

Аугментация данных

Корпусова Софья
Афанасьев Андрей



Что такое **аугментация** данных?

- Процесс искусственного увеличения объема и разнообразия данных
- Создание измененных версий существующих данных
- Важность в машинном обучении при ограниченных или несбалансированных данных

Цели и недостатки аугментации

01

**Увеличение
объема данных**

02

**Повышение
разнообразия
данных**

03

**Снижение риска
переобучения**

01

**Увеличение
времени
обучения**

02

**Риск создания
неестественных
данных**

03

**Необходимость
тщательной
настройки**

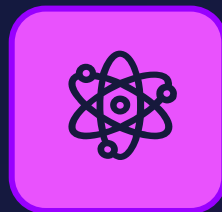
Методы аугментирования



Аугментация изображений с Keras



**Легкая интеграция в
процесс обучения**



**Множество
трансформаций**



**Аугментация в
реальном времени**

Пример

```
from keras.preprocessing.image import ImageDataGenerator

datagen = ImageDataGenerator(

    rotation_range=40,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'

)
```


Заключение

**Аугментация – важный
инструмент**

**Улучшает качество
моделей**

**Требует внимательного
подхода**

```
train_generator = datagen.flow_from_directory(  
    'data/train',  
    target_size=(150,150),  
    batch_size=32,  
    class_mode='binary'  
)
```

```
model.fit_generator(  
    train_generator,  
    steps_per_epoch=2000,  
    epochs=50  
)
```