

Бакалаврская работа

**Транскодирование JPEG-изображений при  
помощи предсказания коэффициентов ДКП на  
основе нейронной сети**

---

Автор: Подцепко Игорь Сергеевич, М34381

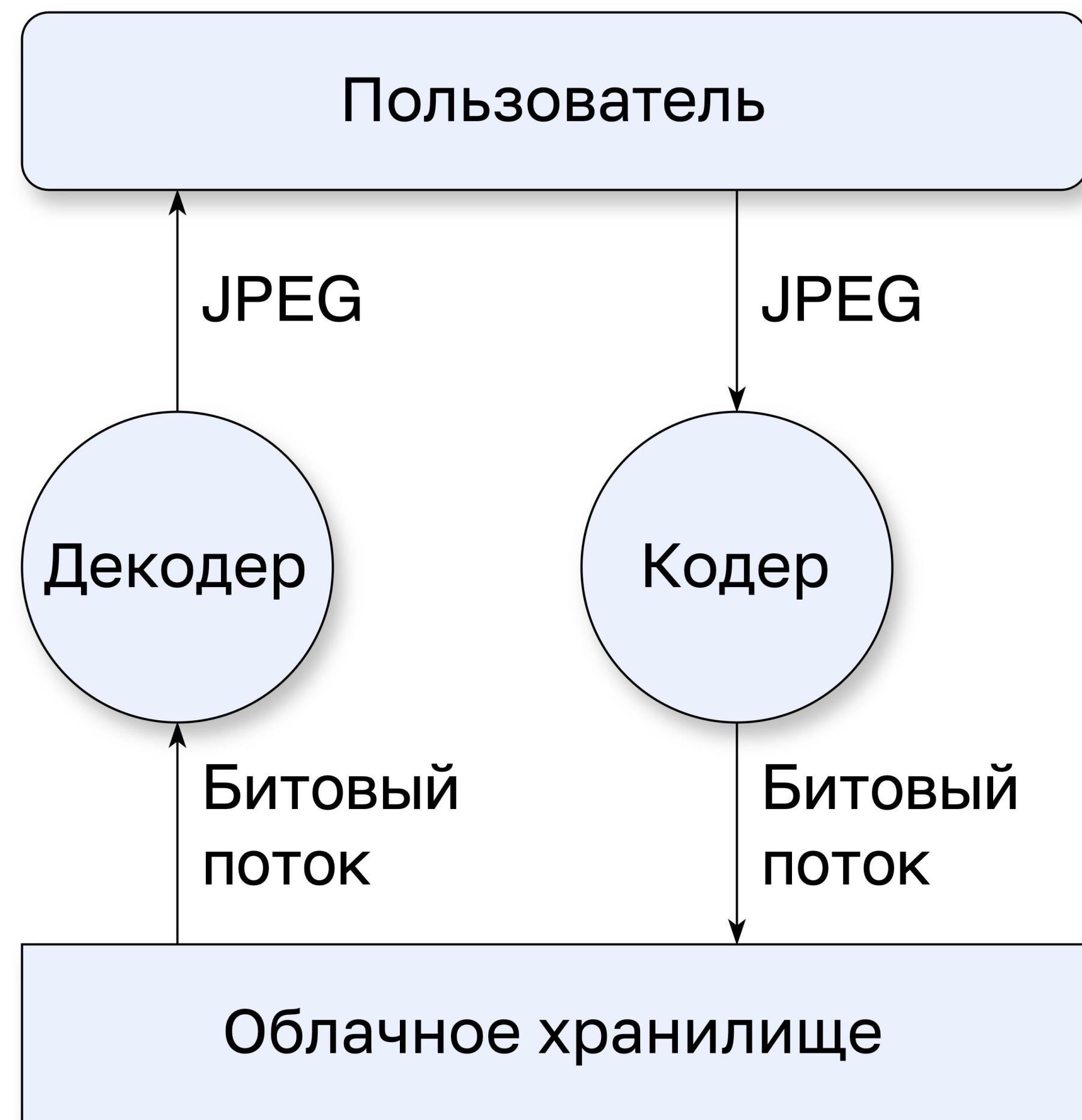
Научный руководитель: Беляев Евгений Александрович, PhD, науки,  
Университет ИТМО, ФИТИП, доцент

Преимущества стандарта JPEG:

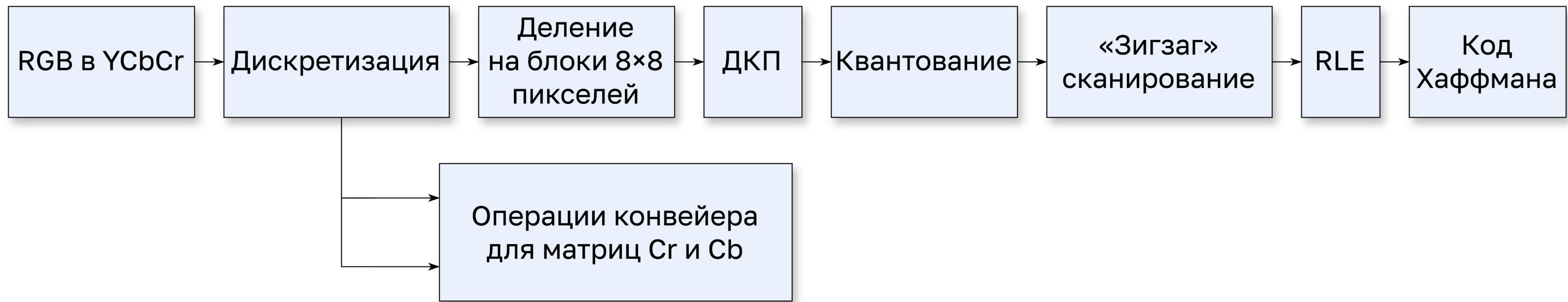
1. Один из первых (1992 г.);
2. С ним «умеет» работать большинство программ;
3. Имеет низкую вычислительную сложность.

Главный его недостаток – это **низкая эффективность сжатия** по сравнению с более современными стандартами (JPEG2000, WebP, BPG, FLIF и другими).

# Архитектура внедрения



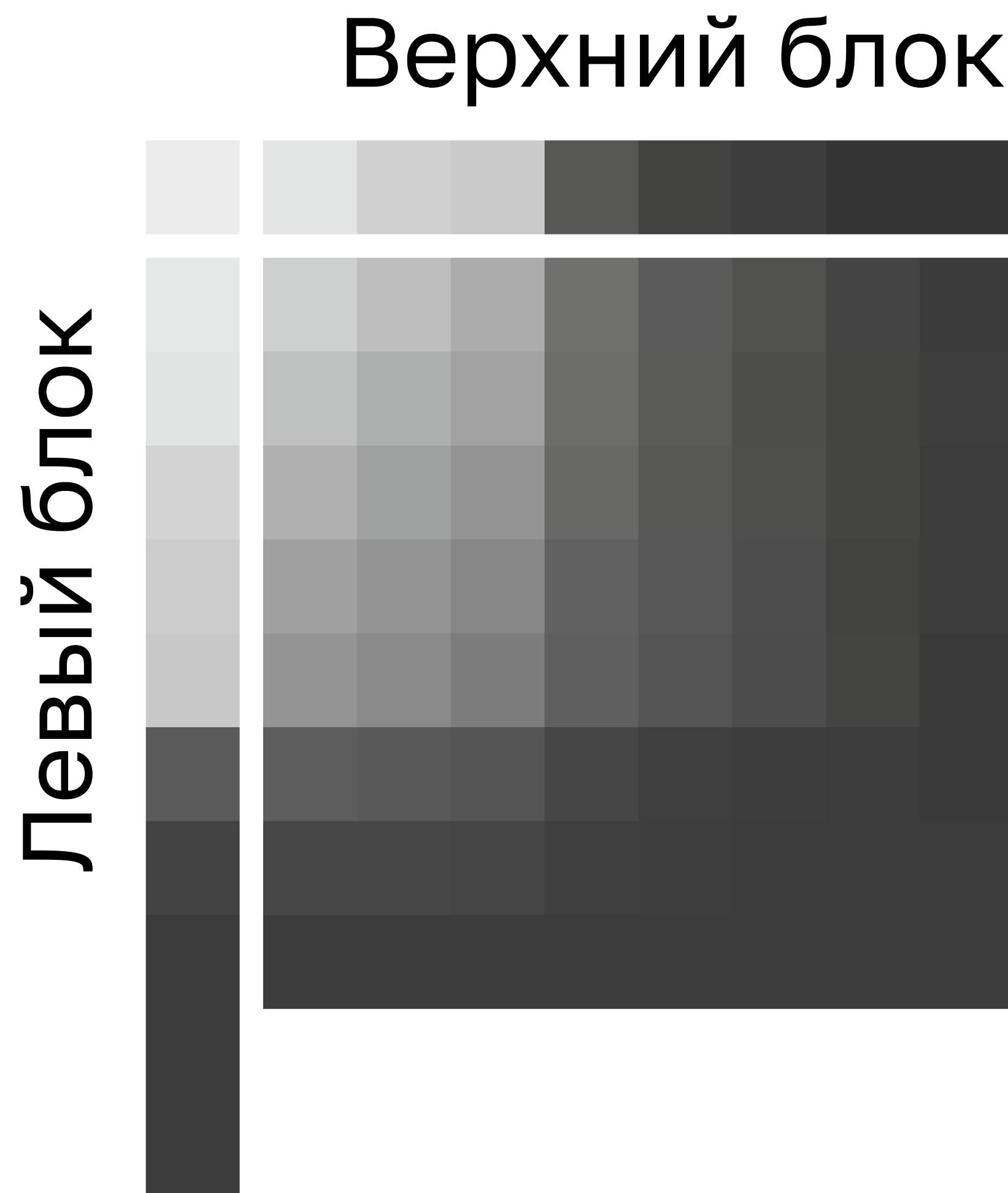
# Алгоритм JPEG



1. Замена кода Хаффмана на арифметический код;
2. Отказ от фиксированного размера блоков;
3. Реализация предсказания коэффициентов ДКП.

# Пример внутреннего предсказания

ИТМО



Значения вычисляются с помощью некоторой функции от коэффициентов верхнего и левого блоков.

Такой подход учитывает только **локальную** схожесть.

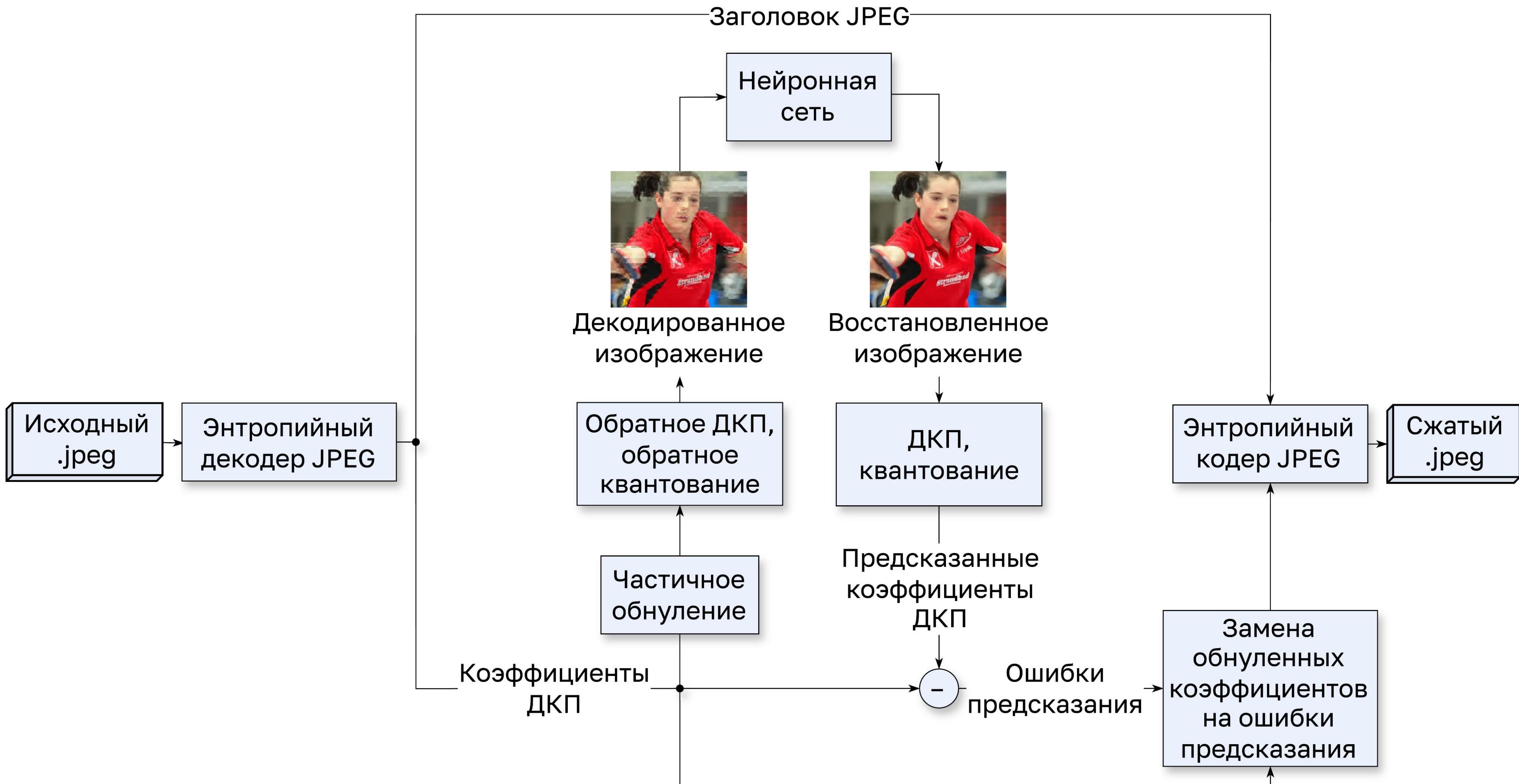
---

**Цель:** изучить возможность и потенциал использования нейронных сетей для реализации этапа предсказания коэффициентов ДКП при транскодировании JPEG-изображений.

## Задачи:

1. Реализовать декодирование JPEG с обнулением части коэффициентов ДКП;
2. Обучить нейронную сеть восстанавливать эти коэффициенты;
3. Реализовать кодирование ошибок предсказания;
4. Оценить степень сжатия, которую удается достичь благодаря нейросетевому предсказанию коэффициентов.

# Схема транскодера



# Блок после квантования

91	21	-5	0	0	2	1	0
-7	12	2	-4	-5	0	0	0
14	-1	-5	0	1	0	-1	0
1	0	3	1	-1	0	-2	0
0	-3	-1	0	0	1	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	-1	0	0	0	0	0	0

# Преобразования над блоком

ІТМО

**Stage A:**

91	21	-5	0	0	2	<b>1</b>	0	0
-7	12	2	-4	<b>-5</b>	0	0	0	0
14	-1	<b>-5</b>	0	1	0	-1	0	0
1	0	3	<b>1</b>	-1	0	<b>-2</b>	0	0
0	<b>-3</b>	-1	0	0	<b>1</b>	0	0	0
0	1	0	0	0	0	0	0	0
0	0	0	<b>0</b>	0	0	<b>0</b>	0	0
0	<b>-1</b>	0	0	0	0	0	0	0

**Stage B:**

91	21	-5	0	0	2	<b>0</b>	0	0
-7	12	2	-4	<b>0</b>	0	0	0	0
14	-1	<b>0</b>	0	1	0	-1	0	0
1	0	3	<b>0</b>	-1	0	<b>0</b>	0	0
0	<b>0</b>	-1	0	0	<b>0</b>	0	0	0
0	1	0	0	0	0	0	0	0
0	0	0	<b>0</b>	0	0	<b>0</b>	0	0
0	<b>0</b>	0	0	0	0	0	0	0

**Stage C:**

91	21	-5	0	0	2	<b>1</b>	0	0
-7	12	2	-4	<b>-4</b>	0	0	0	0
14	-1	<b>-5</b>	0	1	0	-1	0	0
1	0	3	<b>0</b>	-1	0	<b>-1</b>	0	0
0	<b>-4</b>	-1	0	0	<b>1</b>	0	0	0
0	1	0	0	0	0	0	0	0
0	0	0	<b>0</b>	0	0	<b>0</b>	0	0
0	<b>-1</b>	0	0	0	0	0	0	0

21	47	-5	0	0	2	<b>0</b>	0	0
-7	12	2	-4	<b>-1</b>	0	0	0	0
14	-1	<b>0</b>	0	1	0	-1	0	0
1	0	3	<b>1</b>	-1	0	<b>-1</b>	0	0
0	<b>1</b>	-1	0	0	<b>0</b>	0	0	0
0	1	0	0	0	0	0	0	0
0	0	0	<b>0</b>	0	0	<b>0</b>	0	0
0	<b>0</b>	0	0	0	0	0	0	0



Исходный набор данных: ImageNet.

**Объем:**

- 35'000 – для обучения;
- 10'000 – для валидации;
- 5'000 – для тестирования.

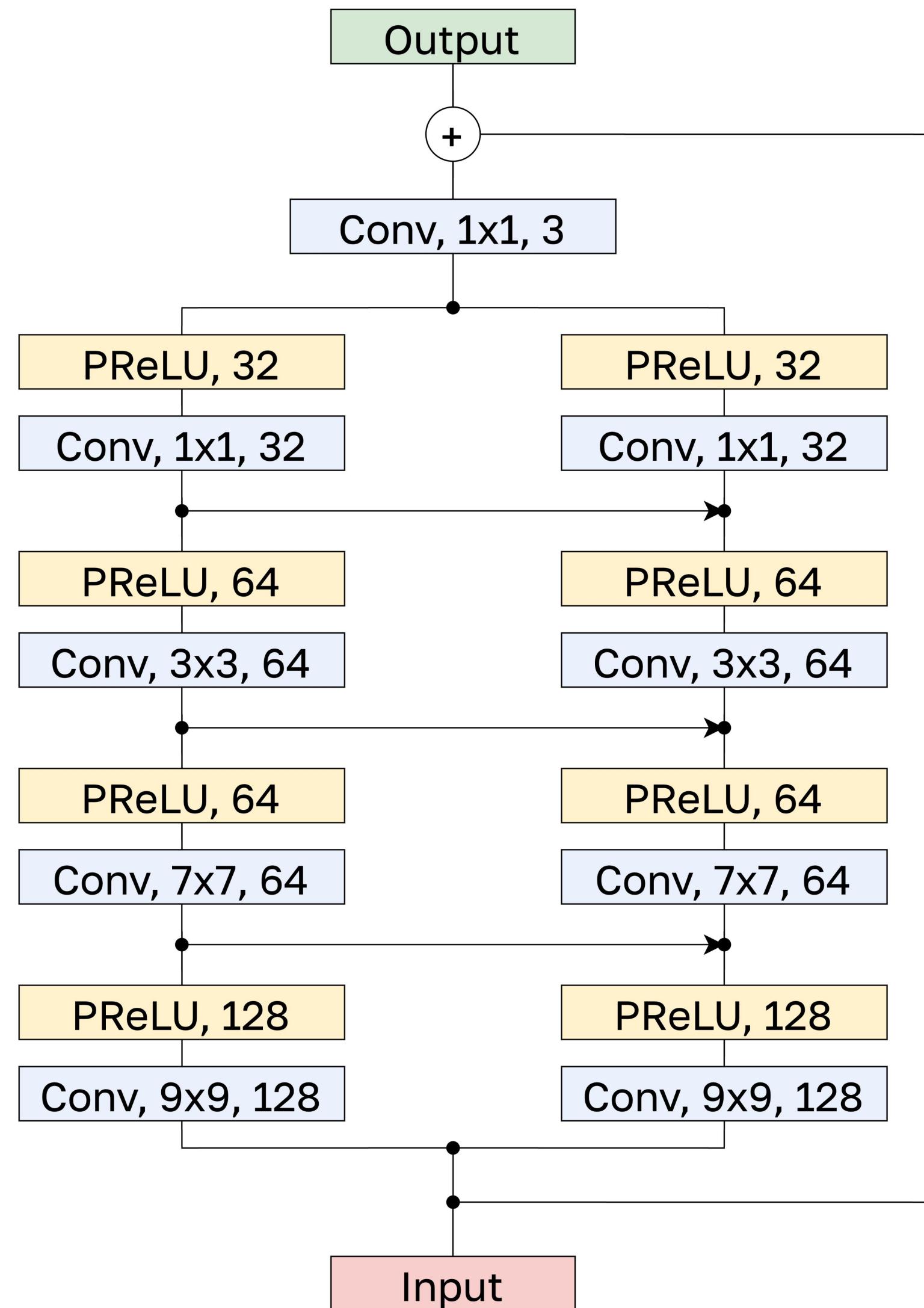


**Подготовка:**

1. Обрезка до размера 320×320;
2. Декодирование в .ррт с обнулением 16 псевдослучайных коэффициентов в каждом блоке.

# Архитектура и реализация нейросети

ИТМО



Каждый сверточный слой сохраняет размерность тензора.

Язык реализации: Python.

Фреймворк для машинного обучения:  
PyTorch.

**Функция потерь:**

`torch.nn.MSELoss()`

**Оптимизатор:**

`torch.optim.Adam(lr=1e-4)`

**Число эпох:** 100

**Машина:**

MacBook Pro 2021 г. (чип M1 Pro)

**Результаты:**

- $\min \text{MSE} = 1,405 \times 10^{-3}$
- $\overline{\Delta \text{YPSNR}} = 3,824$

# Визуальная оценка работы нейросети

ИТМО



# Результаты и тестирование

ИТМО

Проведено функциональное тестирование и выполнено сравнение степени сжатия для различных подходов к транскодированию JPEG:

Методы	Метрики		
	Среднее	Медиана	По всему набору данных
1. Предобученная модель	2,57	2,68	2,76
2. QE-CNN-P	3,33	3,38	3,38
3. Jpegtran	10,05	10,00	10,12
4. QE-CNN-P + Jpegtran	13,16	13,11	13,18
5. LLJPEG	16,14	15,45	15,59
6. QE-CNN-P + LLJPEG	15,44	15,08	15,14

- 
1. Реализовано декодирование JPEG с обнулением части коэффициентов ДКП;
  2. Обучена экспериментальная нейронная сеть, способная восстанавливать обнуленные коэффициенты;
  3. Реализовано дополнительное сжатие JPEG за счет кодирования с предсказанием для обнуленных коэффициентов;
  4. Транскодирование основанное на нейросетевом предсказании коэффициентов дало сжатие тестового набора данных на 3,38%;
  5. Предложенный метод эффективно комбинируется с Jpegtran: суммарное сжатие составило 13,18%, что отстает от LLJPEG всего на 2,41%.



---

**Спасибо за внимание!**

---