

# Восстановление правильного положения горизонта на изображении

Мельникова Алина Денисовна



Репозиторий с кодом



Telegram-бот с моделью

# Задача



«Завал» горизонта — частая проблема, приводящая к неестественному восприятию фото или неправильному чтению документов.

Было бы хорошо автоматически исправлять горизонт на фотографиях не требуя ручной обработки.

**Задача:** построить модель, определяющую по изображению угол его поворота относительно естественного положения.



Основная метрика: средняя абсолютная ошибка (МАЕ)

# Датасет



Рис. 1 примеры из датасета

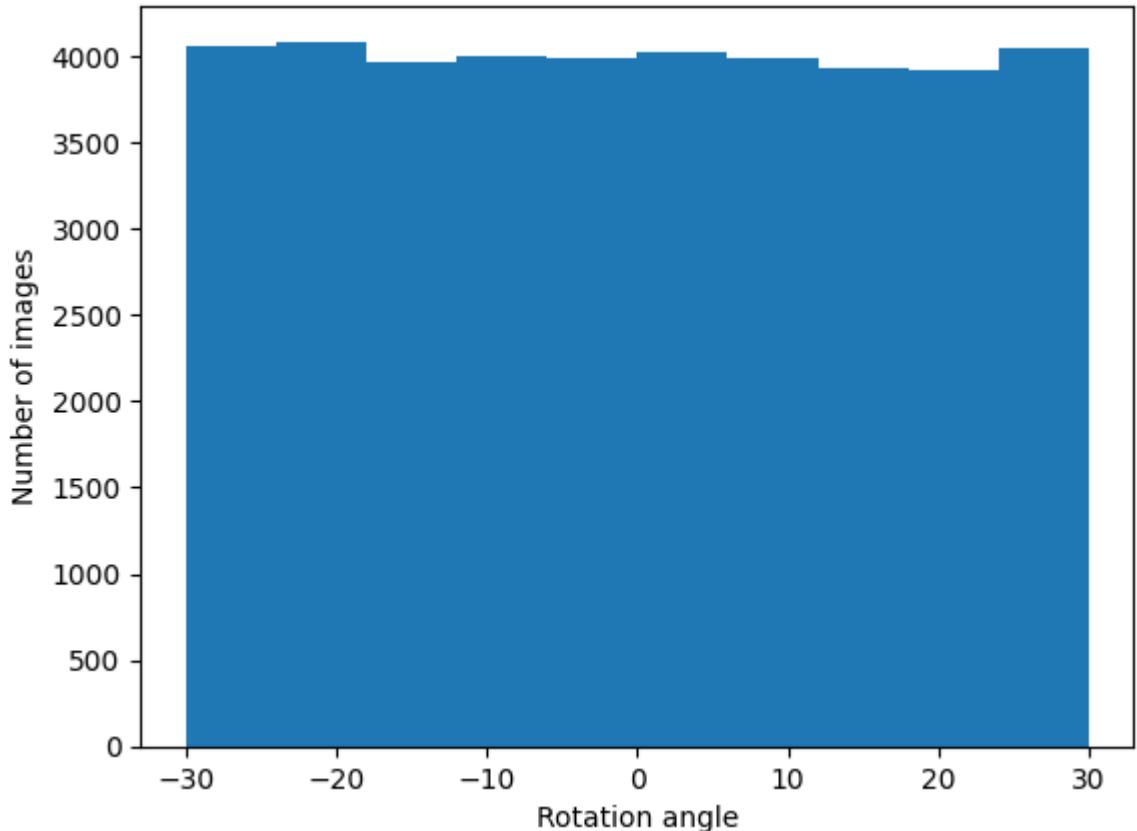


Рис. 2. Гистограмма распределения целевой переменной

# Датасет



Рис 3. Среднее изображение



Рис 4. Усреднённое стандартное отклонение по изображению

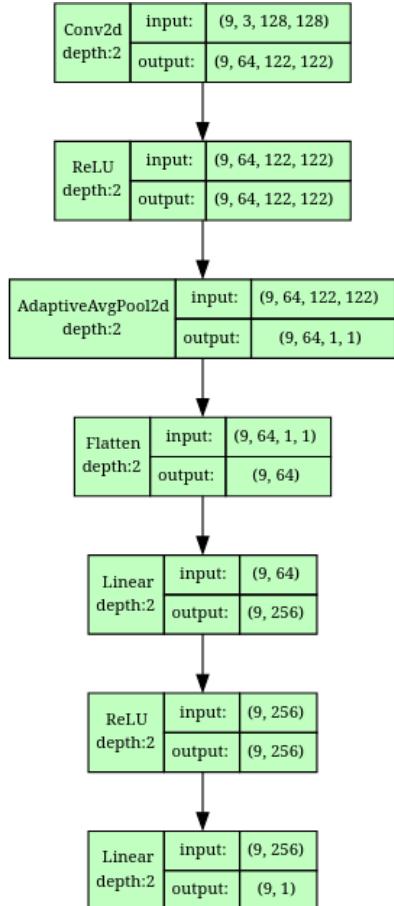
Выборка	Число примеров
Train	28000
Val	6000
Test	6000

Табл 1. Разбиение выборки

# Бейзлайн

```
Sequential(  
    Conv2d(3, 64, 7),  
    ReLU(),  
    AdaptiveAvgPool2d((1,  
1)),  
    Flatten(),  
    Linear(64, 256),  
    ReLU(),  
    Linear(256, 1),  
)
```

Схема 1. Архитектура модели



Параметр обучения	Значение
Оптимизатор	AdamW
Learning Rate	0,001
Weight Decay	0,01
Планировщик	Reduce LR on plateau
Patience	10
Coldown	0
Scale	0,1
КОЛ-ВО ЭПОХ	200

Табл. 2. Параметры обучения

# Бейзлайн

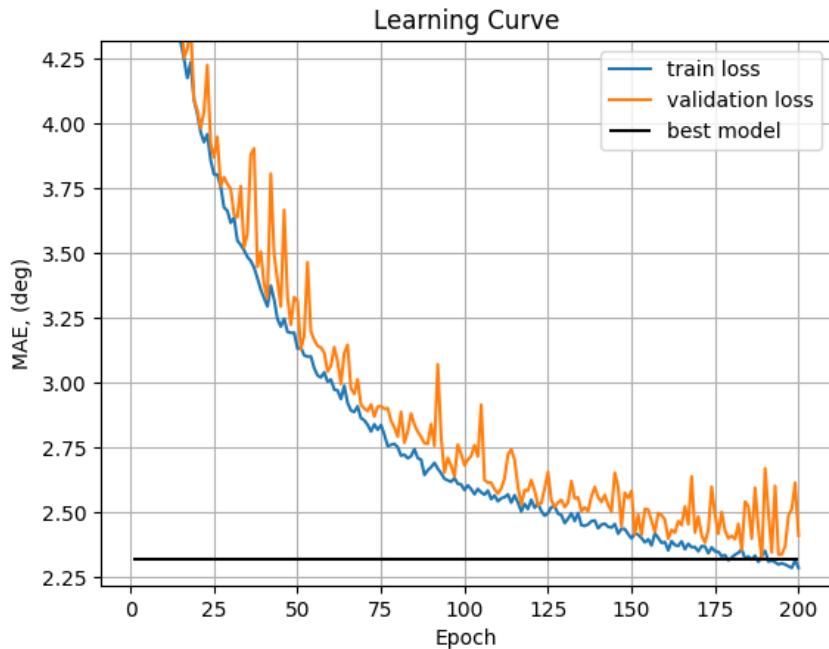
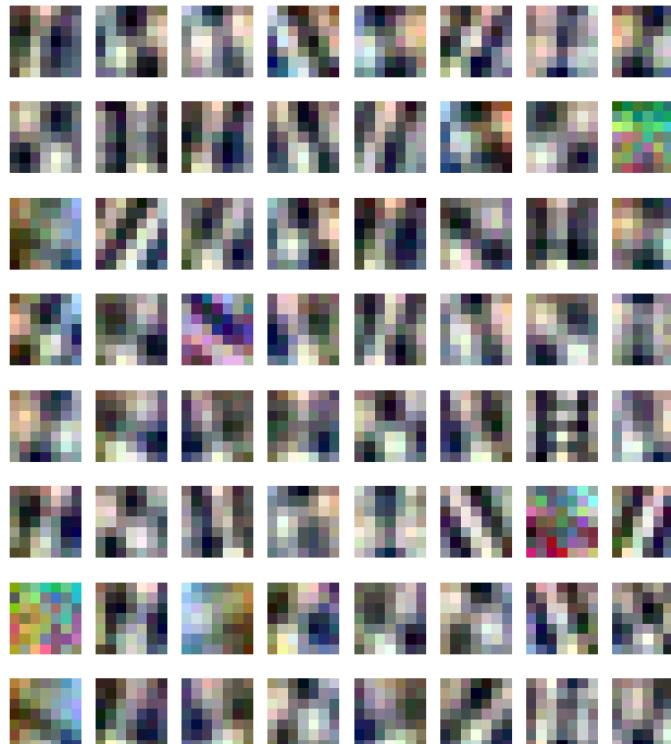


Рис 5. Кривая обучения



МАЕ на  
тестовой  
выборке

2,26

Рис 6. Визуализация фильтров

# Бейзлайн

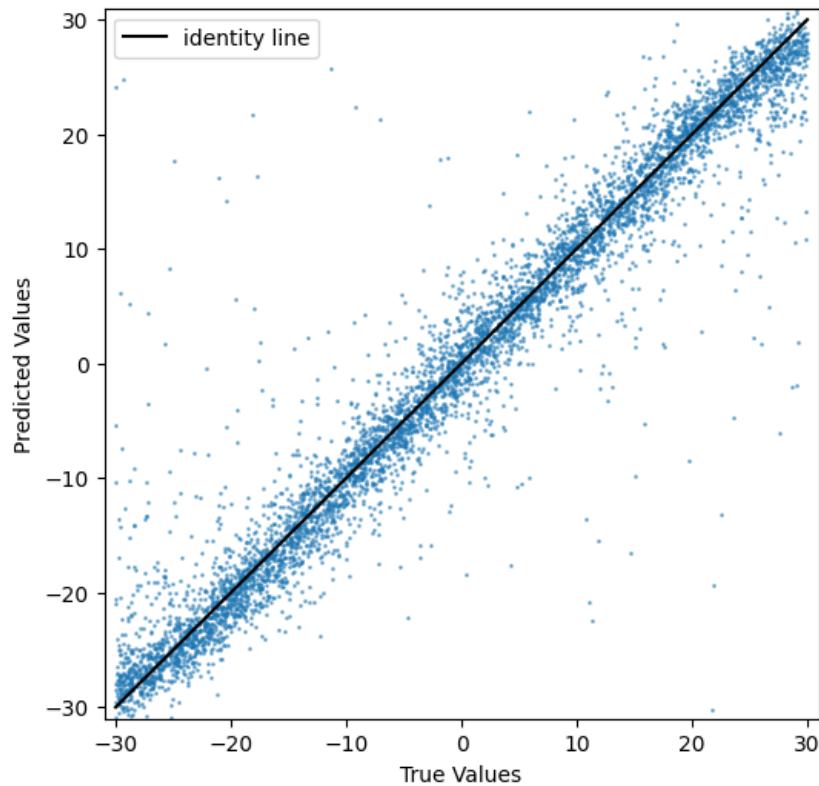


Рис 7. График Регрессии

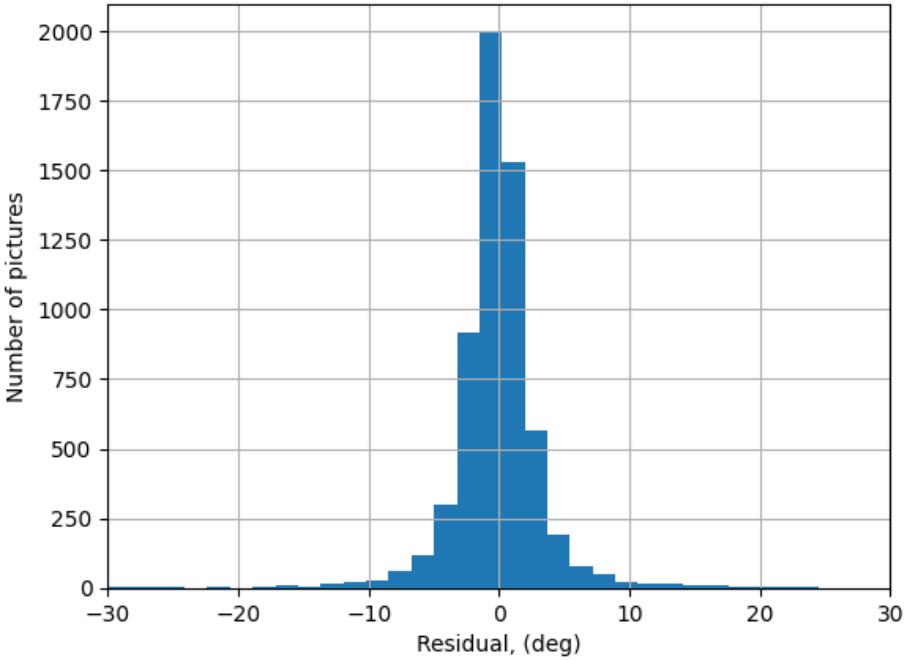


Рис 8. Гистограмма остатков

# Бейзлайн

True: -29.3 Pred: 24.8	True: -30.0 Pred: 24.1	True: 21.8 Pred: -30.2	True: -24.9 Pred: 17.6	True: 21.9 Pred: -19.4	True: -18.1 Pred: 21.6	True: -21.1 Pred: 16.2	True: -11.3 Pred: 25.7	True: 22.6 Pred: -13.2	True: -29.5 Pred: 6.2
---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	--------------------------



Rotated according to True



Rotated according to Pred



Рис 9. Визуализация выбросов

# Выбор модели

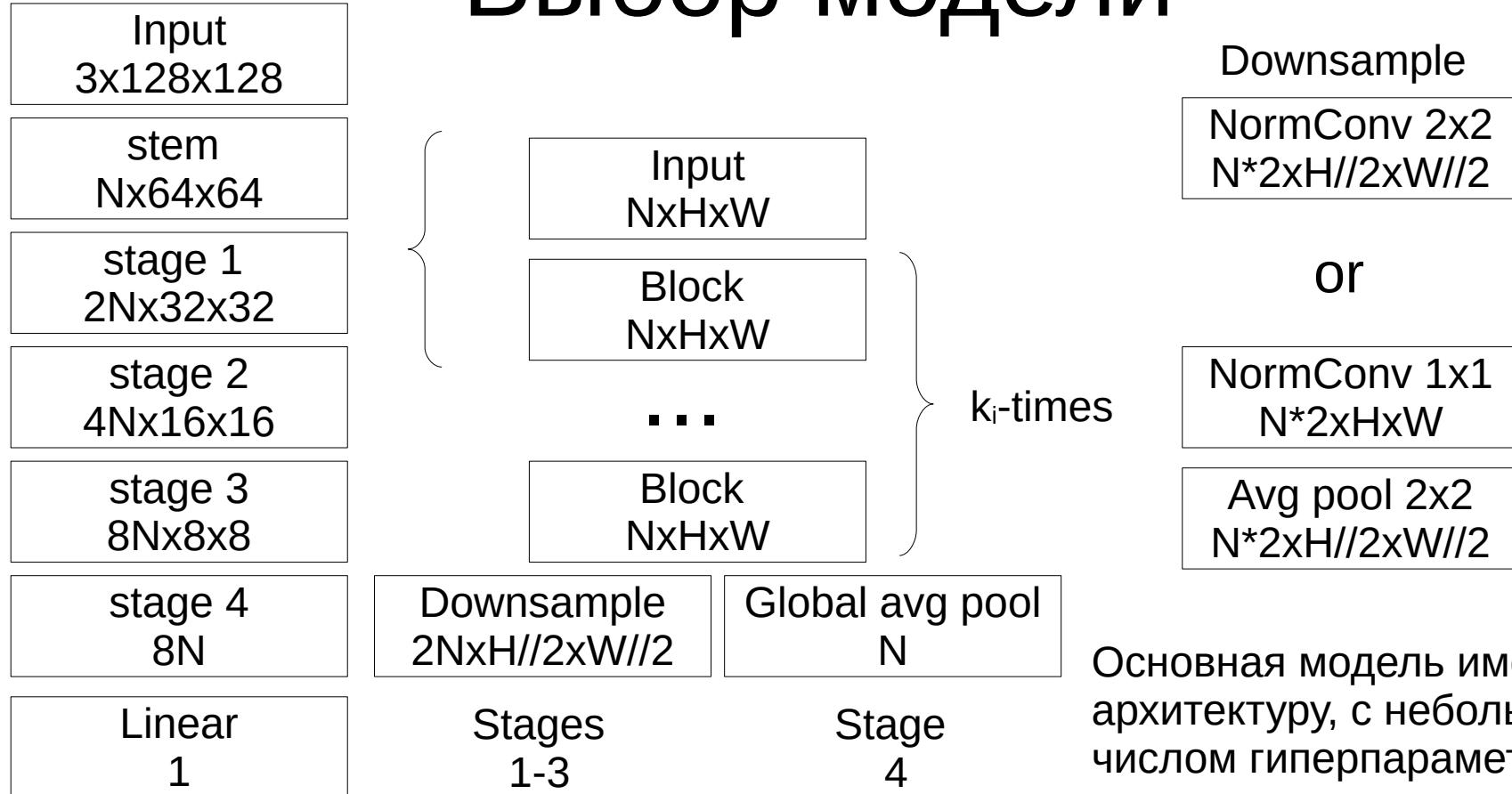


Схема 2. Глобальная архитектура сети, ступень сети, слой понижения размерности

# Выбор модели

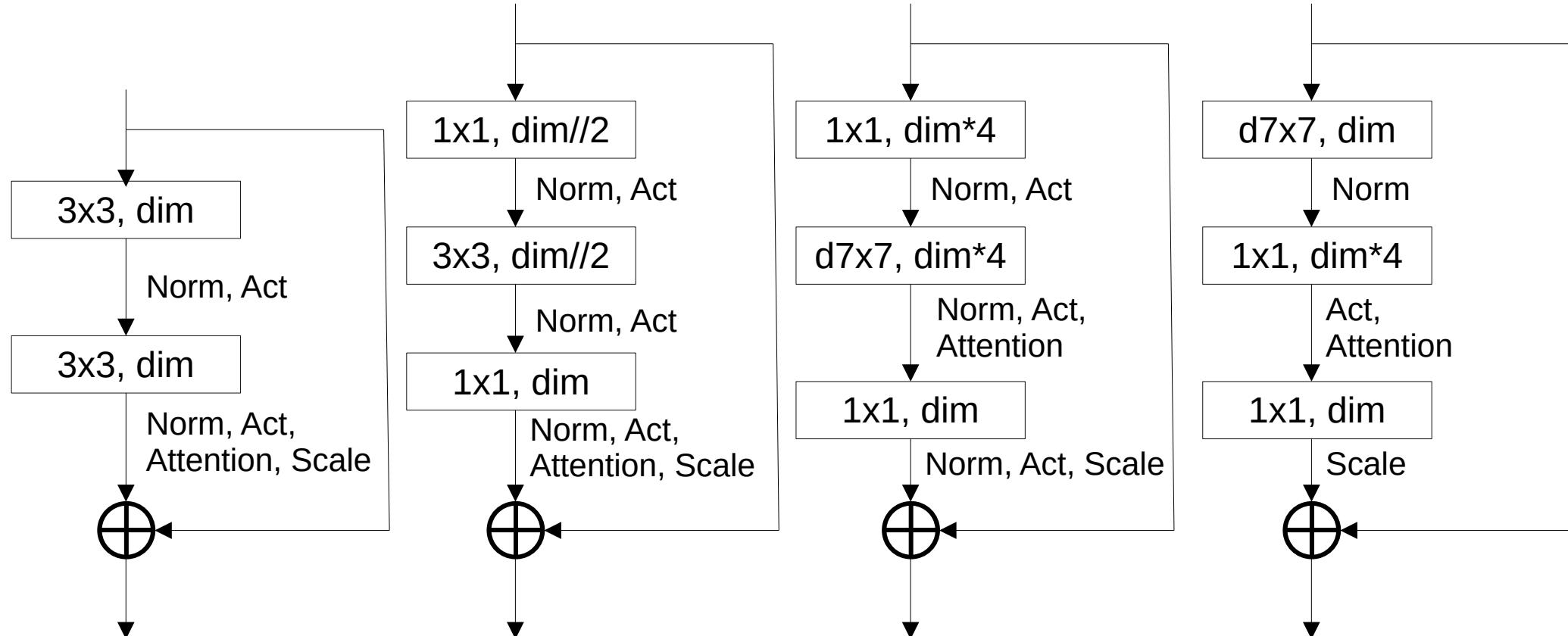


Схема 3. Визуализация составных блоков нейросети, слева-направо:  
ResNet, Bottleneck, Inverted bottleneck, ConvNeXT

# Выбор модели

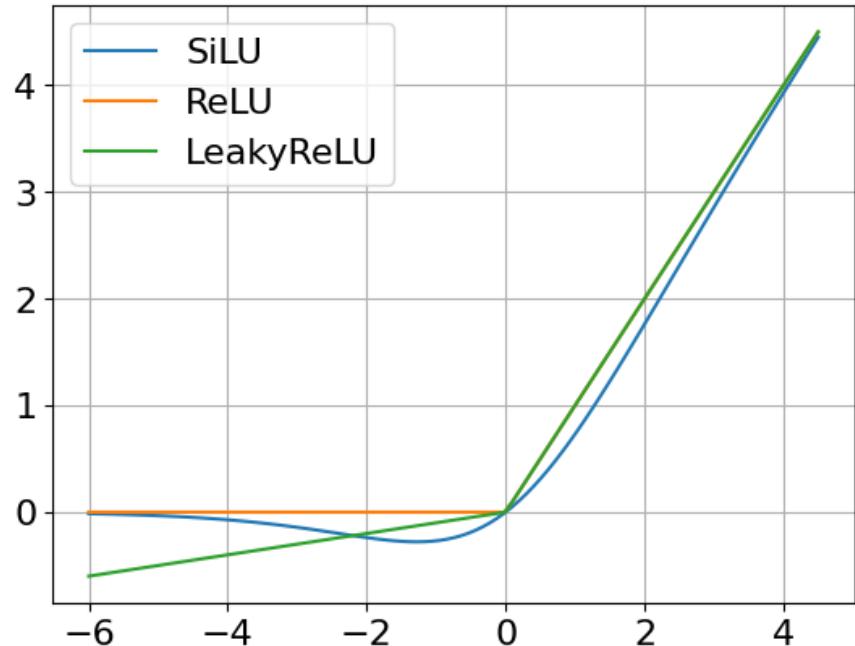


Рис 10. Визуализация функций активации

No Norm	-
Batch Norm	$(B, C, H, W)$ $y = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}} * \gamma + \beta$
Layer Norm	$(B, H, W, C)$ $y = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}} * \gamma + \beta$

Табл. 2. Слои нормировки

# Выбор модели

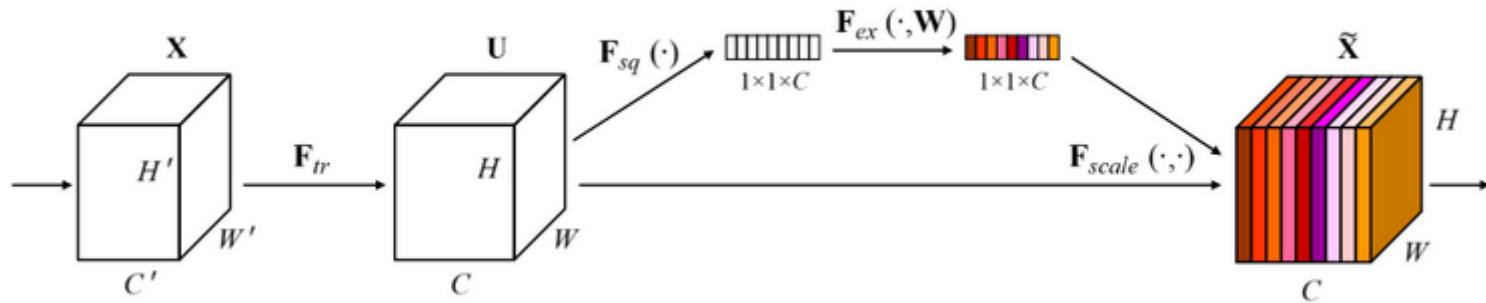


Рис 11. Схема работы squeeze-excitation

$$\mathcal{N}(\|X_i\|) := \|X_i\| / \sum \|X_i\|$$

$$X_i = X_i * \mathcal{N}(\mathcal{G}(X)_i)$$

Формула 1. Global Response  
Normalization

Attention-механизмы, позволяют модели «обратить внимание» модели на глобальный контекст

# Выбор модели

Параметр	Значения
N	16, 24, 32, 48, 64, 80, 96
Downsample	Conv, Pool
$k_1$	2,3,4
$k_2$	2,3,4,5,6
$k_3$	2,3,4,5,6,7,8
$k_4$	2,3,4,5,6
Block	ResNet, Bottleneck Inverted Bottleneck, ConvNeXT
Activation	ReLU, SiLU, LeakyReLU
Norm	No, BatchNorm, LayerNorm
Attention	No, Squeeze-Excitation, Global Response Normalization
Scale	No, Yes

Табл. 3. Список гиперпараметров модели

# Модель

Параметр	Значение
N	16
Downsample	Pool
$k_1$	4
$k_2$	2
$k_3$	8
$k_4$	3
Block	Bottleneck
Activation	LeakyReLU
Norm	BatchNorm
Attention	Squeeze-Excitation
Scale	Yes

*MAE на тестовой  
выборке*

**1,34**

Табл. 4. Список гиперпараметров найденной модели

# Модель

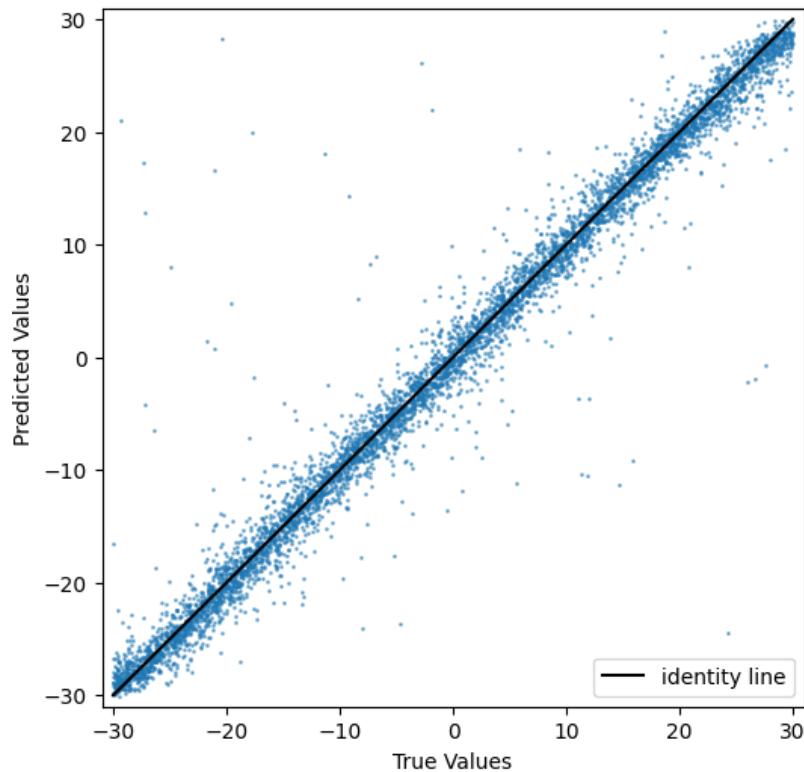


Рис 12. График Регрессии

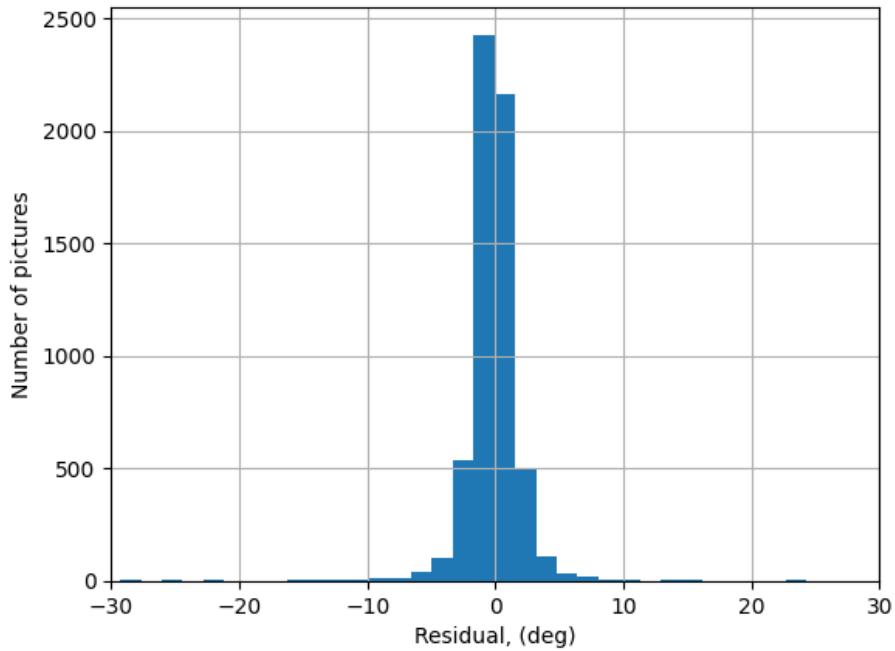


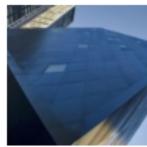
Рис 13. Гистограмма остатков

# Модель

True: -29.3  
Pred: 21.0



True: 24.3  
Pred: -24.5



True: -20.5  
Pred: 28.3



True: -27.3  
Pred: 17.2



True: -27.1  
Pred: 12.8



True: -21.1  
Pred: 16.6



True: -17.7  
Pred: 19.9



True: -24.9  
Pred: 8.0



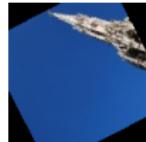
True: -11.3  
Pred: 18.0



True: -2.8  
Pred: 26.2



Rotated according to True



Rotated according to Pred

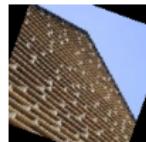
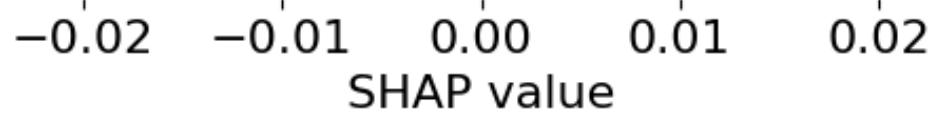
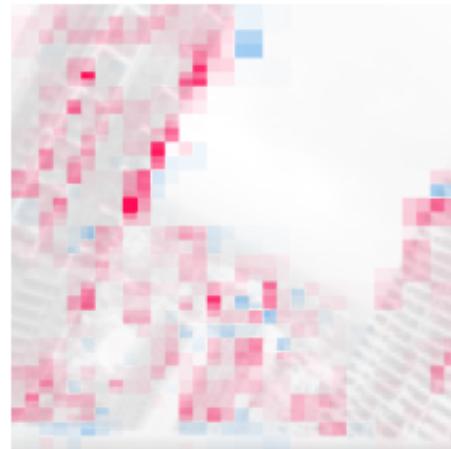
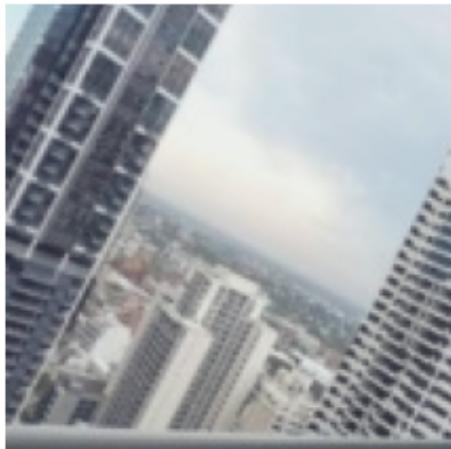


Рис 14. Визуализация выбросов

# Модель



Predicted: 28.14  
True: 27.25



Predicted: -18.50  
True: -15.67

Рис 15. Оценка предсказаний модели на примерах из тестовой выборки

# Выводы

1. Свёрточные нейронные сети хорошо справляются с задачей определения наклона изображения — полученная модель в абсолютном большинстве случаев даёт погрешность меньше 2 градусов
2. Для небольшого количества данных оптимальной оказалась узкая архитектура с bottleneck блоками хорошо справляющаяся с переобучением.
3. Модель в первую очередь обращает внимание на длинные контрастные прямые

Спасибо за  
внимание!