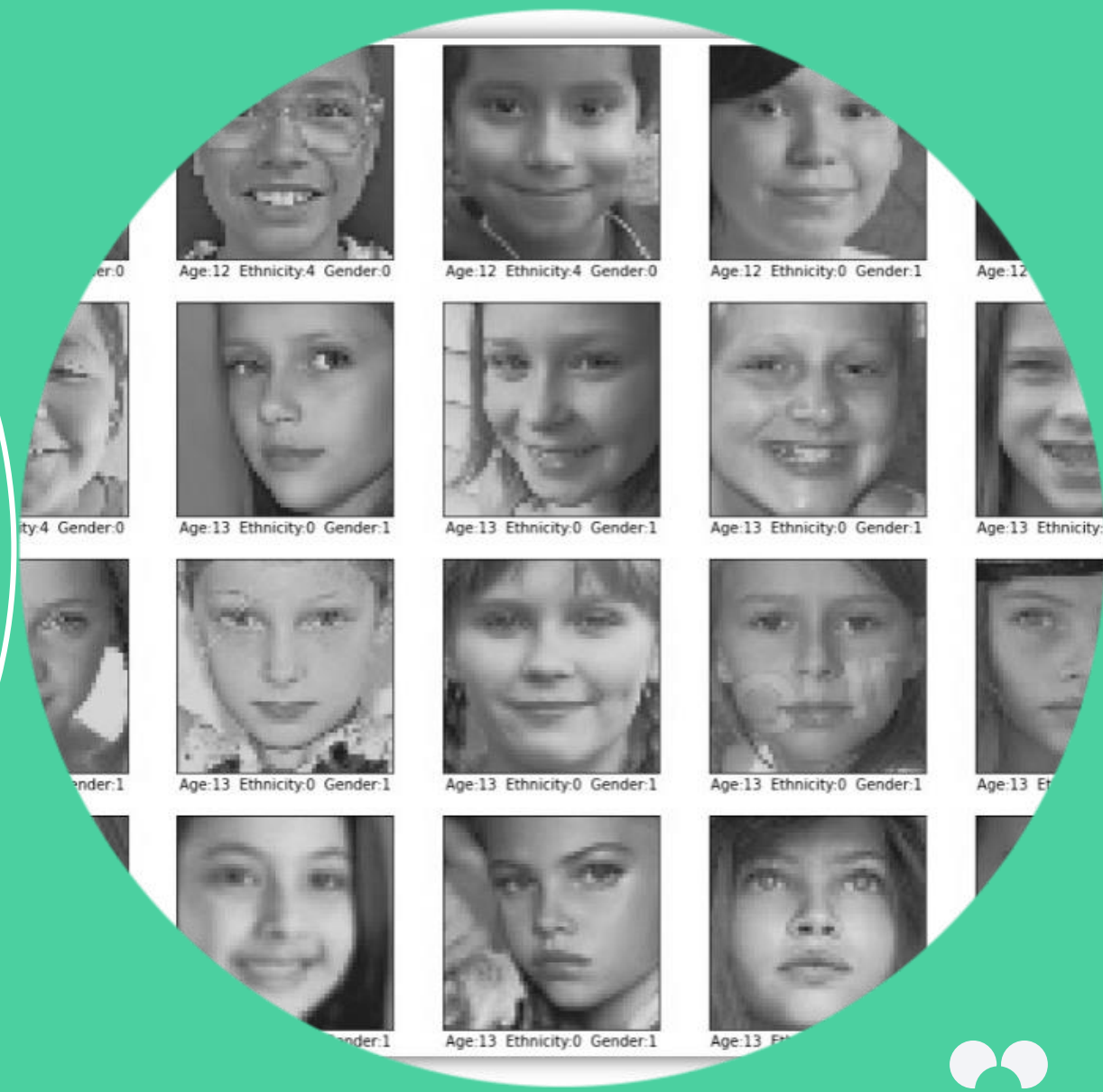
5

Выводы



Постановка задачи для глубокого обучения

1



Описание проекта

Исходная задача:

1. Определение заданных атрибутов (пол, возраст, этнос) человека по фотографии
2. Генерация лиц по заданным атрибутам

Актуальность задачи, её место в предметной области:

Определение заданных атрибутов (пол, возраст, этнос) по изображению лица может применяться в решении следующих задач:

- оборот товаров, имеющих возрастные ограничения
- реклама человеку релевантных товаров и услуг
- соблюдение правил доступа несовершеннолетних лиц к информации
- поиск лиц находящихся в розыске и т.д.

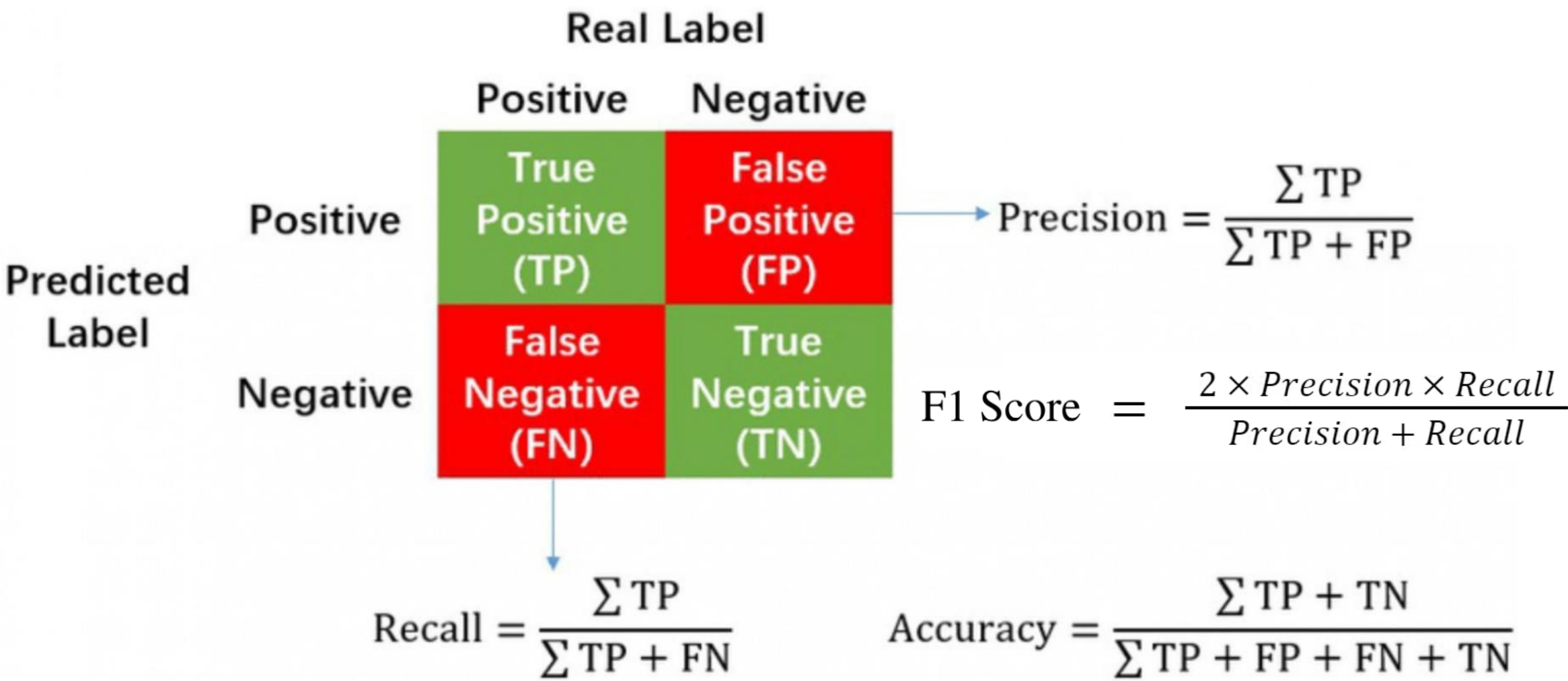
Генерация лиц по заданным атрибутам может применяться в:

- маркетинге и рекламе
- разработке и дизайне
- образовании



Целевые метрики

Задача классификации (определение пола и этноса)



Задача регрессии (определение возраста)

$$MAE = \frac{\sum |(y_i - y_p)|}{n}$$

$$RMSE = \sqrt{\frac{\sum (y_i - y_p)^2}{n}}$$

Метрика	Определение пола	Определение этноса	Определение возраста	Генерация лиц
Accuracy	+	+		
Precision	+			+
Recall	+			+
F1 score	+	+		
MAE			+	
RMSE			+	
FID				+

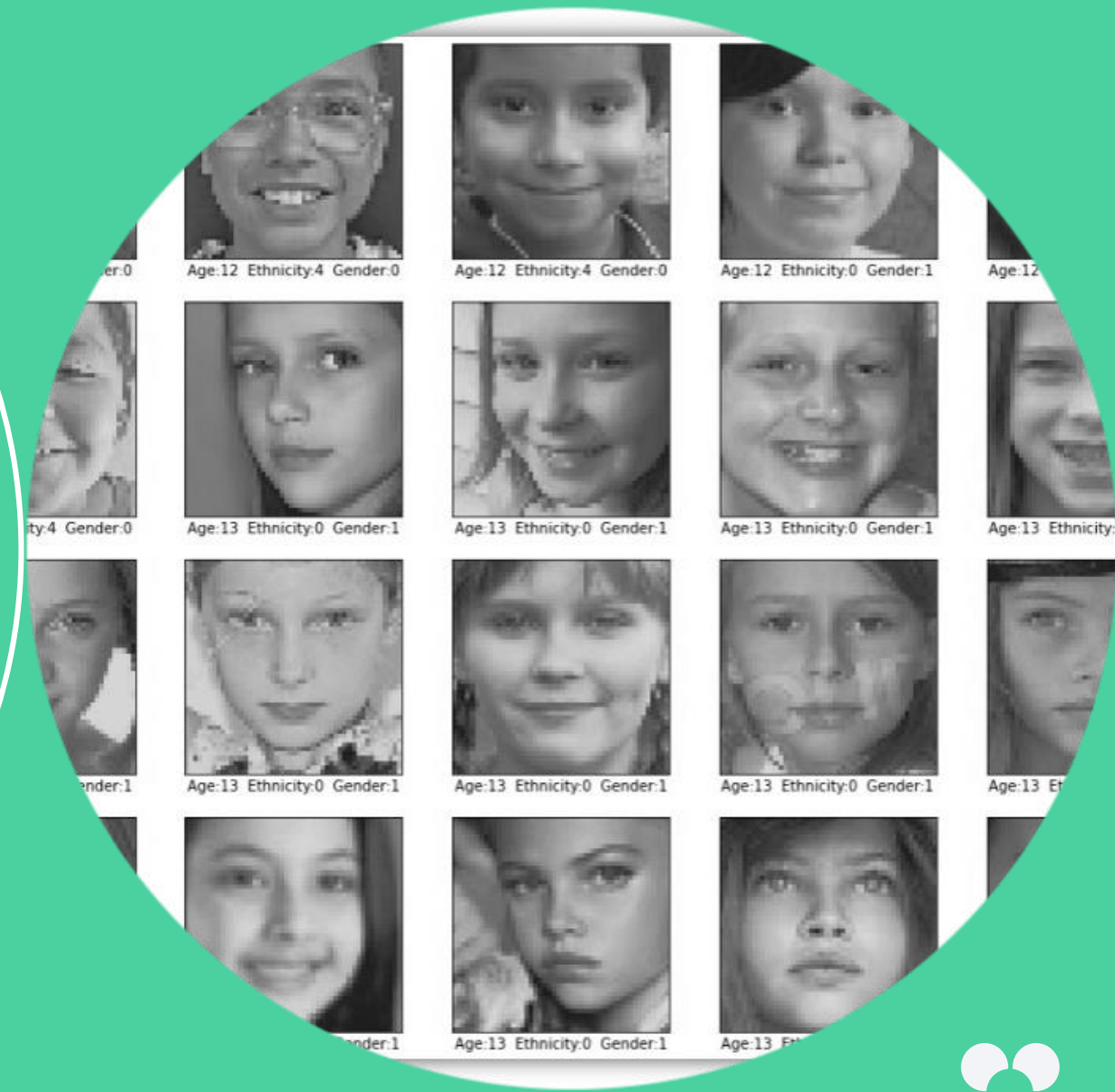
Генерация лиц по заданным атрибутам

$$FID = \|\mu - \hat{\mu}\|^2 + \text{Tr}(\Sigma + \hat{\Sigma} - 2(\Sigma\hat{\Sigma})^{1/2})$$



Анализ исходных данных, преобразования и очистка

2



Описание датасета

Название датасета	AGE, GENDER AND ETHNICITY (FACE DATA) CSV
Источник данных	https://www.kaggle.com/datasets/nipunarora8/age-gender-and-ethnicity-face-data-csv
Тип данных	Int64, object
Размер датасета	5 столбцов, 23705 строк
Размер изображений	48 x 48 пикселей

```
1 # загрузим файл
2 df = pd.read_csv('/content/age_gender.csv')
3 df.head()
```

	age	ethnicity	gender	img_name	pixels
0	1	2	0	20161219203650636.jpg.chip.jpg	129 128 128 126 127 130 133 135 139 142 145 14...
1	1	2	0	20161219222752047.jpg.chip.jpg	164 74 111 168 169 171 175 182 184 188 193 199...
2	1	2	0	20161219222832191.jpg.chip.jpg	67 70 71 70 69 67 70 79 90 103 116 132 145 155...
3	1	2	0	20161220144911423.jpg.chip.jpg	193 197 198 200 199 200 202 203 204 205 208 21...
4	1	2	0	20161220144914327.jpg.chip.jpg	202 205 209 210 209 209 210 211 212 214 218 21...

Датасет представляет собой CSV-файл с изображениями лиц, маркированными по возрасту, полу и этнической принадлежности.

Набор данных содержит 23705 строк и 5 столбцов:

- Возраст
- Этнос
- Пол
- Имя изображения
- Пиксели (массив в строку пикселей изображения)

```
1 df.shape
(23705, 5)

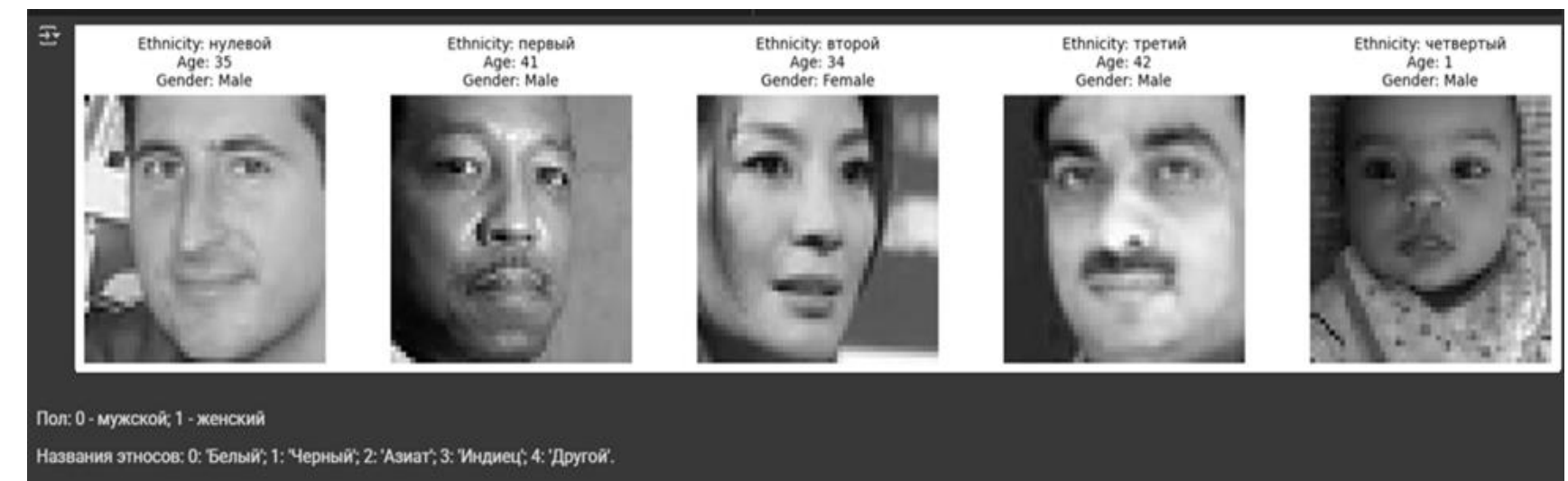
1 print(df.columns)
Index(['age', 'ethnicity', 'gender', 'img_name', 'pixels'], dtype='object')

1 df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23705 entries, 0 to 23704
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0    age         23705 non-null  int64
1    ethnicity   23705 non-null  int64
2    gender      23705 non-null  int64
3    img_name    23705 non-null  object
4    pixels      23705 non-null  object
dtypes: int64(3), object(2)
memory usage: 926.1+ KB
```

```
1 # определим размеры изображений
2 pixels = df['pixels'].iloc[0]
3 pixel_list = list(map(int, pixels.split())) # преобразуем строку в список чисел
4
5 # Предположим, что изображение квадратное, grayscale (1 канал)
6 if height == width == int(np.sqrt(len(pixel_list))):
7     print(f"Размер изображения: {height} x {width}")
8 else:
9     print("Ошибка: изображение не квадратное")

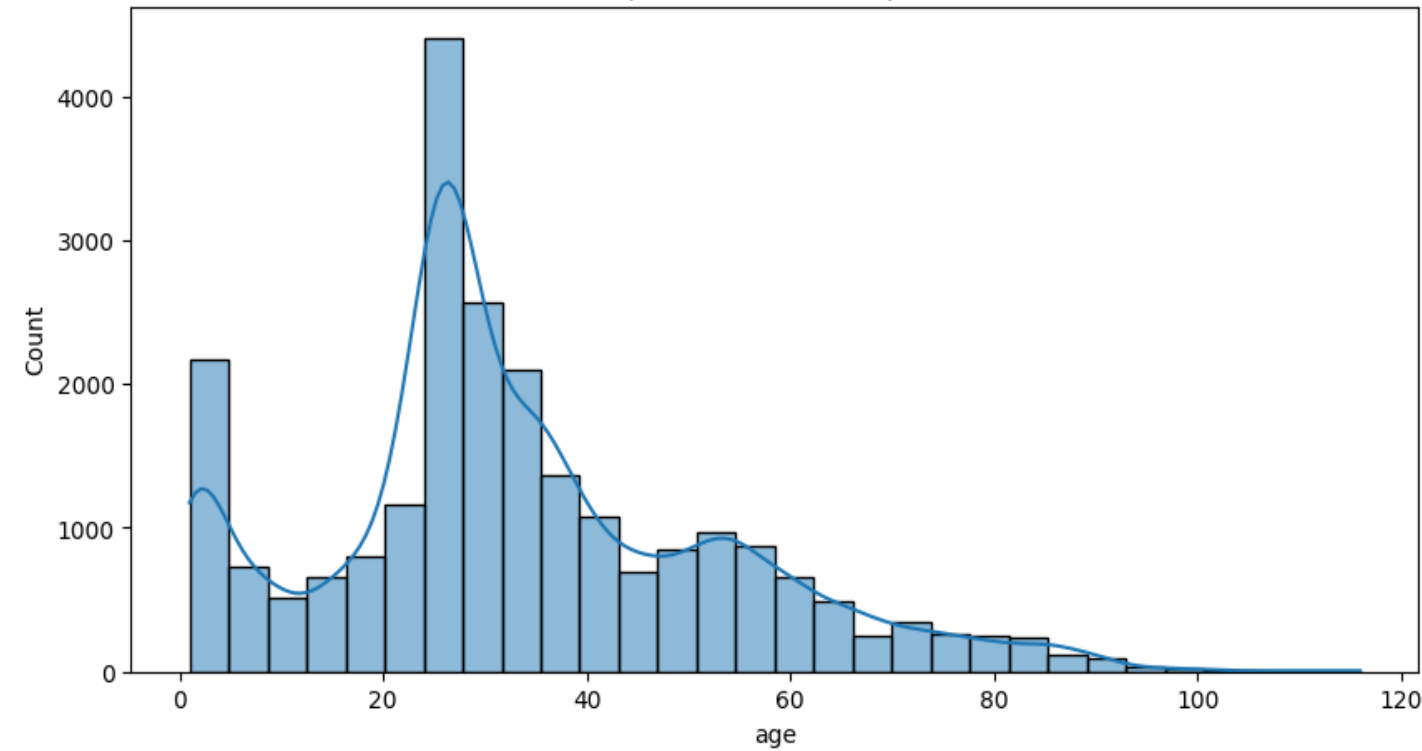
Размер изображения: 48 x 48
```

```
1 df.isnull().sum()
0
age      0
ethnicity 0
gender   0
img_name 0
pixels   0
dtype: int64
```

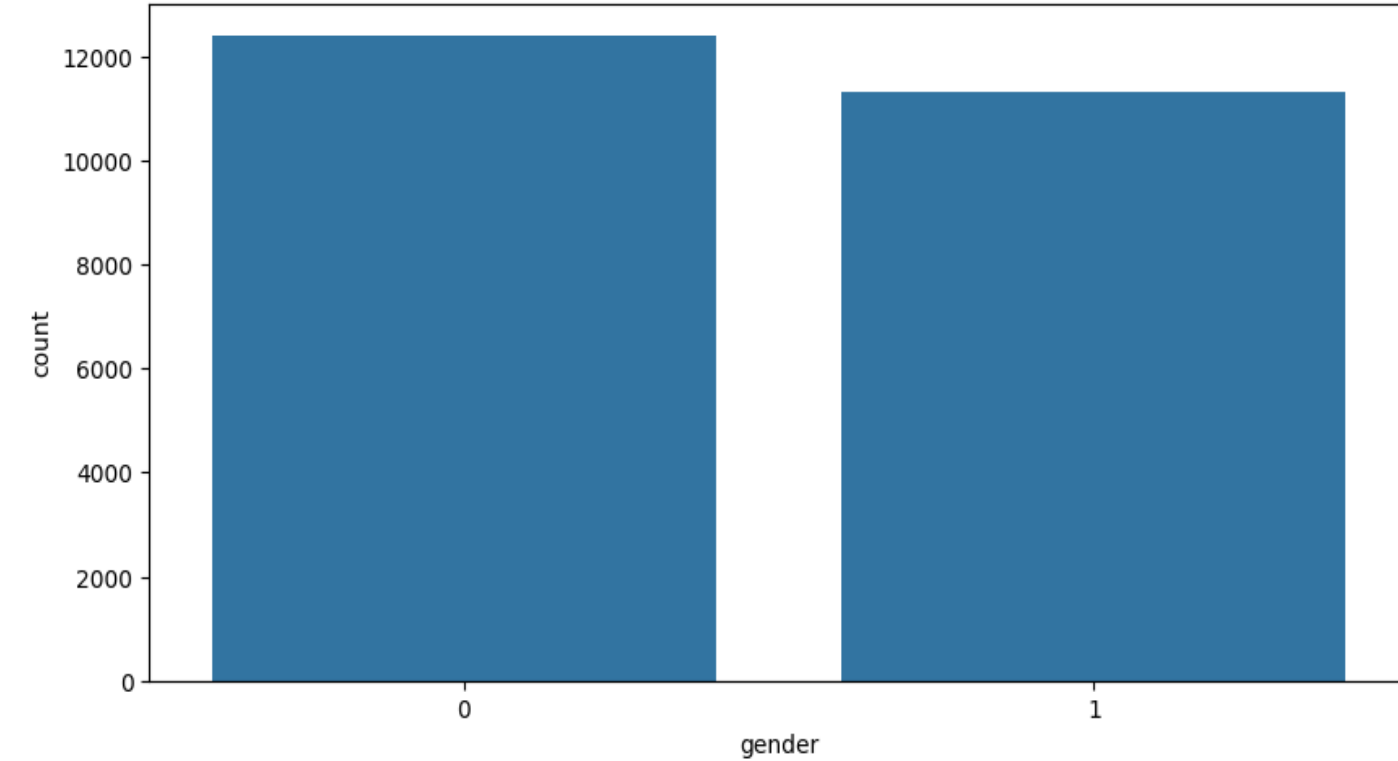


Разведочный анализ данных

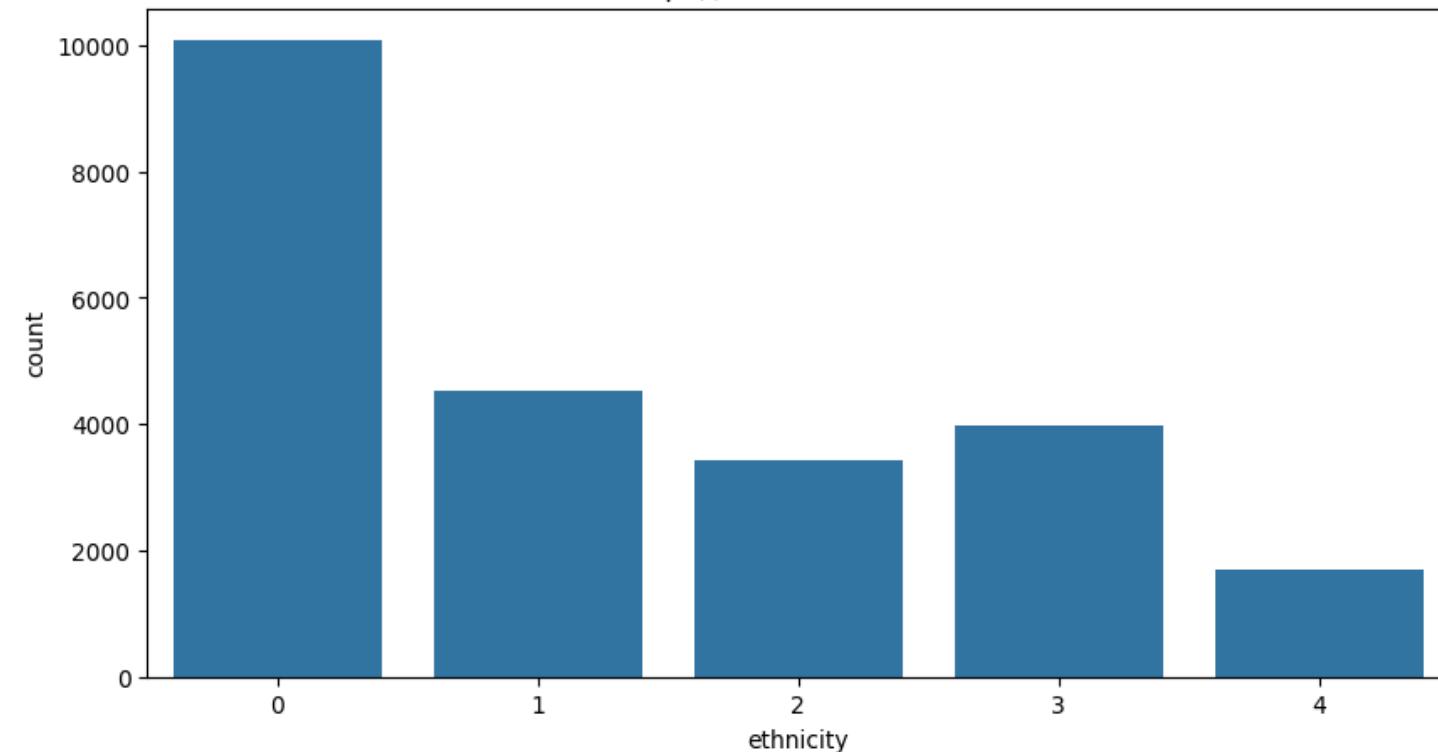
Распределение по возрастам



Распределение по полу



Распределение по этносам



Основные выводы по разведочному анализу данных:

- Пропуски отсутствуют
- Явных выбросов или аномалий не обнаружено
- Датасет не сбалансирован

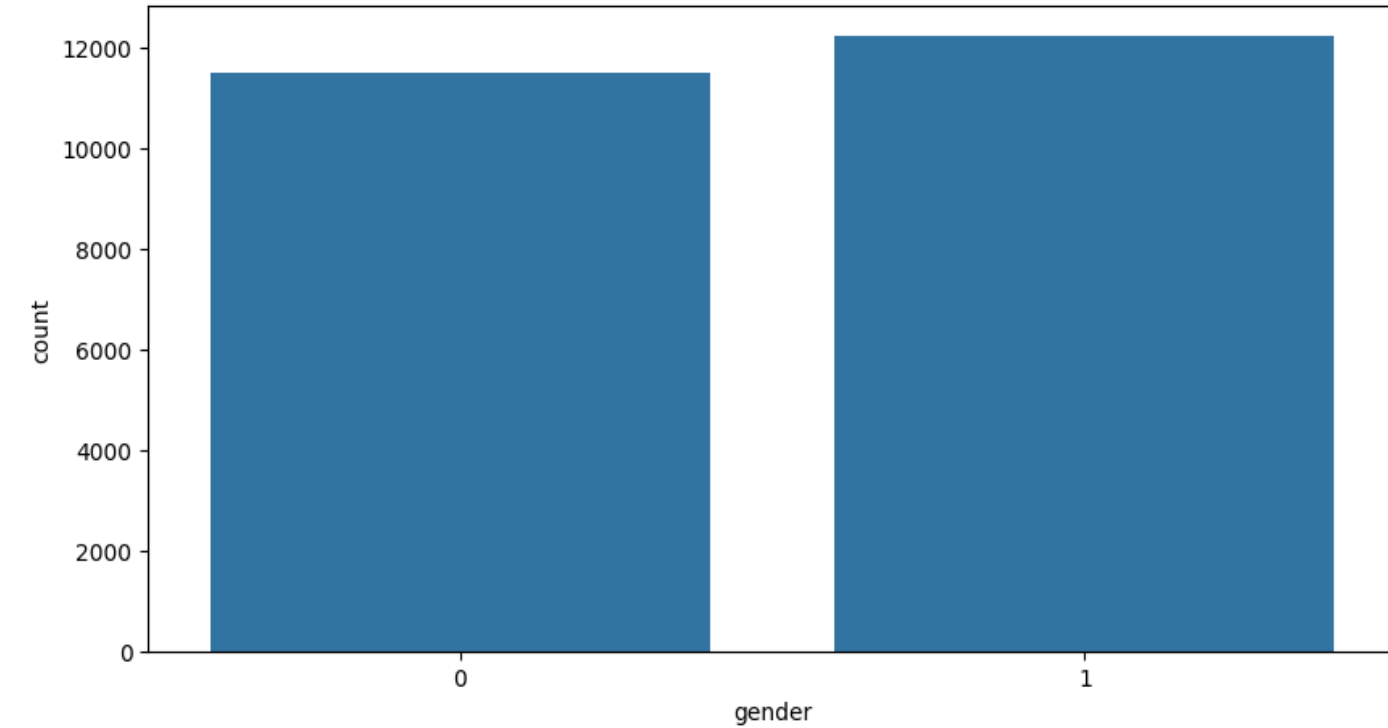


Преобразования и очистка

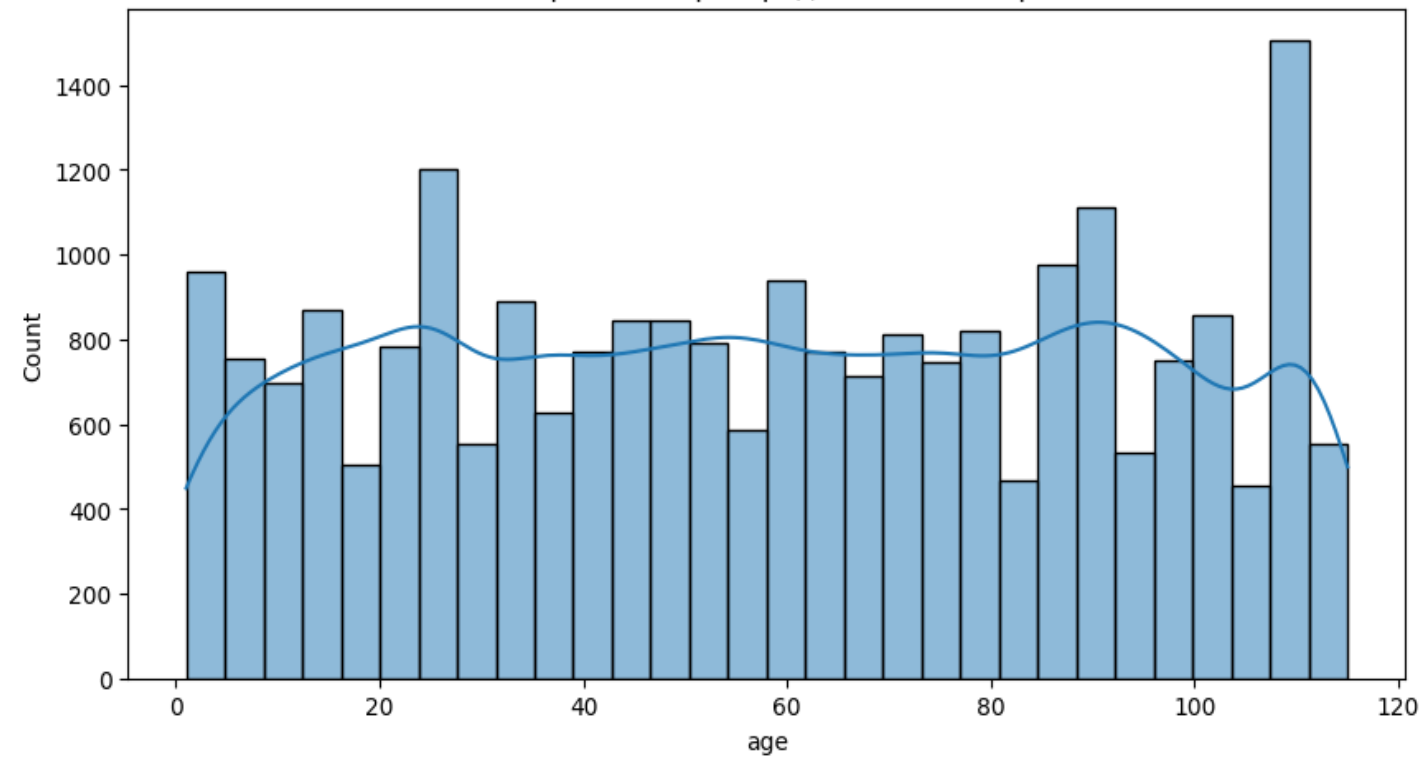
Аугментация данных

```
1 # Инициализация ImageDataGenerator
2 datagen = ImageDataGenerator(
3     rotation_range=15,      # случайный поворот на ±15 градусов
4     width_shift_range=0.1,  # случайный сдвиг по ширине на ±10%
5     height_shift_range=0.1, # случайный сдвиг по высоте на ±10%
6     shear_range=0.1,       # случайный сдвиг
7     zoom_range=0.1,        # случайное увеличение/уменьшение
8     horizontal_flip=True,   # случайное отражение по горизонтали
9     fill_mode='nearest'    # заполнение пикселей при трансформациях
10 )
```

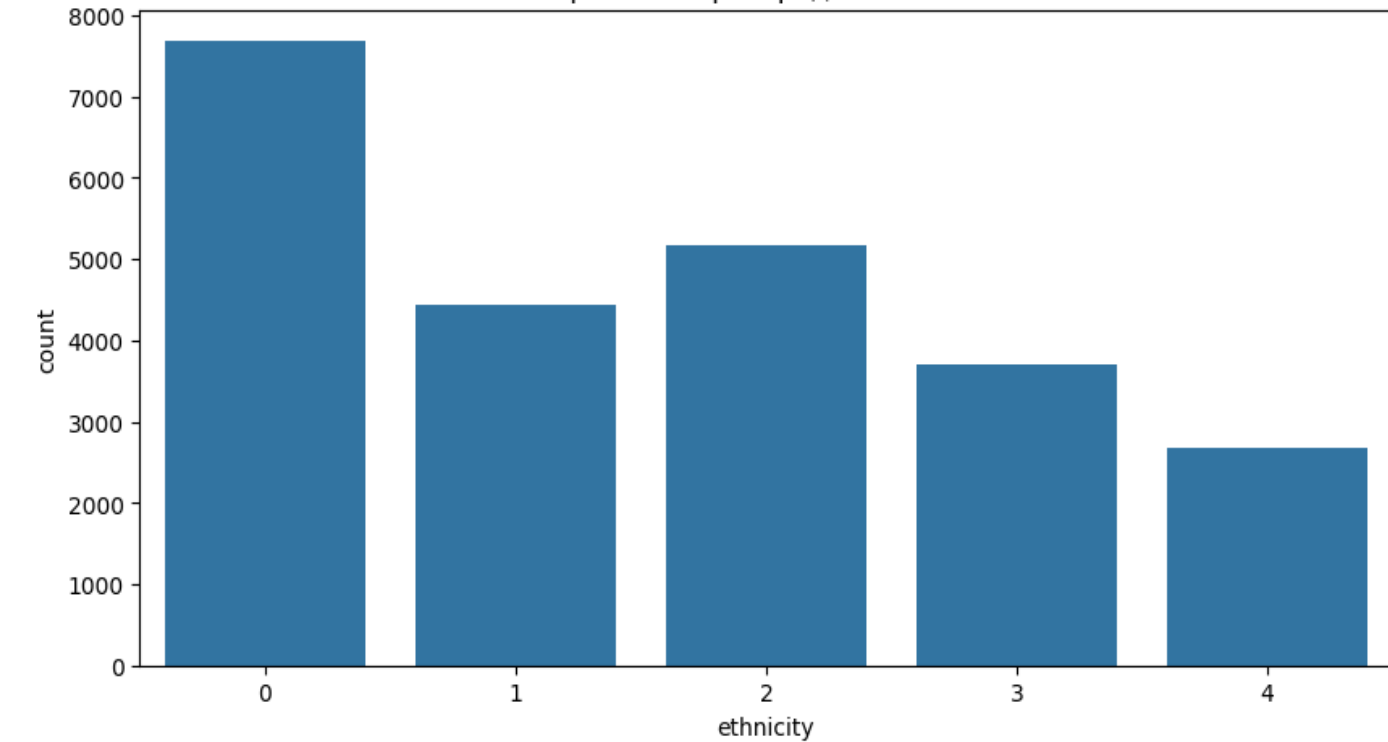
Сбалансированное распределение по полу



Сбалансированное распределение по возрастам

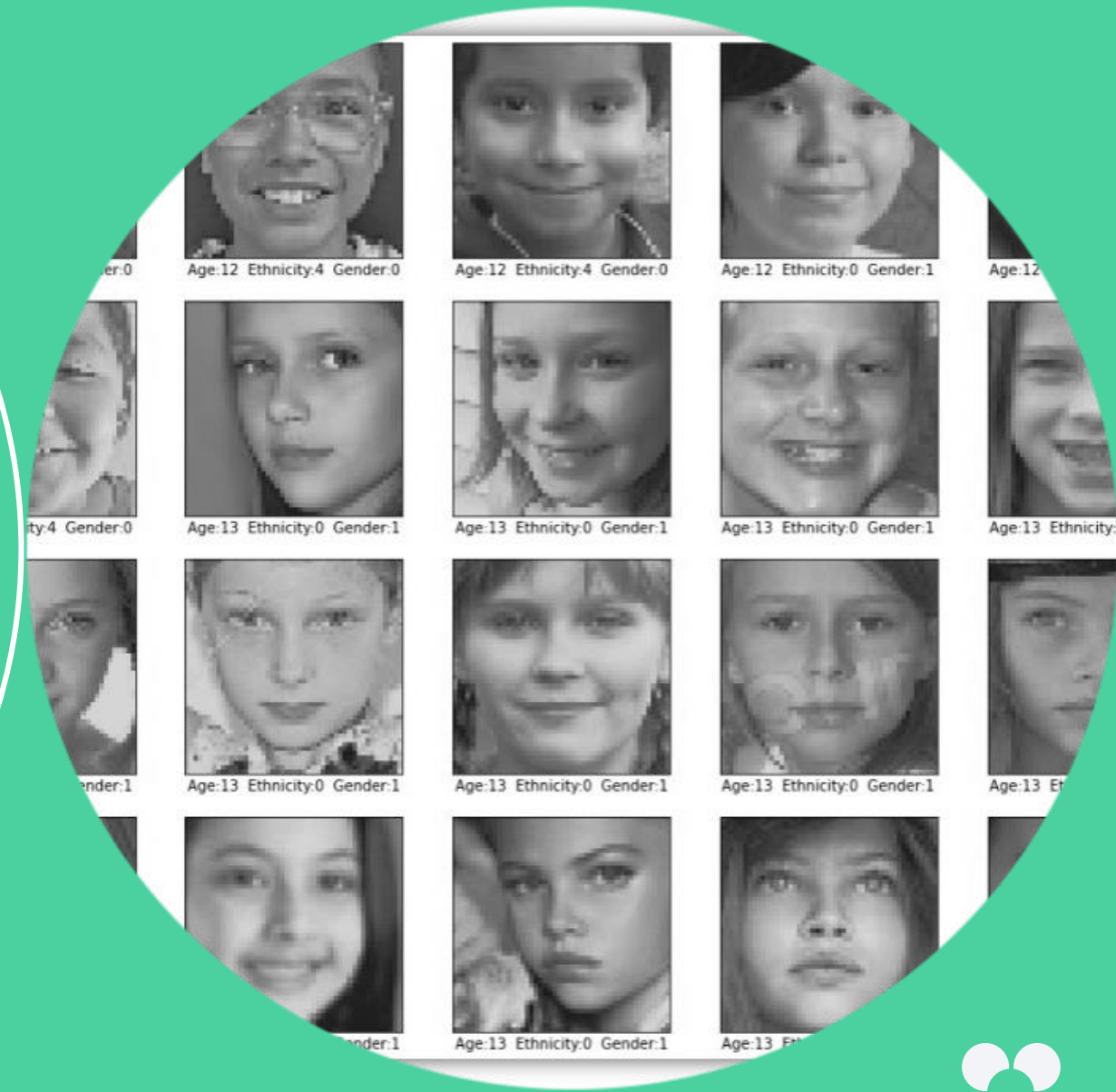


Сбалансированное распределение по этносам



Определение пола, возраста и этноса человека по фотографии

3



Параметры моделей

Параметры

Модель	Параметры	Input Shape	Предобучение	Attention	Особенности
ResNet50	~23.5M	224x224x3	ImageNet	Нет	Базовая transfer learning модель
EfficientNetB0	~4.0M	224x224x3	Нет	Нет	Эффективная архитектура
CNN (Multi-Task)	~1.2M	48x48x1	Нет	Нет	Residual-блоки, простота
CNN + Attention	~1.3M	48x48x1	Нет	Да	Spatial + Channel Attention
SNN (EfficientNetB0)	~4.1M	48x48x1	Нет	Нет	Общие признаки + индивидуальные ветки

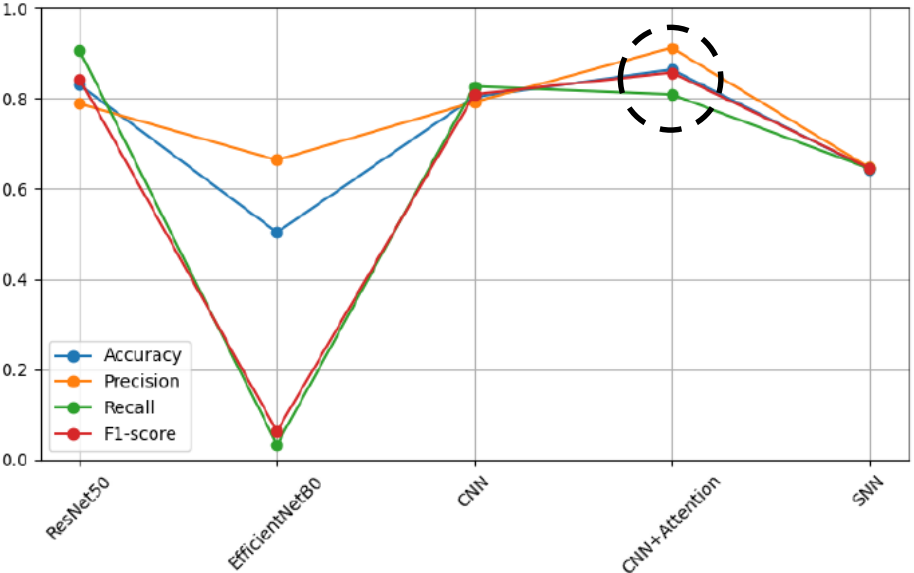
Гиперпараметры

Модель	Оптимизатор (LR)	Batch Size	Loss Weights (Пол/Этнос/Возраст)	Loss Functions (Пол/Этнос/Возраст)	Метрики (Пол/Этнос/Возраст)	Регуляризация	Attention	Input Shape	Особенности
ResNet50	Adam (0.001)	32	1.0 / 1.5 / 0.8	BCE / SparseCE / MSE	Accuracy / Accuracy / MAE	Dropout 0.5	Нет	224x224x3	Fine-tuning, предобучение на ImageNet
EfficientNetB0	Adam (0.0005)	32	1.0 / 1.2 / 0.8	BCE / SparseCE / MAE	Accuracy / Accuracy / MAE	Dropout 0.5, L2(0.01)	Нет	224x224x3	Обучение с нуля
CNN (Multi-Task)	Adam (0.001)	32	1.0 / 1.5 / 0.8	BCE / SparseCE / MAE	Accuracy / Accuracy / MAE	Dropout 0.3	Нет	48x48x1	Residual-блоки
CNN + Attention	Adam (0.001)	32	1.0 / 1.5 / 0.8	BCE / SparseCE / MAE	Accuracy / Accuracy / MAE	Dropout 0.3	Да (Spatial + Channel)	48x48x1	Улучшенная интерпретируемость признаков
SNN (на базе EfficientNetB0)	Adam (0.001)	32	1.0 / 1.5 / 0.8	BCE / SparseCE / MAE	Accuracy / Accuracy / MAE	Dropout 0.3-0.5, L2(0.01)	Нет	48x48x1 → 3 канала	Callbacks (EarlyStopping, ReduceLR)

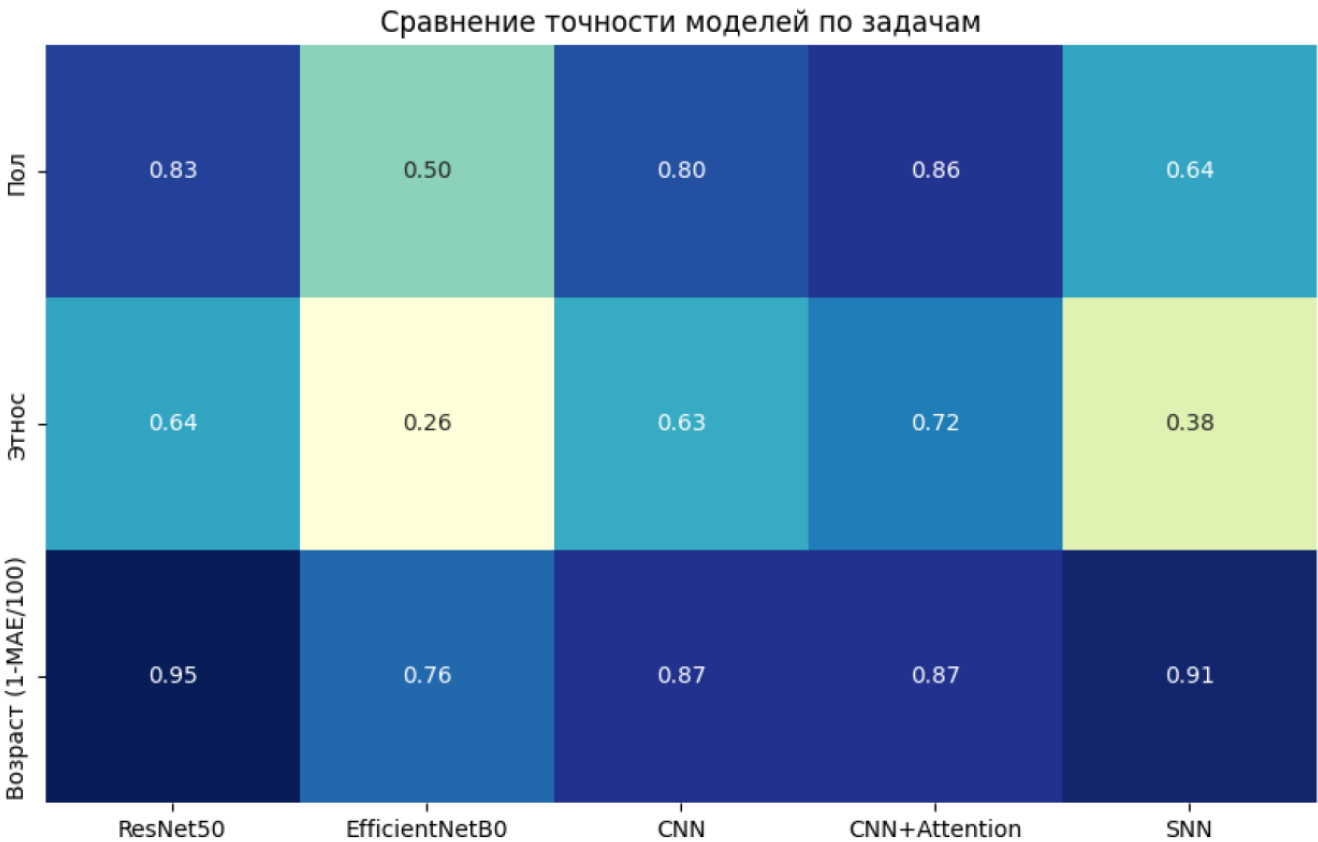


Итоги обучения

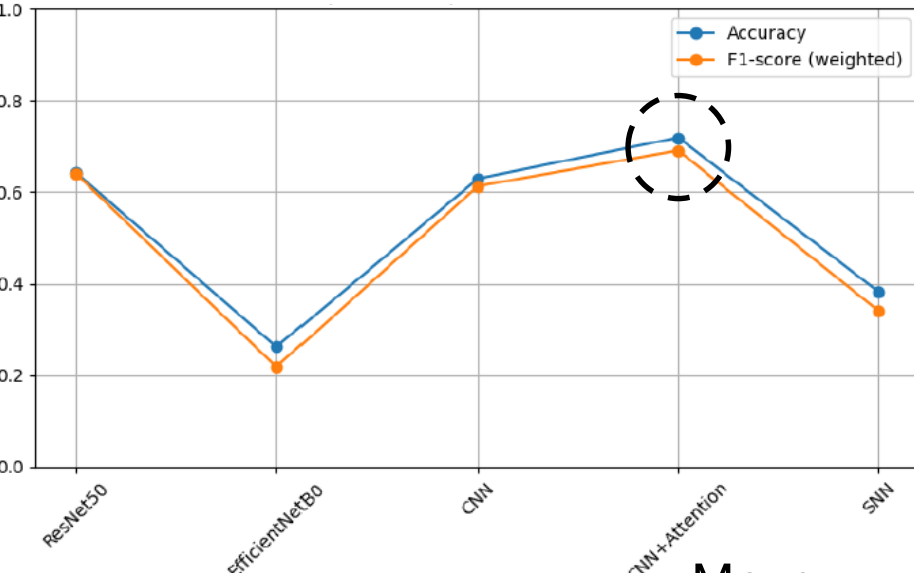
Метрики для предсказания пола



Модель	Accuracy	Precision	Recall	F1 score
ResNet50	83,05%	78,87%	90,72%	84,38%
EfficientNetB0	50,36%	66,39%	3,30%	6,29%
CNN	80,35%	79,25%	82,73%	80,95%
CNN + Attention	86,45%	91,23%	80,93%	85,77%
SNN	64,35%	64,74%	64,49%	64,61%

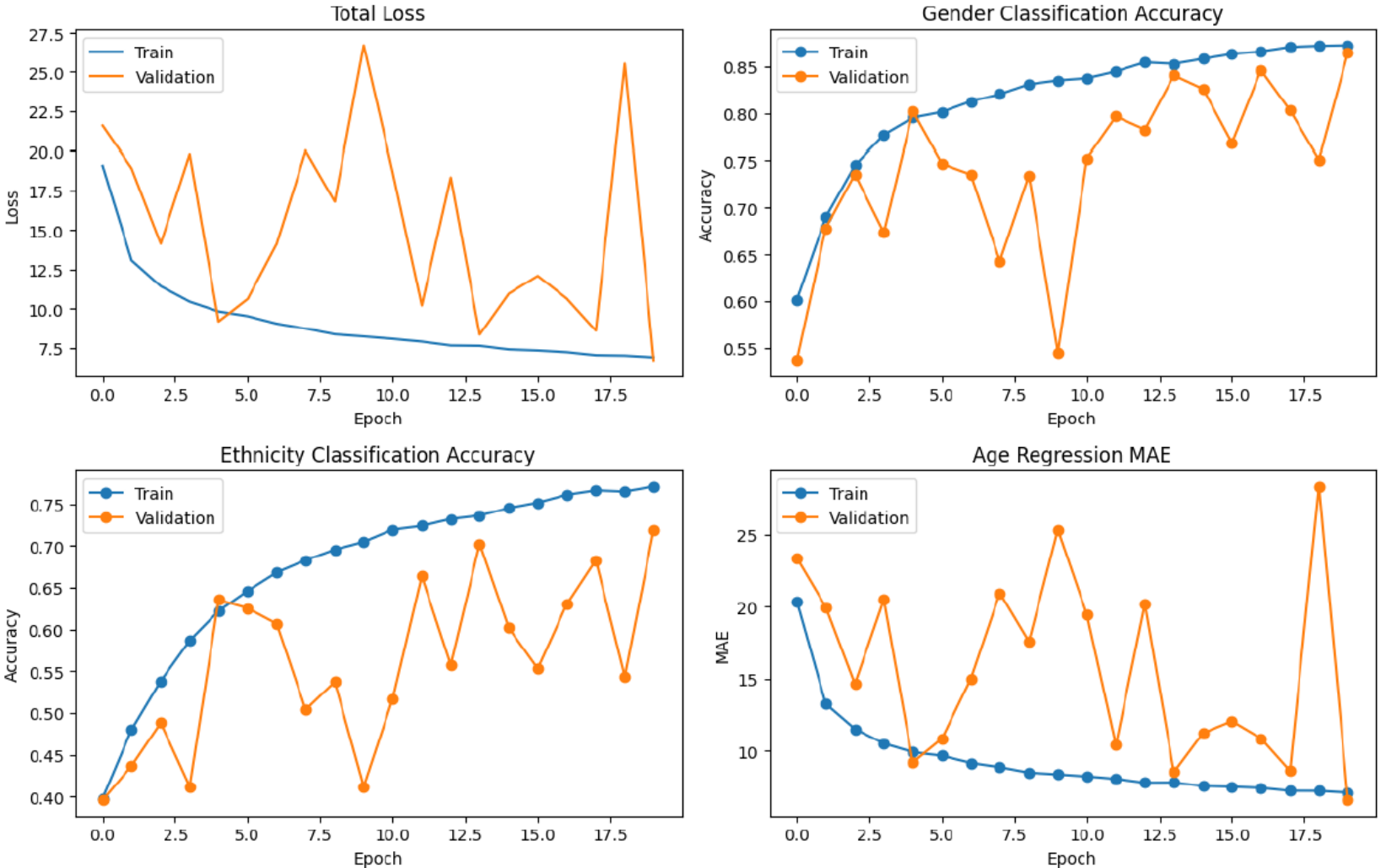


Метрики для предсказания этноса

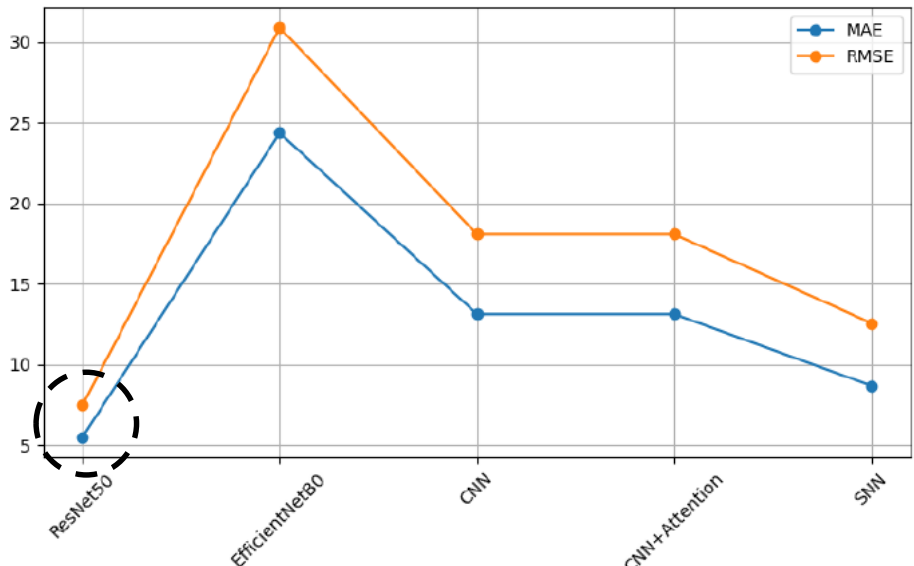


Модель	Accuracy	F1 score
ResNet50	64,33%	64,09%
EfficientNetB0	26,19%	21,82%
CNN	62,96%	61,33%
CNN + Attention	71,93%	69,11%
SNN	38,18%	34,01

Графики обучения модели CNN + Attention



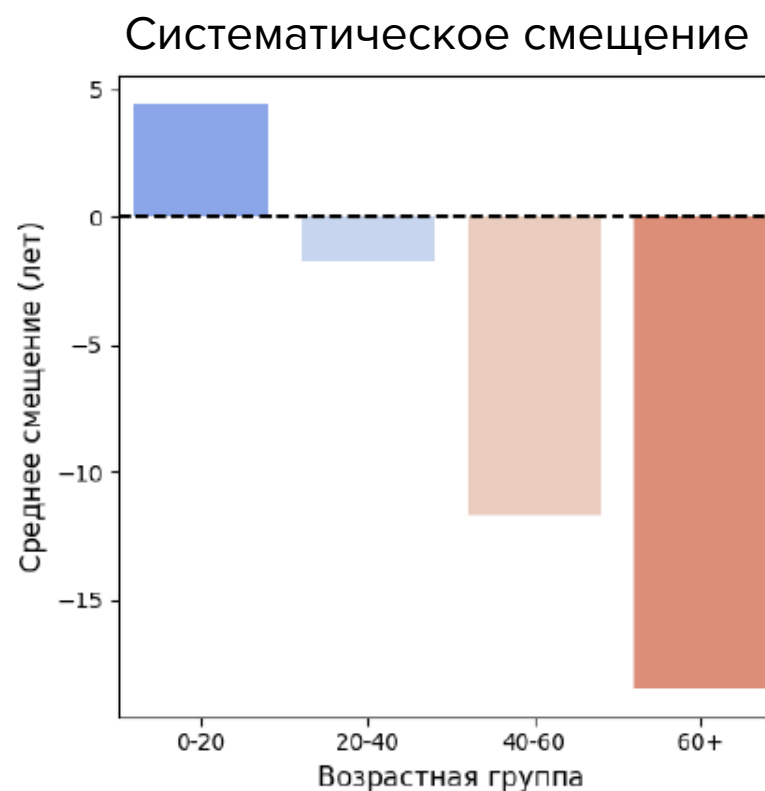
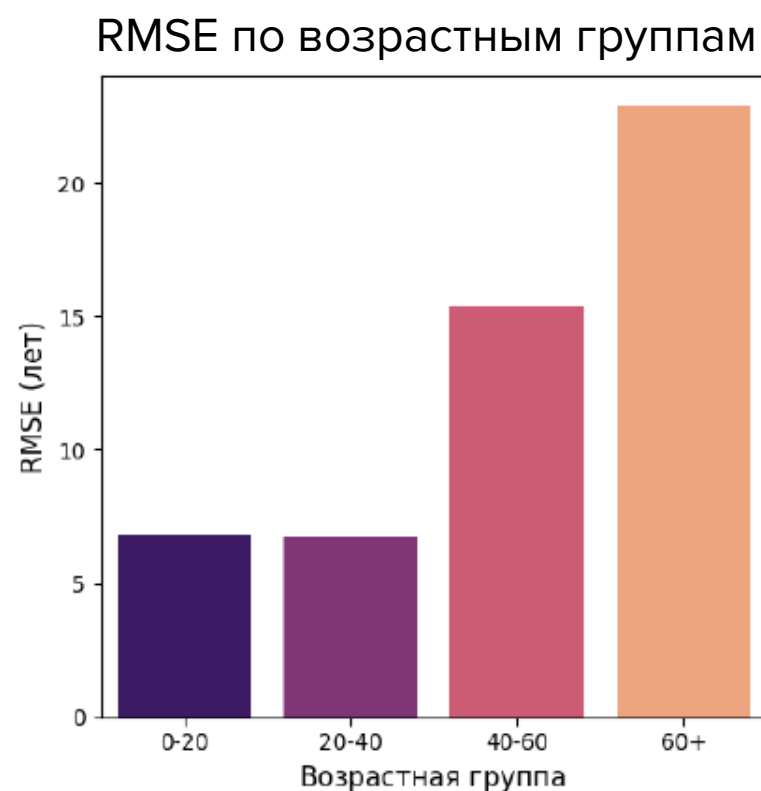
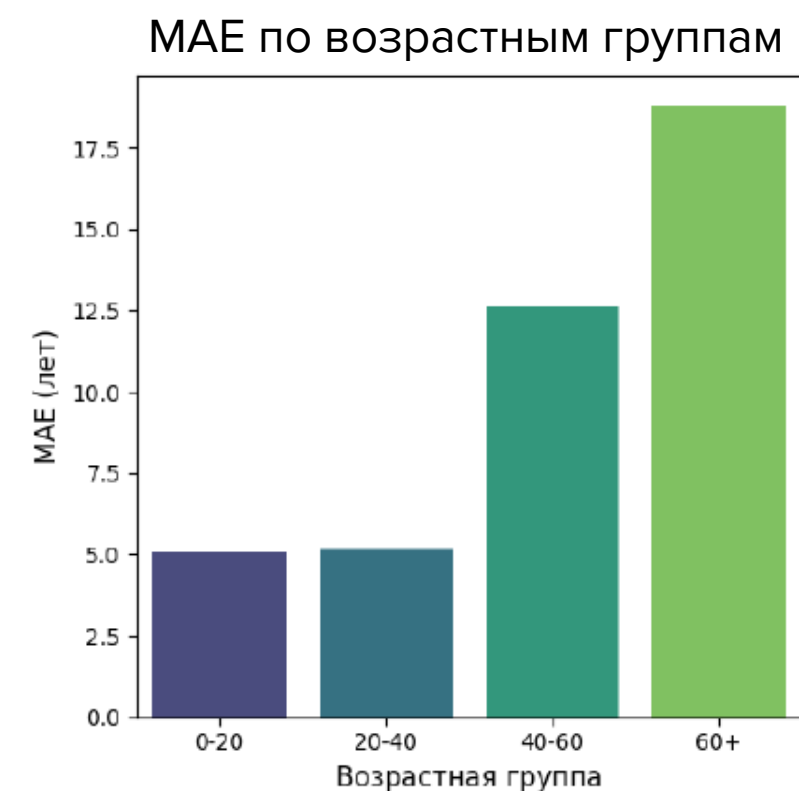
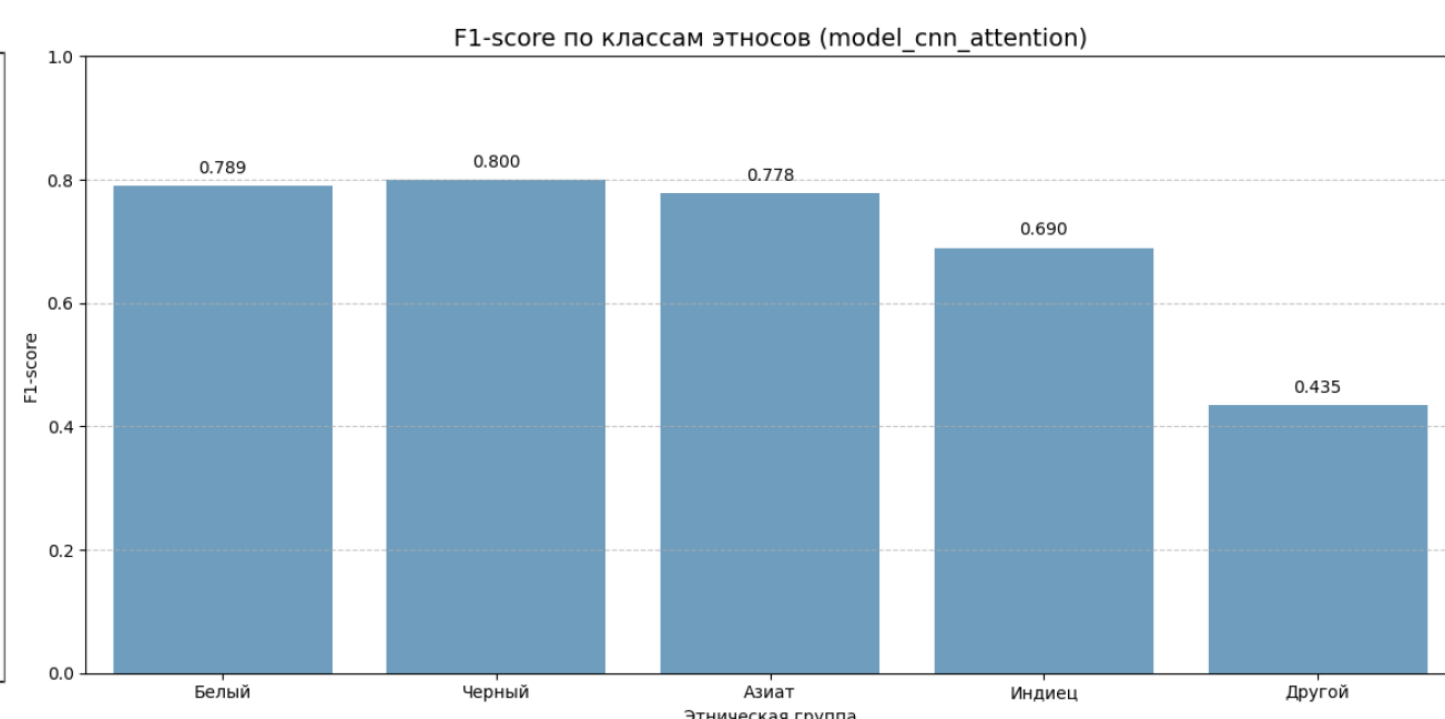
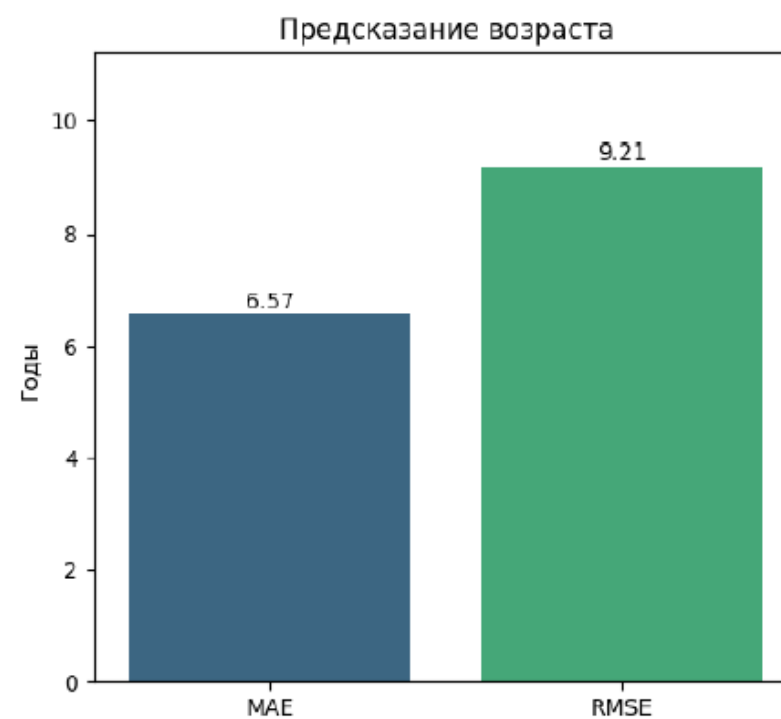
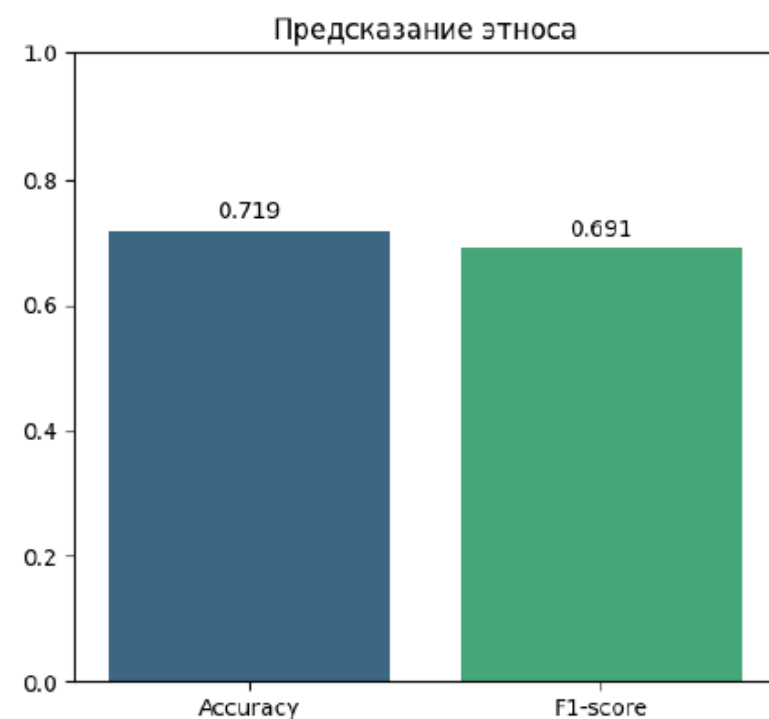
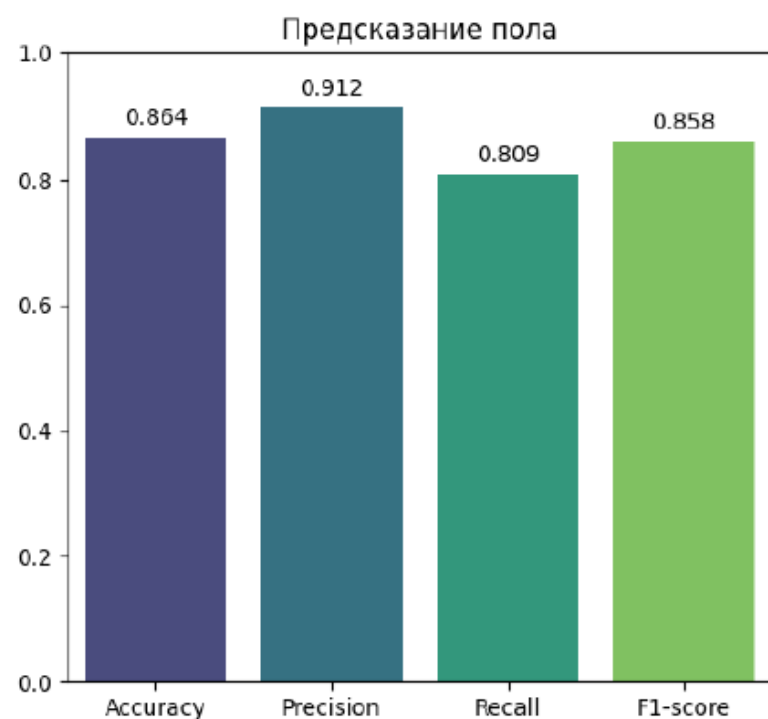
Метрики для возраста



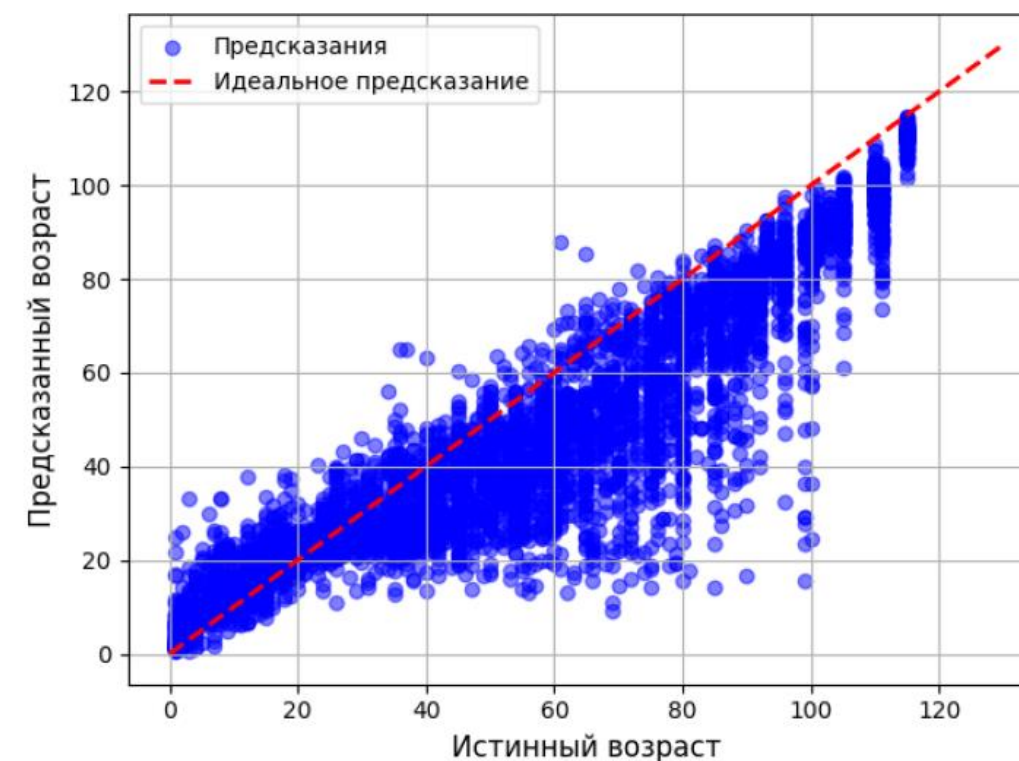
Модель	MAE	RMSE
ResNet50	5,49	7,46
EfficientNetB0	24,37	30,91
CNN	13,14	18,09
CNN + Attention	13,14	18,09
SNN	8,68	12,58



Анализ метрик

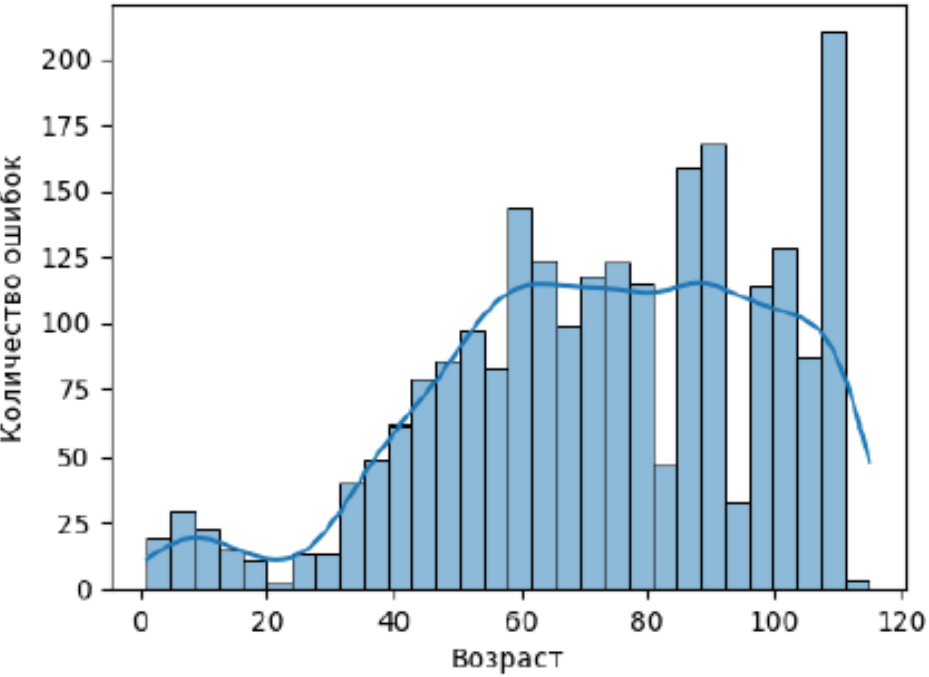


Кросс плот предсказанный vs истинный возраст

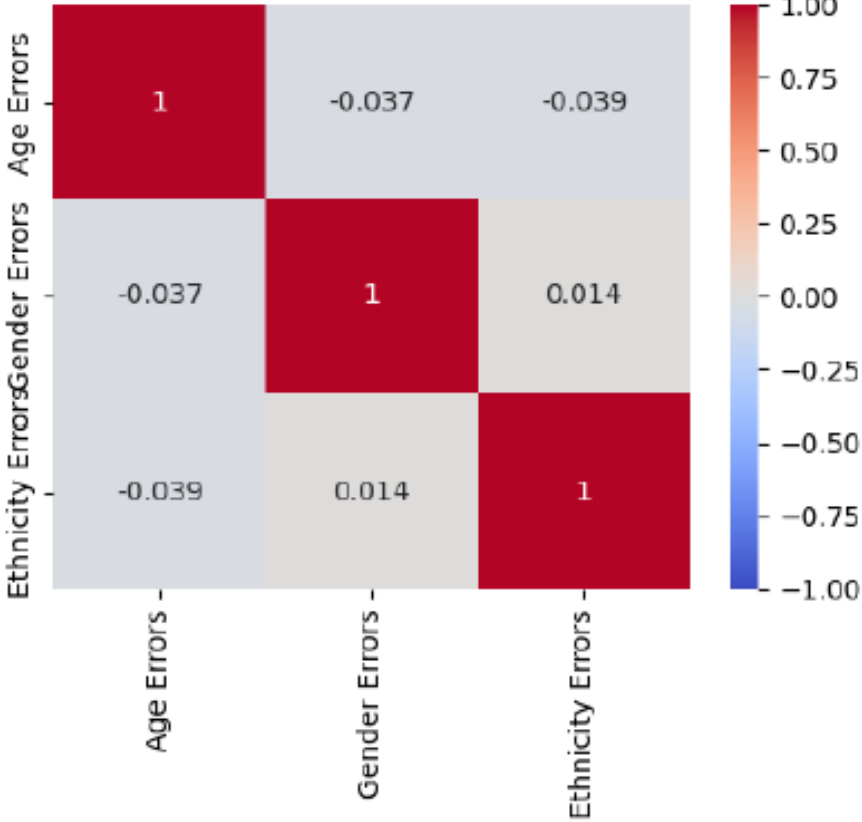


Анализ ошибок

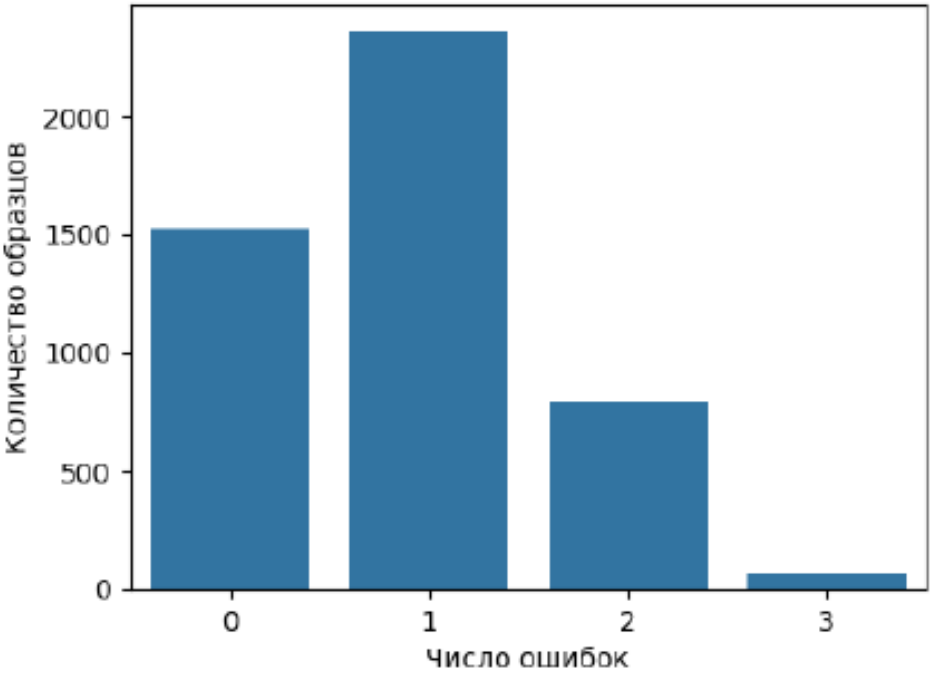
Грубые ошибки возраста (> 10 лет)



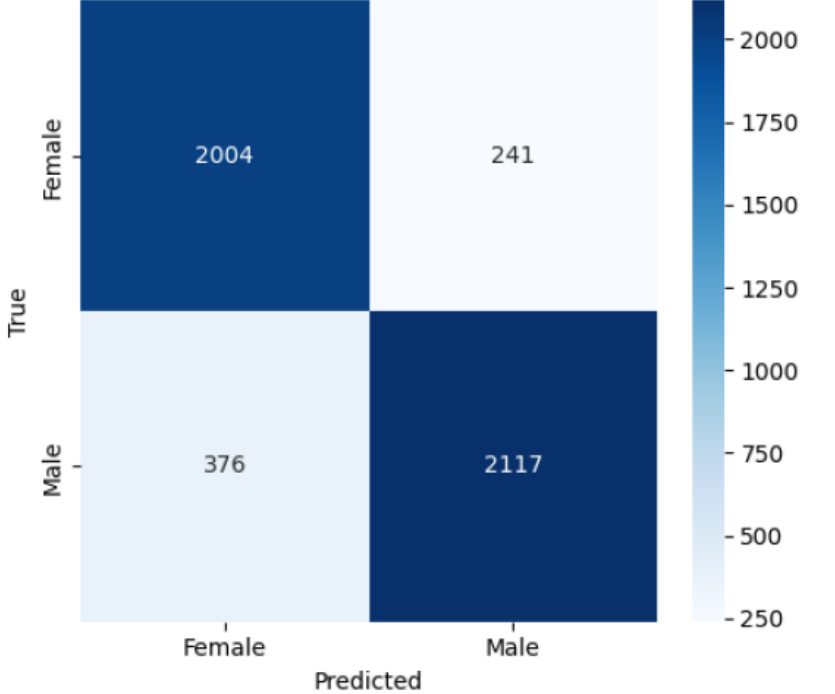
Корреляция между ошибками



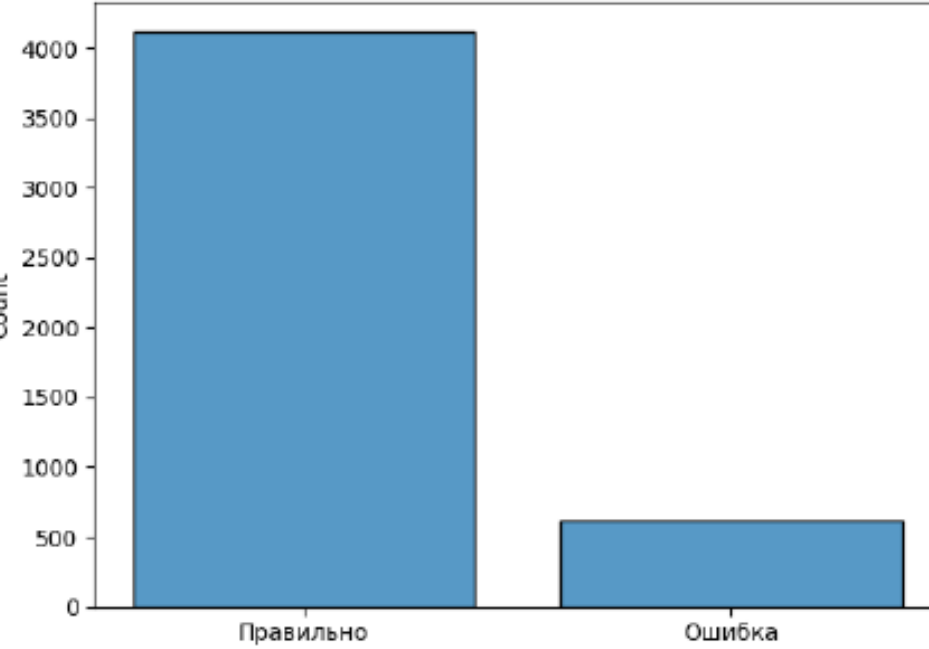
Количество ошибок на один образец



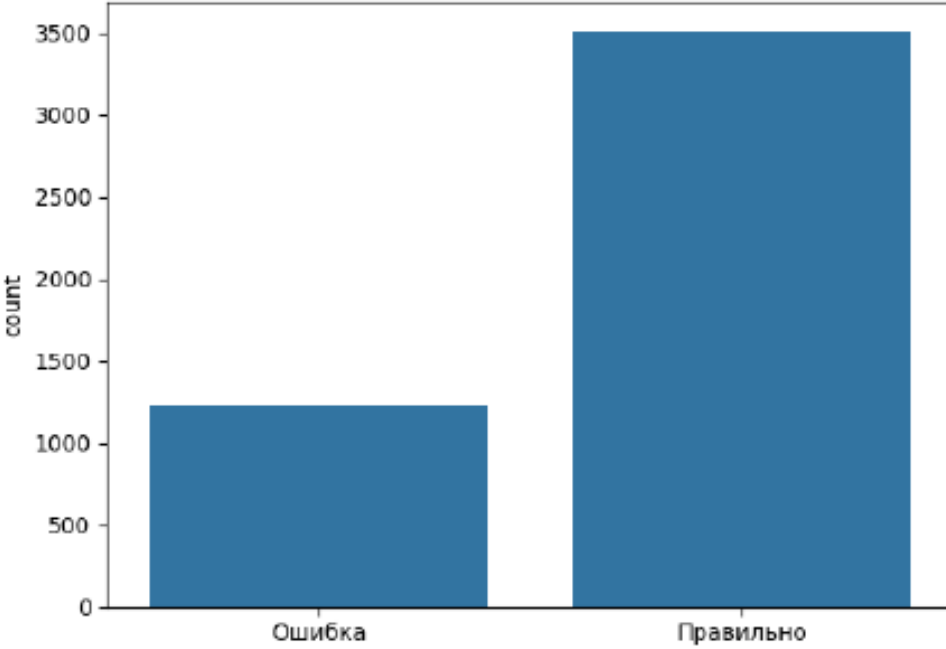
Gender Classification



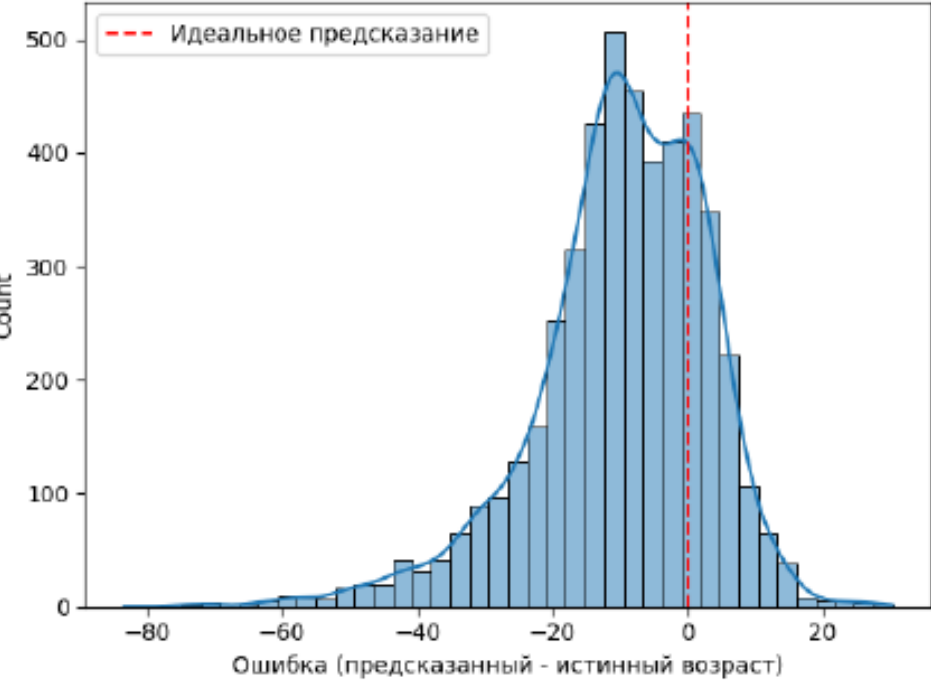
Ошибки предсказания пола



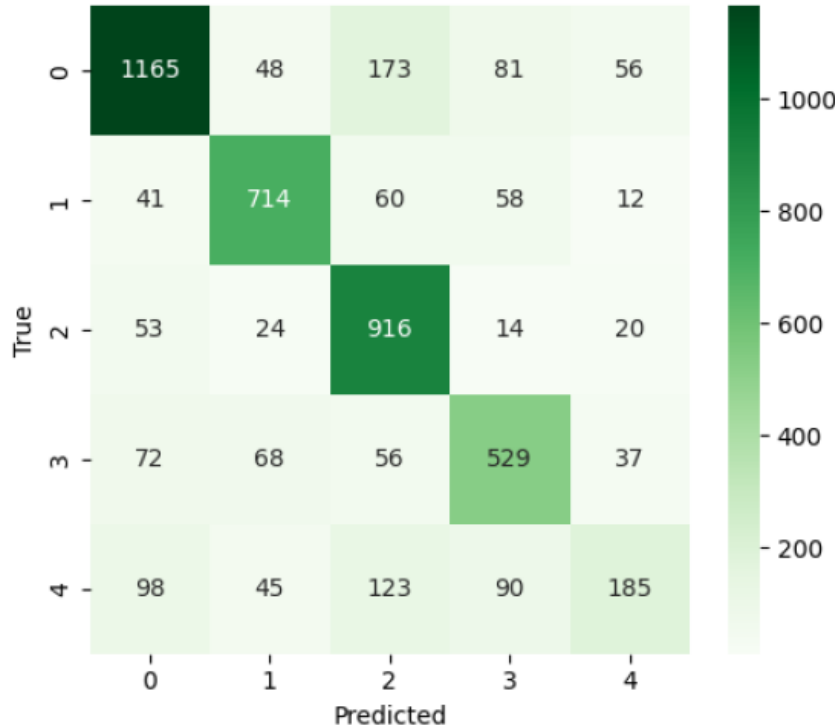
Ошибки предсказания этноса



Распределение ошибок возраста



Ethnicity Classification



Визуальный анализ

Истинно:
Пол: Мужчина
Этнос: Другой
Возраст: 50

Предсказано:
Пол: Мужчина
Этнос: Белый
Возраст: 45



Истинно:
Пол: Женщина
Этнос: Черный
Возраст: 32

Предсказано:
Пол: Женщина
Этнос: Черный
Возраст: 26



Истинно:
Пол: Мужчина
Этнос: Черный
Возраст: 46

Предсказано:
Пол: Мужчина
Этнос: Черный
Возраст: 46



Истинно:
Пол: Мужчина
Этнос: Индиец
Возраст: 70

Предсказано:
Пол: Мужчина
Этнос: Черный
Возраст: 54



Истинно:
Пол: Женщина
Этнос: Белый
Возраст: 47

Предсказано:
Пол: Женщина
Этнос: Белый
Возраст: 34



Истинно:
Пол: Мужчина
Этнос: Индиец
Возраст: 70

Предсказано:
Пол: Мужчина
Этнос: Индиец
Возраст: 40



Истинно:
Пол: Женщина
Этнос: Индиец
Возраст: 110

Предсказано:
Пол: Женщина
Этнос: Индиец
Возраст: 102



Истинно:
Пол: Мужчина
Этнос: Другой
Возраст: 82

Предсказано:
Пол: Мужчина
Этнос: Черный
Возраст: 64



Истинно:
Пол: Женщина
Этнос: Черный
Возраст: 54

Предсказано:
Пол: Мужчина
Этнос: Черный
Возраст: 56



Истинно:
Пол: Женщина
Этнос: Другой
Возраст: 6

Предсказано:
Пол: Мужчина
Этнос: Азиат
Возраст: 8



Проверка работоспособности

Оригинал

Пол: Мужчина
Этнос: Белый
Возраст: 26



Пол: Мужчина
Этнос: Белый
Возраст: 32



Пол: Женщина
Этнос: Белый
Возраст: 46



Пол: Женщина
Этнос: Белый
Возраст: 14



Пол ✗
Этнос ✓
Возраст ✗

Пол: Мужчина
Этнос: Белый
Возраст: 33



Пол ✓
Этнос ✓
Возраст ✓

Пол: Женщина
Этнос: Белый
Возраст: 71



Пол ✓
Этнос ✓
Возраст ✗

Модель



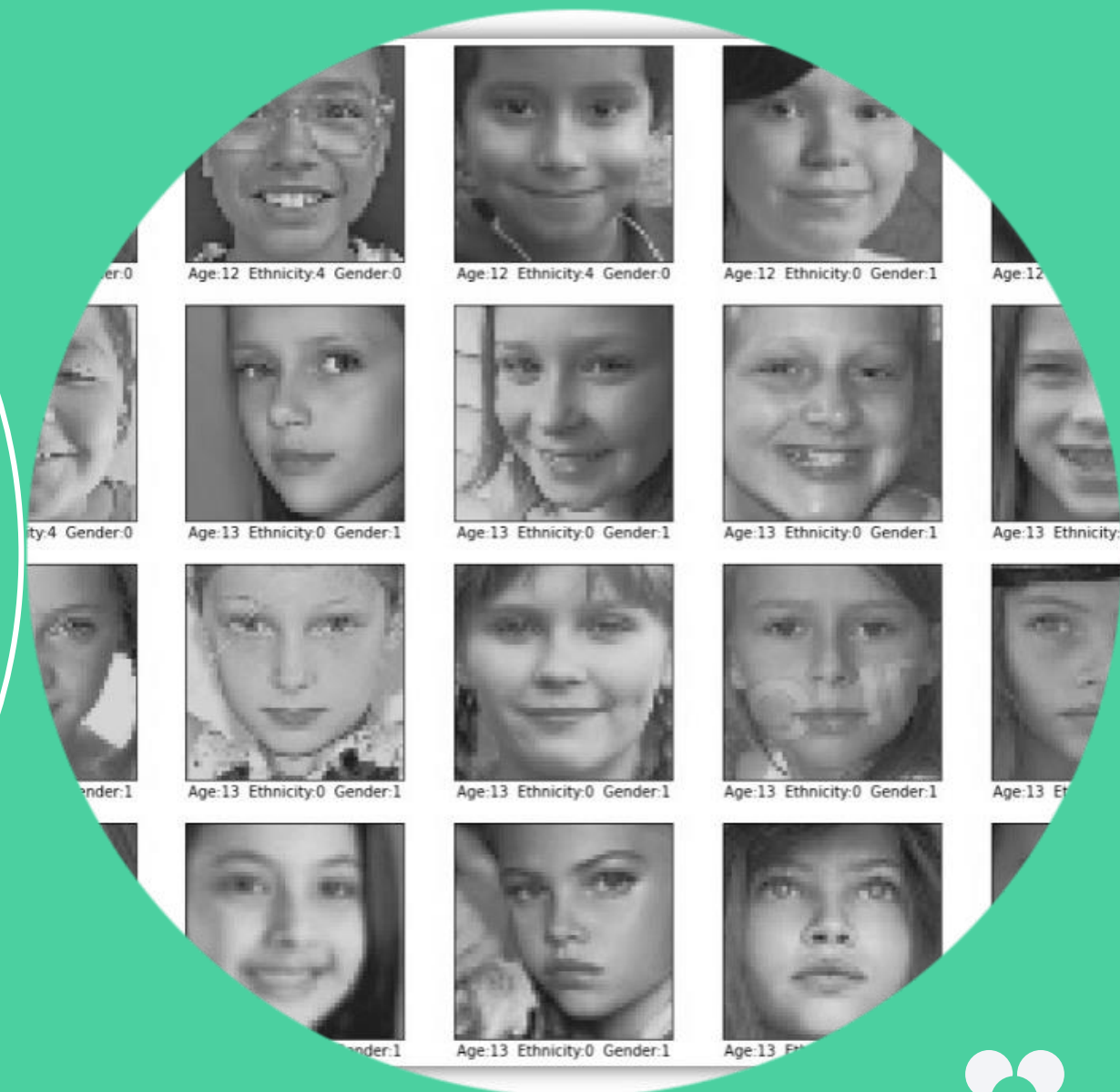
Направления по улучшению

- Использование более современных архитектур с механизмами внимания (ConvNeX и др.)
- Использование трансформеров (ViT, Swin Transformer)
- Раздельные энкодеры для разных задач
- Более тщательная балансировка данных по всем задачам
- Раздельная оптимизация для разных задач (возможно, с разными learning rate)
- Увеличение количества эпох обучения



Генерация лиц по заданным атрибутам

4



Модели GAN

1



Параметры моделей

Модель	LATENT_DIM	BATCH_SIZE	EPOCHS	Attention	SpectralNorm	LR Generator	LR Discriminator	Оптимизатор
GAN-1	100	256	200	Нет	Нет	1e-4	1e-4	Adam
GAN-2	256	256	200	Нет	Нет	1e-4	2e-4	Adam
GAN-3	256	64	135	Self + Cross	Нет	1e-4	2e-4	Adam
GAN-4	256	32	130	Self + Cross	Нет	1e-5	2e-5	Adam
GAN-5	256	32	140	Self + Cross	Нет	1e-5	4e-5	Adam

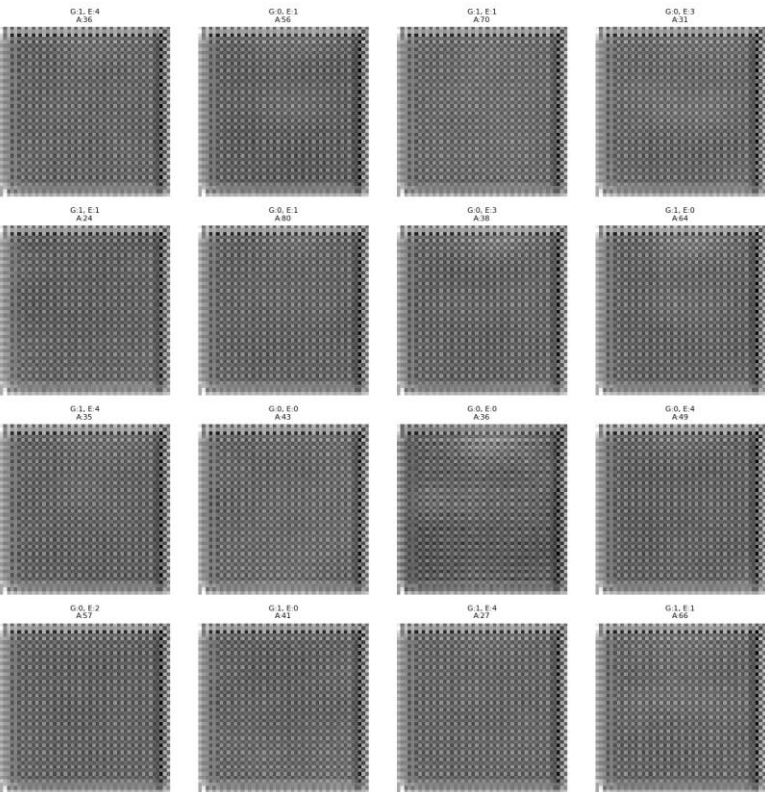
- Модели учитывают возраст, пол и этническую принадлежность при генерации
- Многозадачный дискриминатор: помимо классификации реальное/фейковое, дискриминатор предсказывает атрибуты



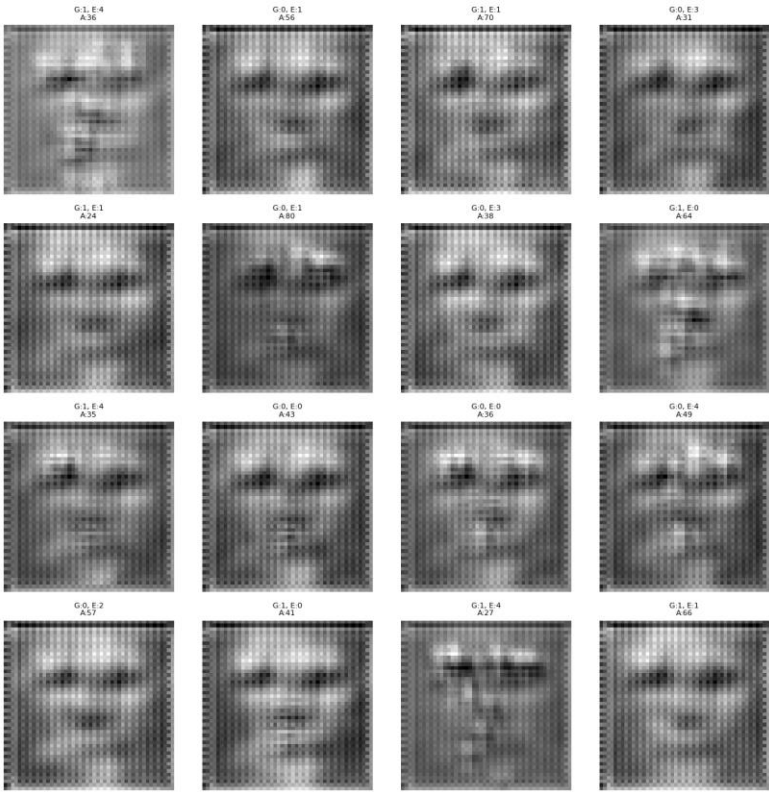
Итоги обучения

GAN-3

Эпоха 1



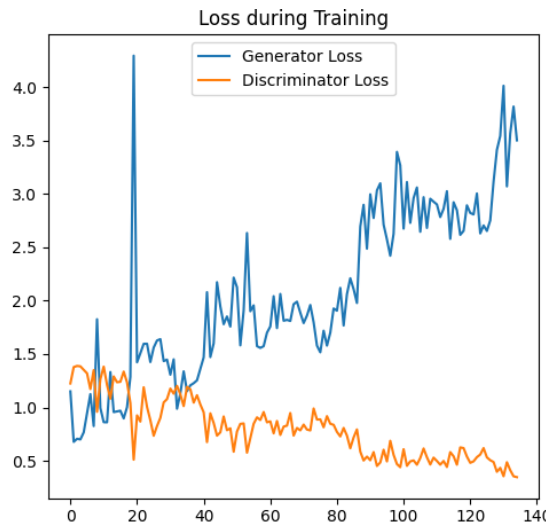
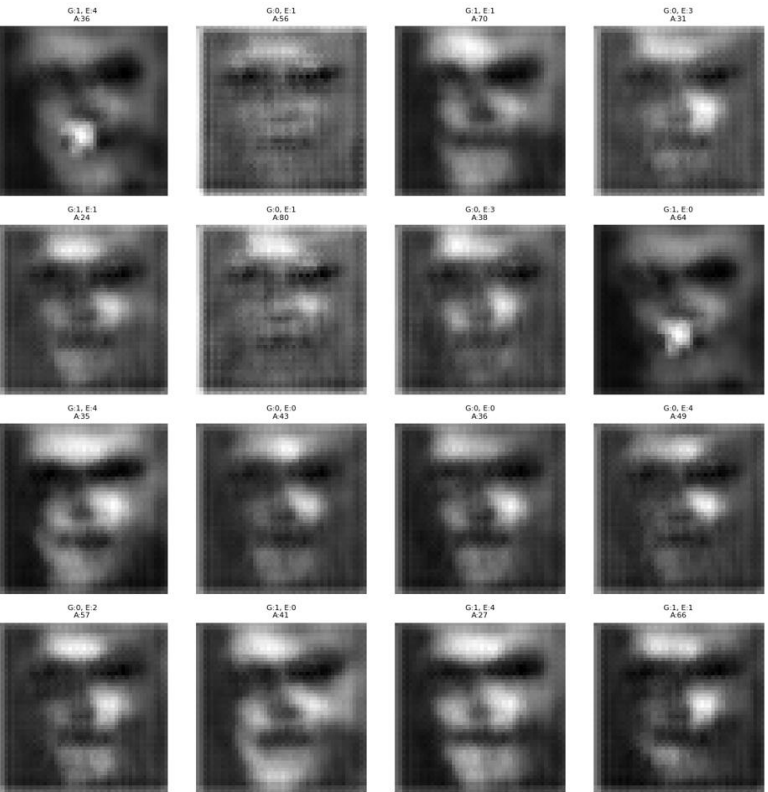
Эпоха 40



Эпоха 125

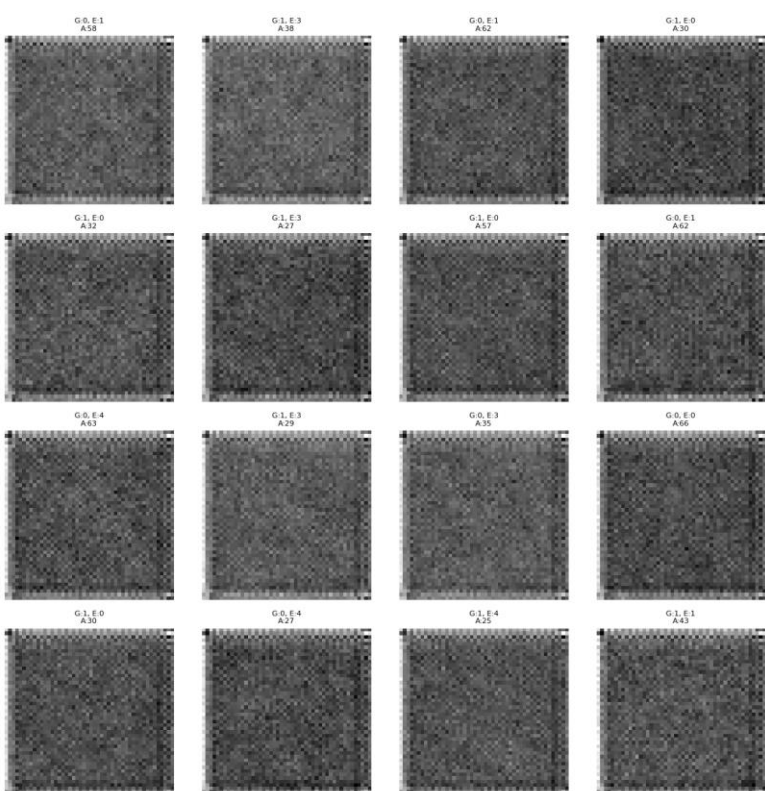


Эпоха 135

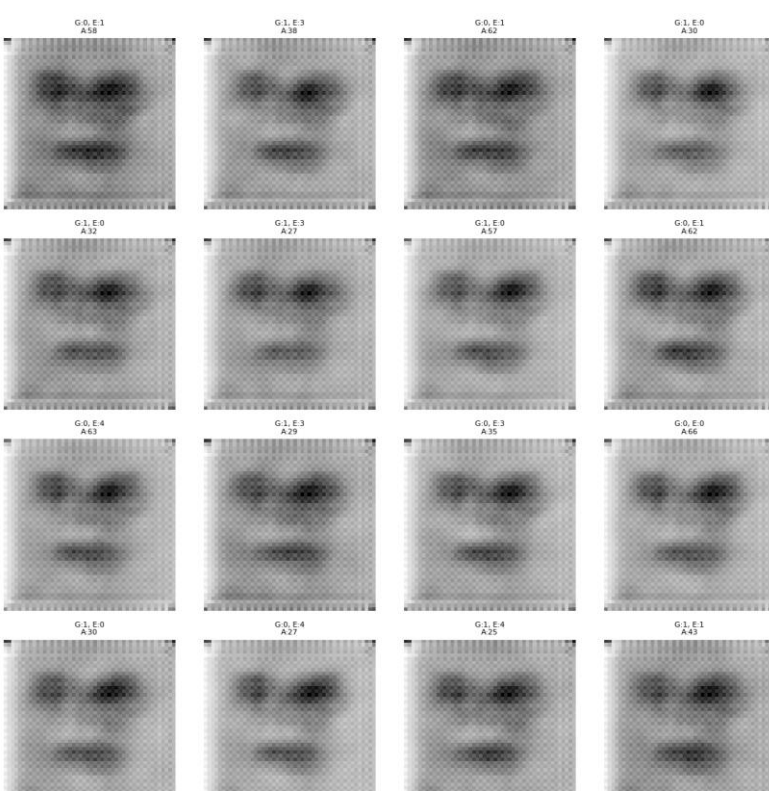


GAN-4

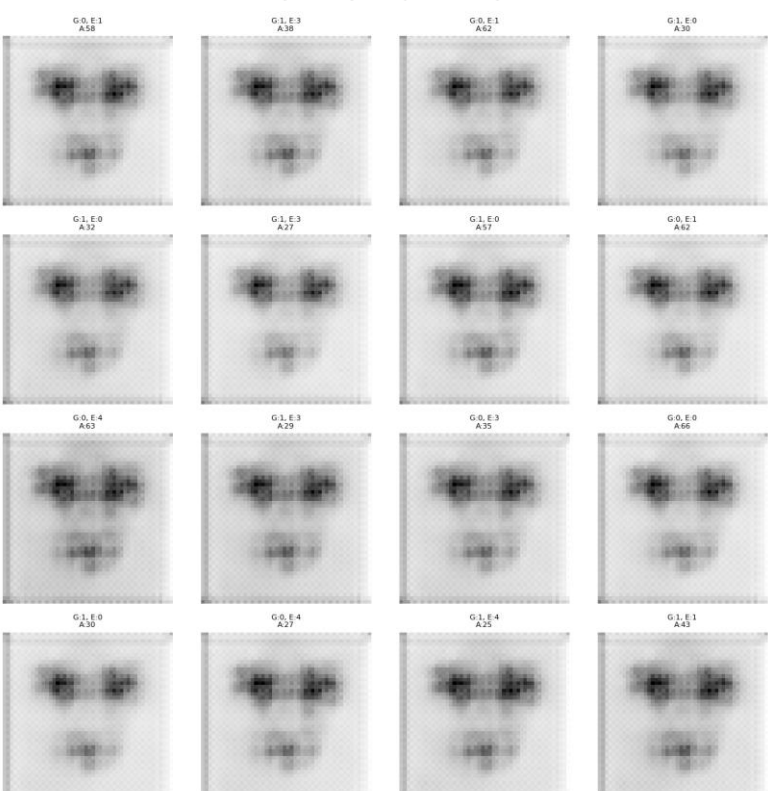
Эпоха 1



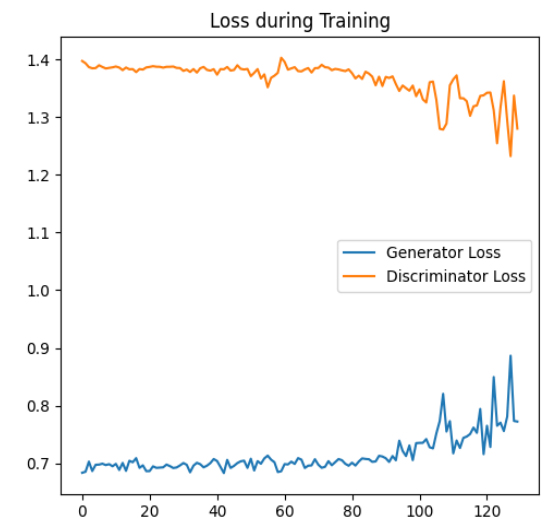
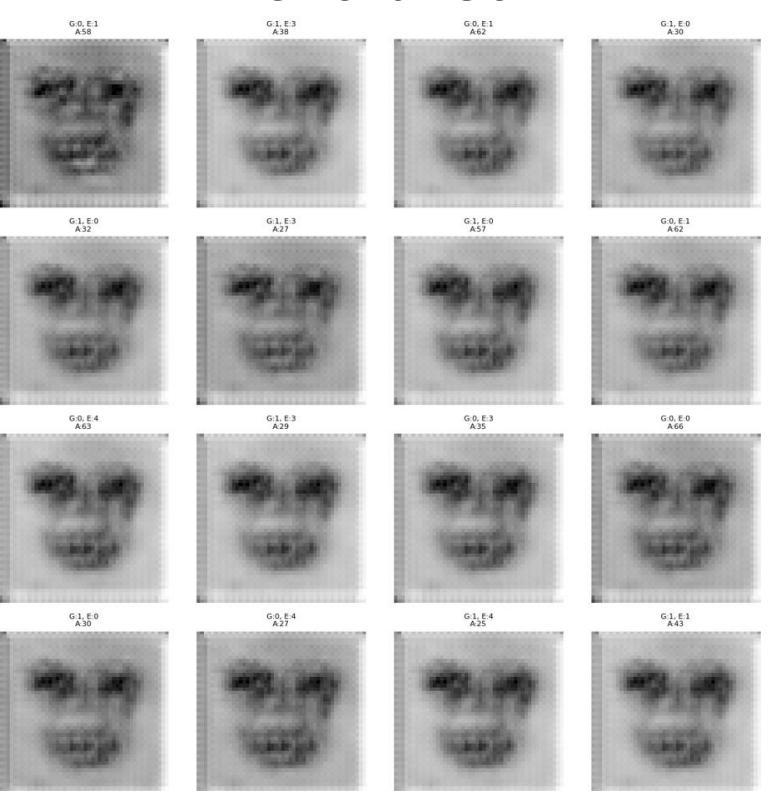
Эпоха 80



Эпоха 110



Эпоха 130



Модели DCGAN

2



Параметры моделей

Параметр	Значение	Описание
Тип модели	Deep Convolutional GAN (DCGAN)	Состязательная сеть со сверточными слоями
Вход генератора	100-мерный вектор	Нормальное распределение $N(0,1)$
Выход генератора	48×48×1	Grayscale изображения в диапазоне [-1, 1]
Вход дискриминатора	48×48×1	Реальные или сгенерированные изображения
Выход дискриминатора	Скаляр [0,1]	Вероятность реальности изображения
Batch Size	32	Оптимальный баланс стабильности/скорости
Learning Rate (G)	0.0003	Быстрее обучение генератора
Learning Rate (D)	0.0001	Медленнее обучение дискриминатора
Beta1 (Adam)	0.5	Стандарт для GAN (меньше momentum)
Эпохи	50	Достаточно для сходимости
Loss Function	Binary Crossentropy	Стандарт для бинарной классификации

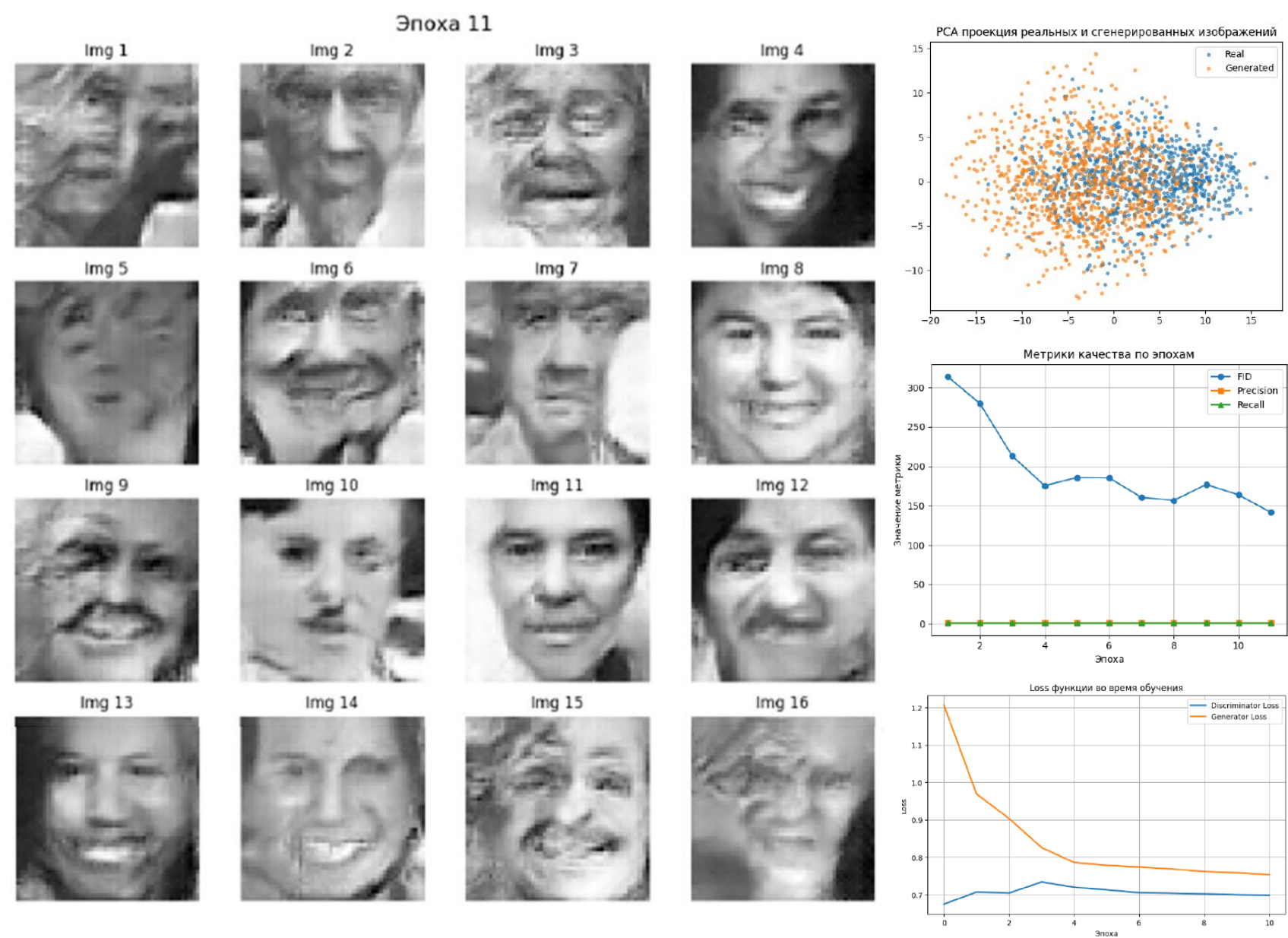
Две модели:

- DCGAN
- DCGAN с Attention:
 - 3 Attention слоя в генератор
 - 2 Attention слоя в дискриминатор

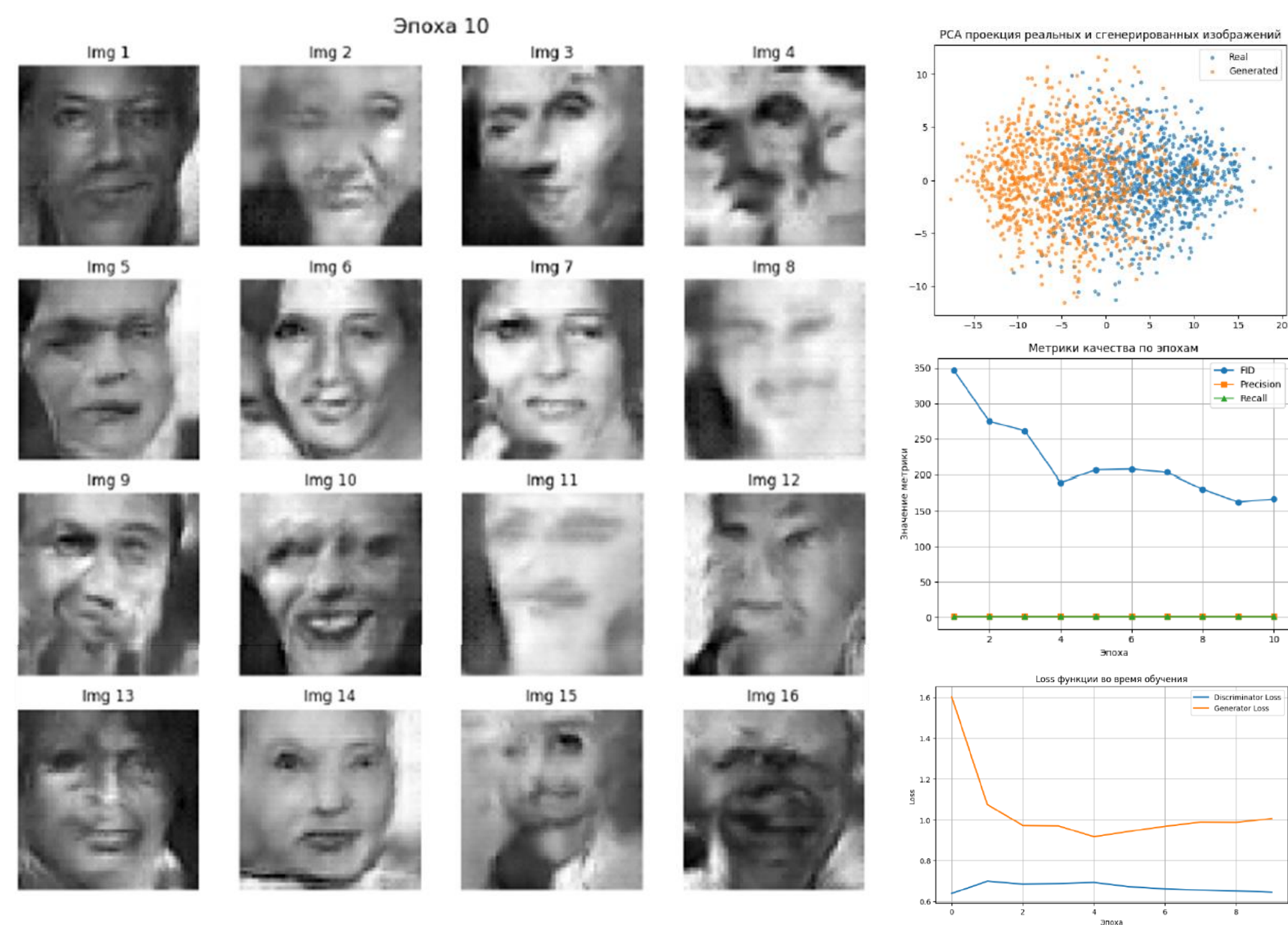


Итоги обучения

Модель DCGAN (без Attention)



Модель DCGAN (с Attention)



Гибридная модель: условная Deep Convolutional GAN (cDCGAN)

3



Параметры модели

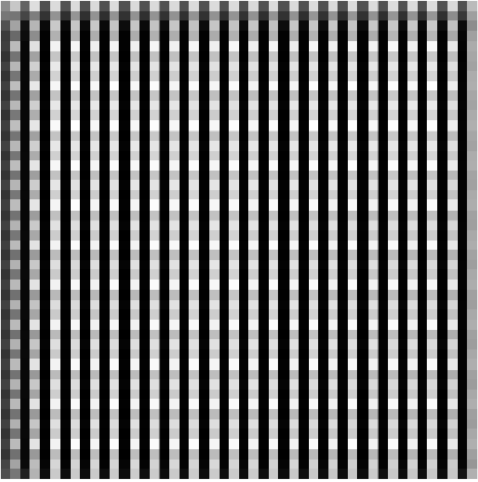
Параметр	Значение		Описание
Основные параметры	IMG_SIZE	48x48	Размер изображения
	CHANNELS	1	Количество каналов (1 - grayscale)
	LATENT_DIM	256	Размерность латентного пространства
	BATCH_SIZE	32	Размер батча для обучения
	EPOCHS	80	Количество эпох обучения
Архитектурные параметры	weight_init	RandomNormal(mean=0.0, stddev=0.02)	Инициализация весов
	NUM_CLASSES['gender']	2	Пол: 0=мужчина, 1=женщина
	NUM_CLASSES['ethnicity']	5	Этническая принадлежность: 0-4 категории
	NUM_CLASSES['age']	1	Возраст: регрессия (нормализованный 0-1)
Генератор	Входной слой	256 + 12 + 12 + 1 = 281	Объединенный вход (шум + эмбединги)
	Dense слой	12×12×512 = 73728 нейронов	Начальный полносвязный слой
	Reshape	(12, 12, 512)	Преобразование в пространственную структуру
	Conv2DTranspose слои	256→128→64 фильтров	Последовательность апсемплинга
	Выходной слой	(48, 48, 1)	Сгенерированное изображение
Дискриминатор	Conv2D слои	64→128→256 фильтров	Последовательность даунсемплинга
	Embedding размер	10	Размерность эмбедингов для атрибутов
	Выходы	4	[validity, gender, ethnicity, age]
Оптимизаторы	generator_optimizer	Adam(lr=0.0001, β ₁ =0.5)	Оптимизатор генератора
	discriminator_optimizer	Adam(lr=0.0002, β ₁ =0.5)	Оптимизатор дискриминатора
Функции потерь	Состязательная	BinaryCrossentropy	Реальный/фейковый
	Пол	BinaryCrossentropy	Бинарная классификация
	Стнос	SparseCategoricalCrossentropy	Многоклассовая классификация
	Возраст	MeanSquaredError	Регрессия
Параметры обучения	Dropout rate	0.3	Вероятность dropout в дискриминаторе
	LeakyReLU alpha	0.2	Коэффициент для LeakyReLU



Итоги обучения

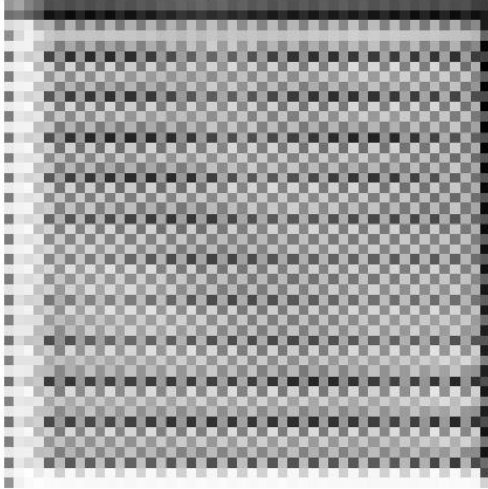
Эпоха 1

G:0, E:3, A:0.90



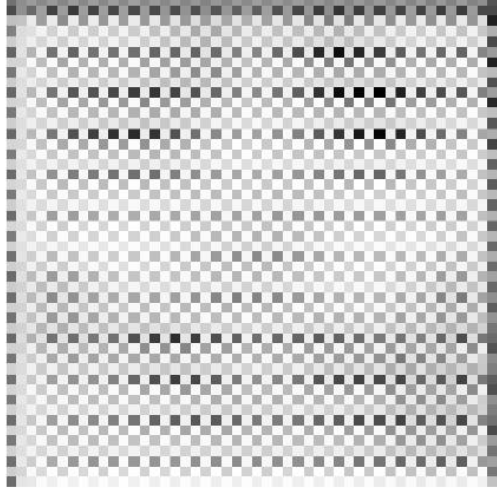
Эпоха 21

G:0, E:3, A:0.90



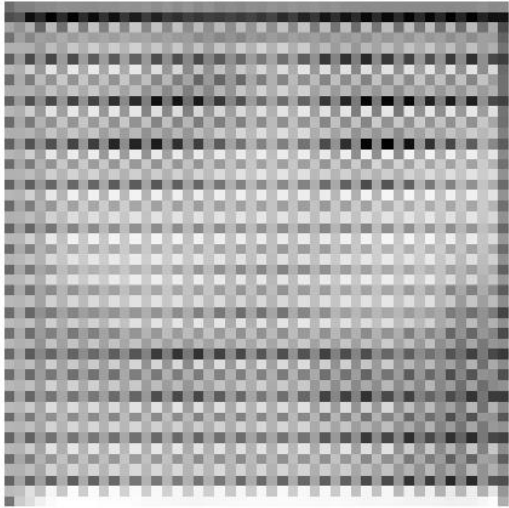
Эпоха 41

G:0, E:3, A:0.90



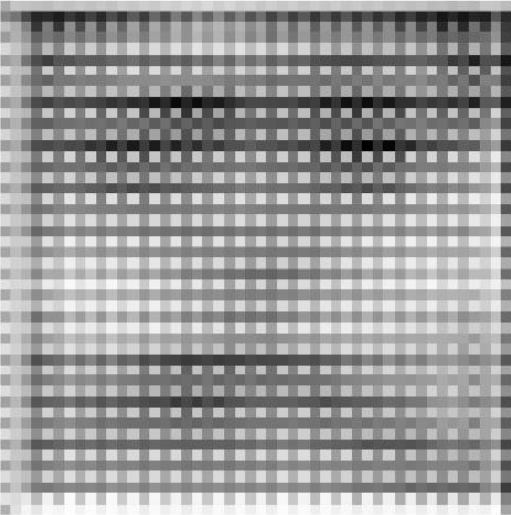
Эпоха 51

G:0, E:3, A:0.90



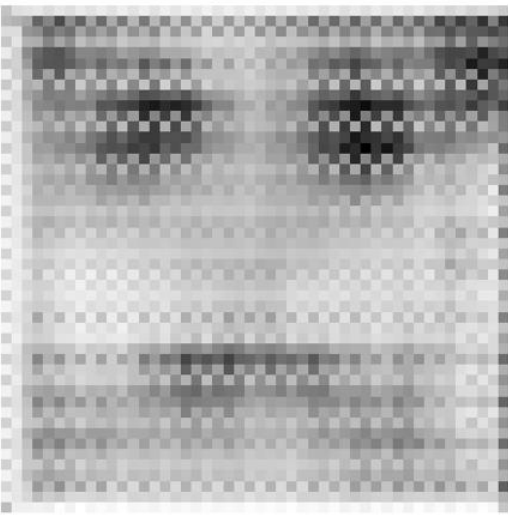
Эпоха 61

G:0, E:3, A:0.90

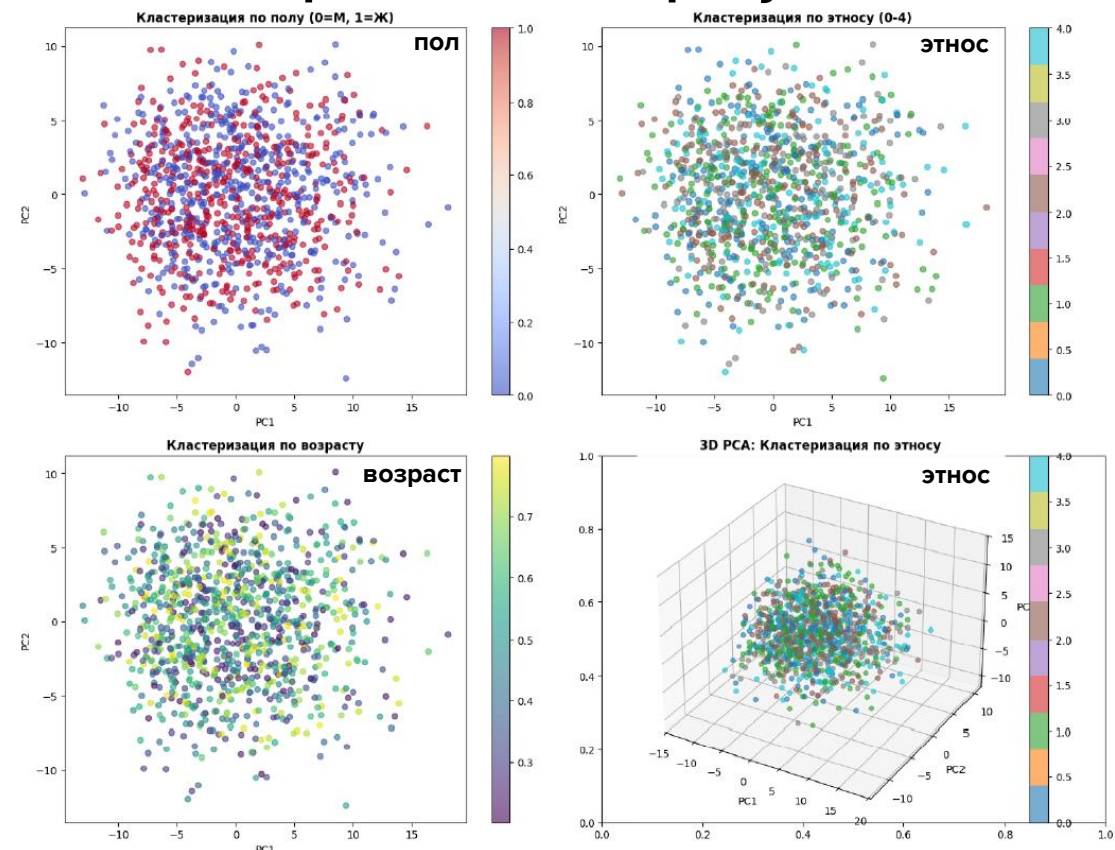


Эпоха 71

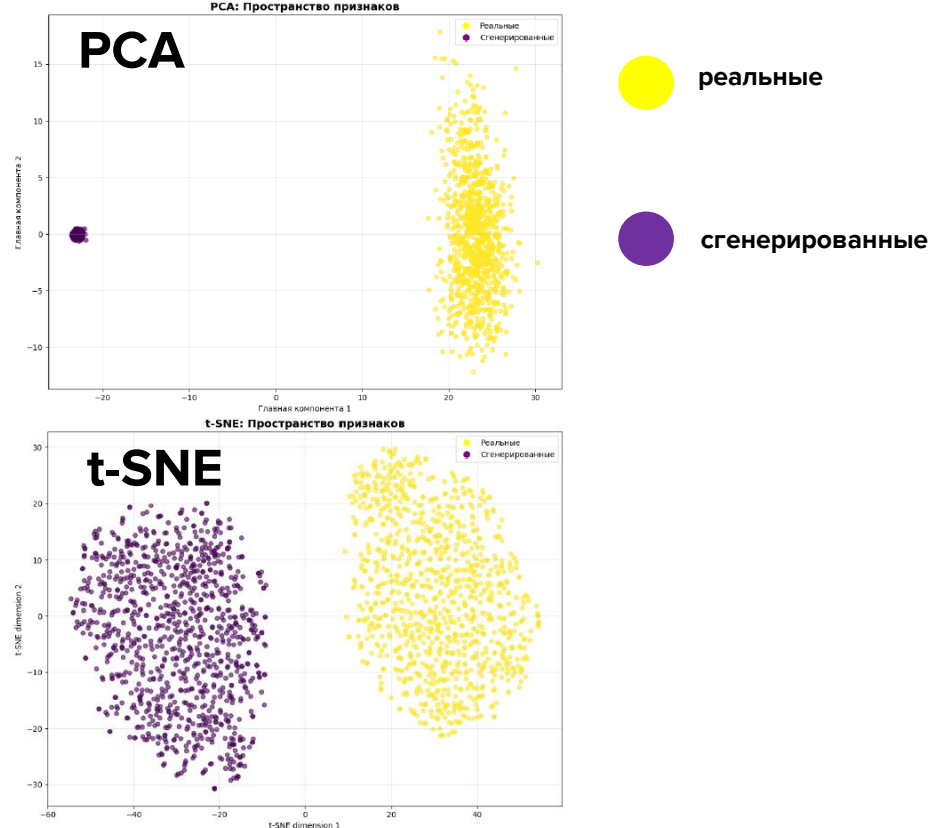
G:0, E:3, A:0.90



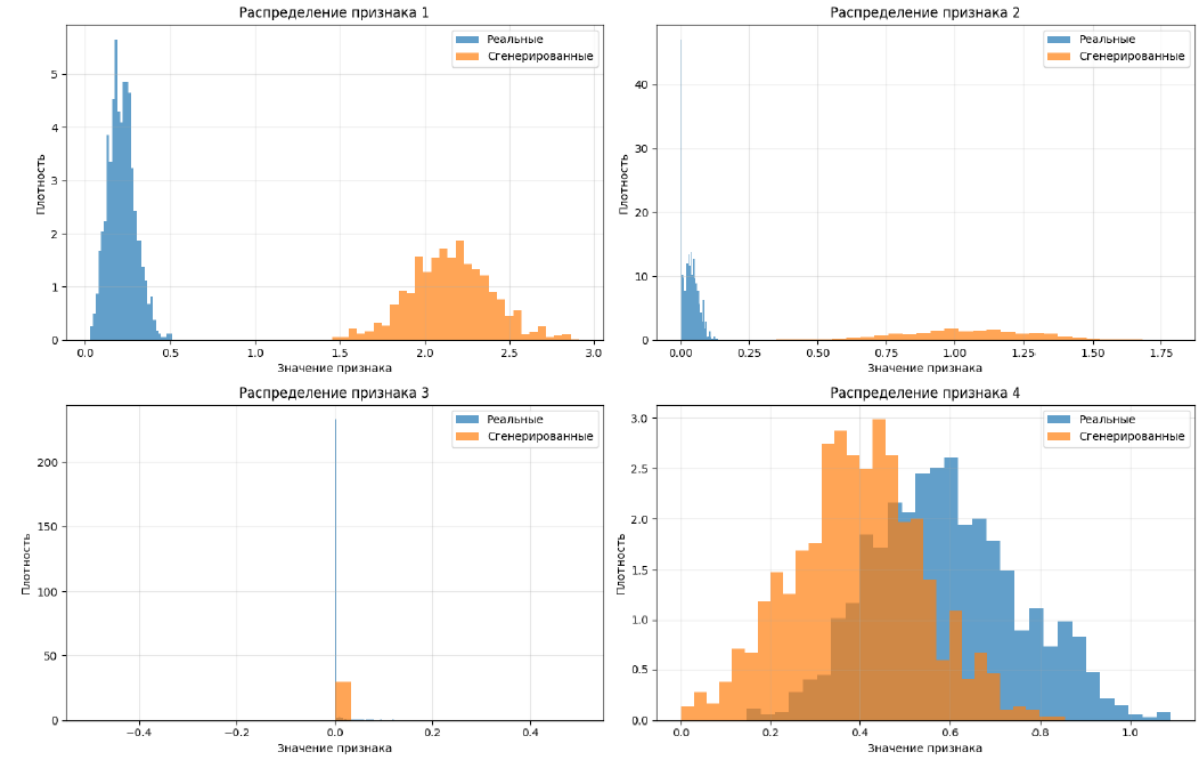
Анализ кластеризации сгенерированных изображений по атрибутам



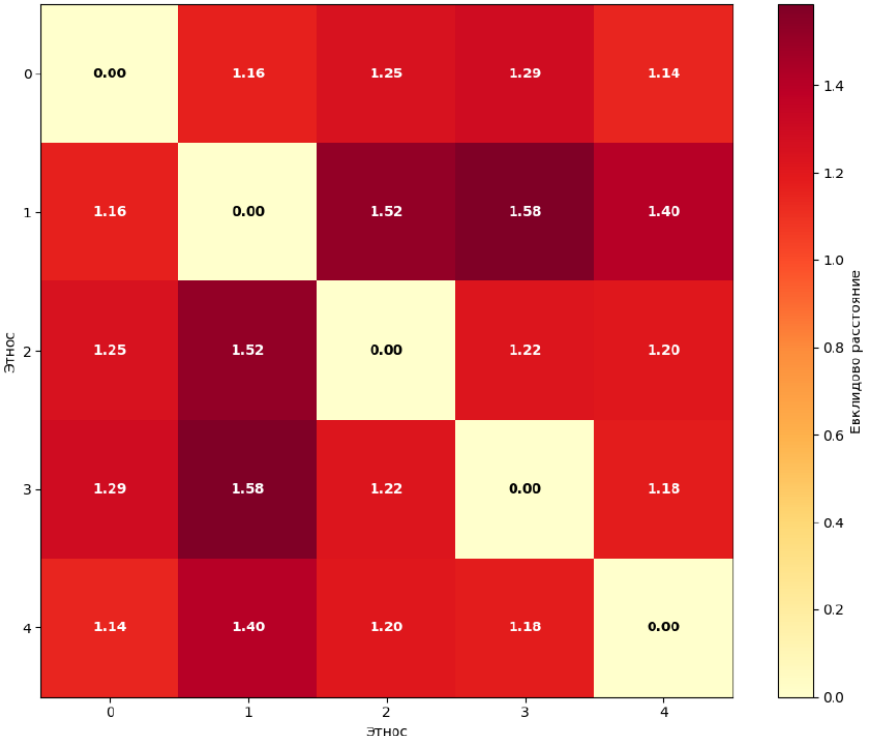
Сравнение реальных и сгенерированных изображений в пространстве признаков



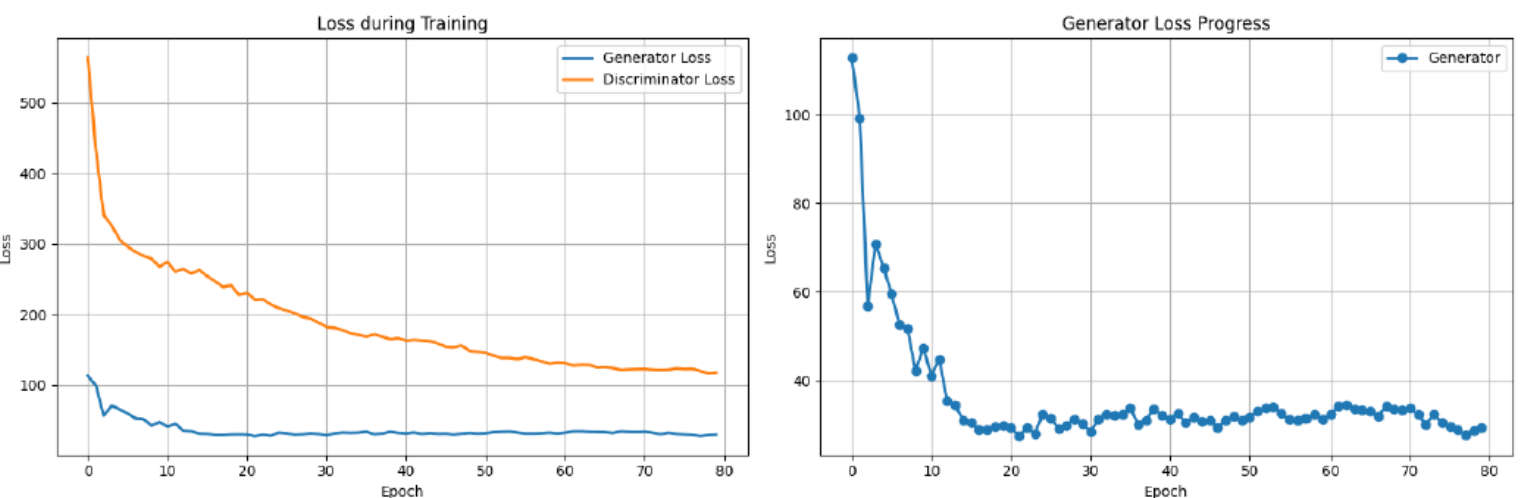
Сравнение распределений признаков



Матрица расстояний между кластерами этносов



Динамика Loss



Проверка работоспособности

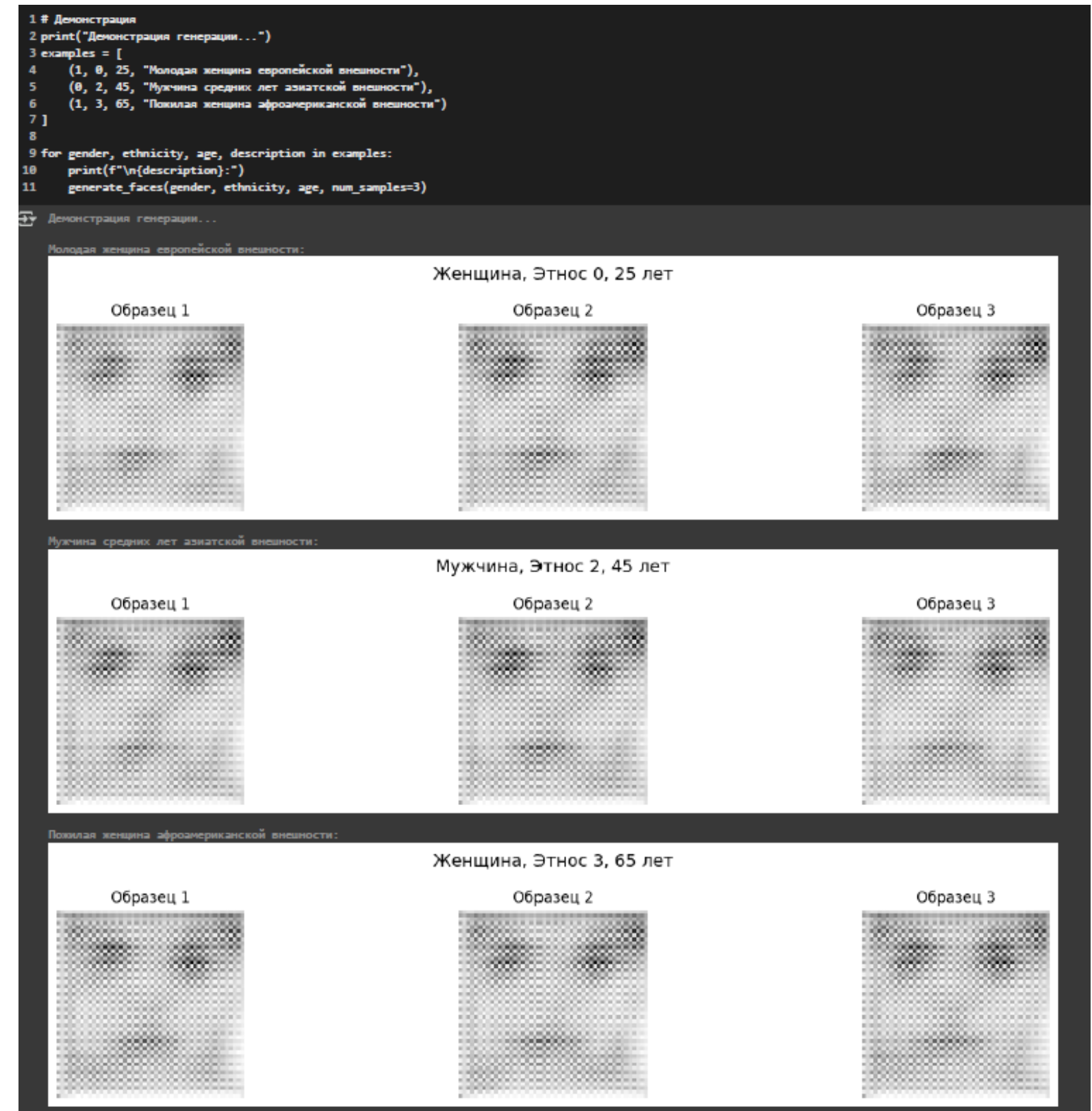
GAN-3



GAN-4



cDCGAN



- Модели учитывают возраст, пол и этническую принадлежность при генерации
- Многозадачный дискриминатор: помимо классификации реальное/фейковое, дискриминатор предсказывает атрибуты
- В целом модели рабочие и генерируют лица по заданным атрибутам, но являются недообученными по причине инфраструктурных ограничений



Направления по улучшению

1. Увеличение разрешения:

- Использовать прогрессивное наращивание разрешения (ProGAN)
- Применить архитектуру StyleGAN для лучшей детализации

2. Улучшение стабильности обучения:

- Использовать WGAN-GP с градиентным штрафом
- Применить TTUR (Two Time-scale Update Rule)

3. Оптимизация гиперпараметров:

- Поиск оптимального размера скрытого пространства
- Настройка коэффициентов для разных компонентов потерь

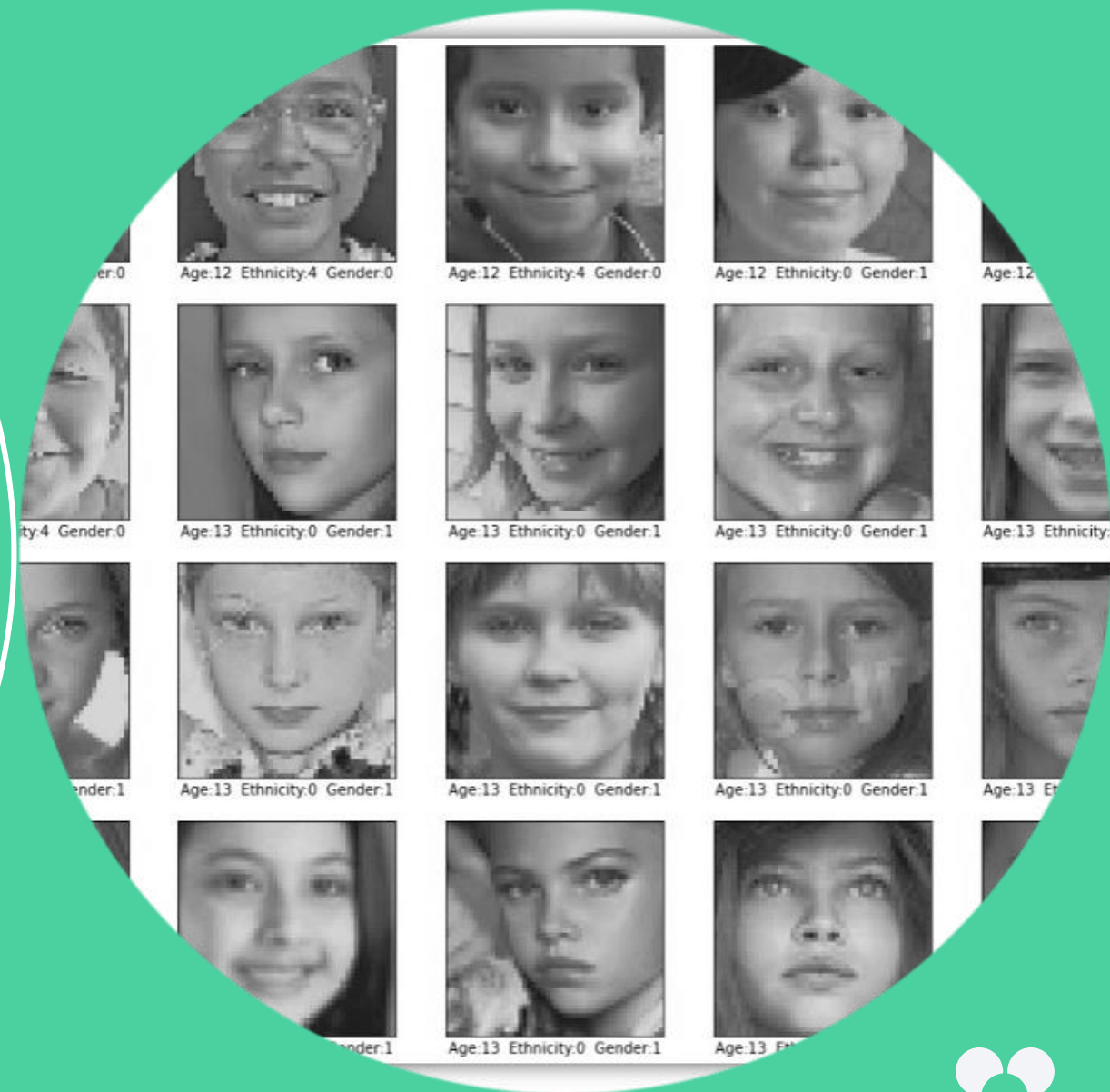
4. Увеличение количества эпох обучения моделей:

- Обучить модели не менее чем на 1000-х эпохах



Выводы

5



Выводы

- В дипломной работе были решены две задачи:
 1. определение пола, возраста и этноса человека по фотографии
 2. генерация лиц по заданным атрибутам
- Качество выполнения задач контролировалось использованием метрик: Accuracy, Precision, Recall, F1 score, MAE, RMSE и FID
- В рамках решения первой задачи было построено пять различных архитектур моделей:
 - ResNet50
 - EfficientNetB0
 - CNN (Multi-Task)
 - CNN + Attention
 - SNN (на базе EfficientNetB0)
- Наилучшей оказалась модель простой обычной сверточной нейронной сети, созданной с нуля специально для данной задачи без предобученных моделей на основе Multi-Task Learning (MTL) (модель CNN) с добавлением Attention-слоя
- Для решения задачи генерации лиц по заданным атрибутам рассмотрены модели GAN, DCGAN и гибридная модель: условная Deep Convolutional GAN (cDCGAN) для генерации лиц по заданным атрибутам
- Даны рекомендации по дальнейшему улучшению лучших моделей



Спасибо
за внимание!



Определение пола, возраста и этноса человека по фотографии и генерация лиц по заданным атрибутам с применением нейронных сетей и методов компьютерного зрения

Дипломный проект по профессии «Data Scientist»

Каторгин Иван
Группа DSU-PROD-67



Ivan.katorgin@gmail.com



t.me/IvanKatorgin

