



Московский государственный технический университет
имени Н.Э. Баумана (национальный исследовательский университет)

Семантическая сегментация изображений рыб

Студент группы ФН12-71Б: А.А. Сысоев
Научный руководитель: Д.А. Фетисов

Москва, 2025 г.

Постановка задачи

Цель работы:

разработка и обучение модели нейронной сети, способной решать задачу семантической сегментации изображений рыб

Задачи:

- 1 Исследование архитектуры нейронной сети U-net.
- 2 Подготовка и предварительная обработка набора изображений.
- 3 Обучение модели на выбранном датасете.
- 4 Оценка качества работы модели с использованием различных метрик.
- 5 Анализ полученных результатов.

Семантическая сегментация – попиксельная разметка изображения, где каждая метка соответствует определенному классу объекта. Для формализации введем следующие обозначения:

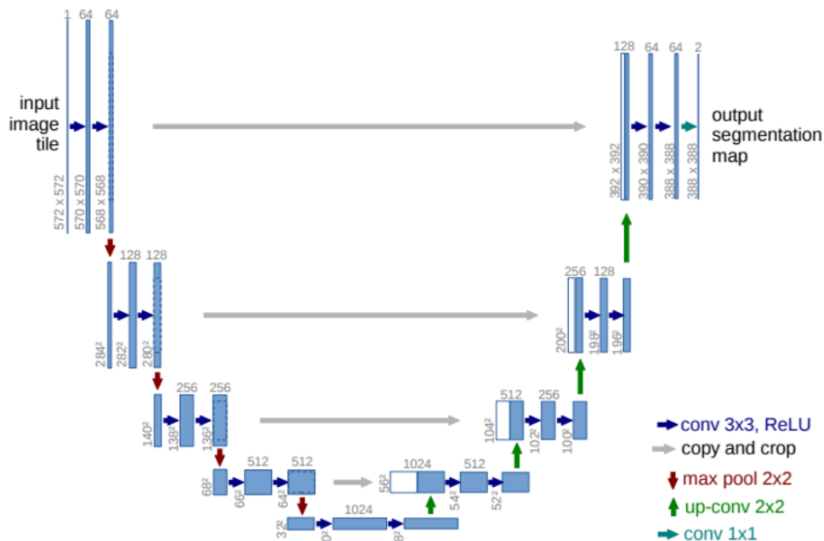
- $X = \{x_i\}_{i=1}^N$ — множество пикселей входного изображения;
- $Y = \{y_i\}_{i=1}^N$ — множество истинных меток пикселей;
- $\hat{Y} = \{\hat{y}_i\}_{i=1}^N$ — множество предсказанных меток;
- $y_i, \hat{y}_i \in \{0, 1\}$, $i = 1, \dots, N$.

Задача семантической сегментации заключается в нахождении такого отображения

$$f_{\theta} : X \rightarrow \hat{Y}, \quad (1)$$

где f_{θ} — нейронная сеть с параметрами θ , которая для каждого пикселя входного изображения предсказывает его принадлежность к соответствующему классу. Обучение модели сводится к подбору параметров θ , минимизирующих функцию потерь между предсказанными метками \hat{Y} и истинной разметкой Y на обучающей выборке.

Архитектура сверточной нейронной сети U-net



- Датасет состоит из 9000 пар изображений рыб и бинарных масок (9 классов).
- Исходное разрешение изображений: 590×445 пикселей.
- Приведение всех изображений и масок к фиксированному размеру 224×224 .
- Преобразование изображений в тензоры и нормализация по каналам RGB (статистика ImageNet).
- Масштабирование масок с использованием интерполяции ближайшего соседа.
- Деление на обучающую, валидационную и тестовую выборки.

- Accuracy — доля правильно классифицированных пикселей.
- Основана на элементах матрицы ошибок:
 - TP — верно предсказанный объект;
 - TN — верно предсказанный фон;
 - FP — ложный объект;
 - FN — пропущенный объект.

- Формула:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Плюсы:** простая и интуитивная метрика.
- **Минусы:** может быть завышена при дисбалансе классов (преобладание фона).

- Intersection over Union (IoU) — основная метрика качества в задачах семантической сегментации.
- Геометрическое определение:

$$\text{IoU} = \frac{A \cap B}{A \cup B}$$

где:

- A — множество пикселей истинной маски;
- B — множество пикселей, предсказанных моделью.
- Определение через матрицу ошибок:

$$\text{IoU} = \frac{TP}{TP + FP + FN}$$

- Метрика не учитывает истинно отрицательные пиксели (TN), что делает её устойчивой к дисбалансу классов.

Функция потерь Dice Loss

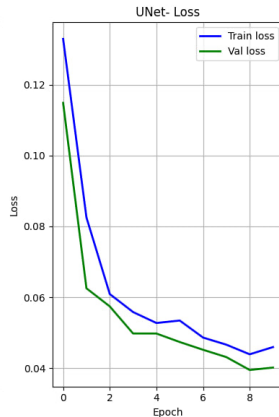
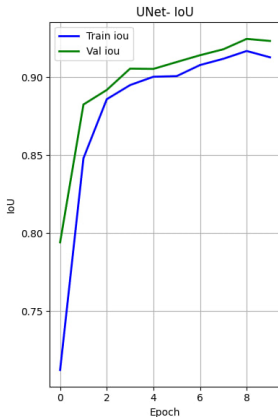
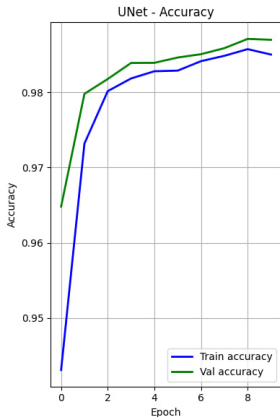
- Dice Loss — функция потерь, используемая в задачах семантической сегментации.
- Основана на коэффициенте Dice, который оценивает степень совпадения предсказанной и истинной масок.
- Предсказания модели преобразуются в вероятности с помощью сигмоидной функции.
- Элементы, используемые в формуле:
 - $TP = \sum_i p_i y_i$ — корректно предсказанные пиксели объекта;
 - $FP = \sum_i p_i (1 - y_i)$ — пиксели, ошибочно отнесённые к объекту;
 - $FN = \sum_i (1 - p_i) y_i$ — пиксели объекта, не распознанные моделью.
- Формула коэффициента Dice:

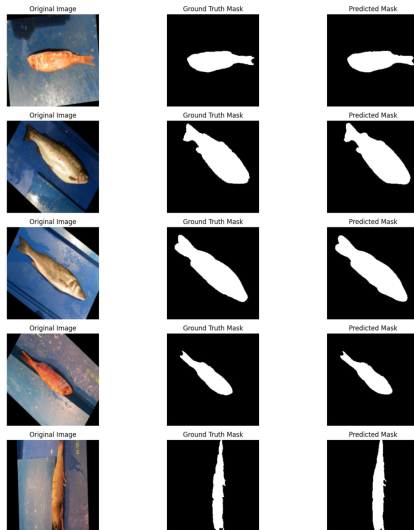
$$\text{Dice} = \frac{2TP}{2TP + FP + FN}$$

- Функция потерь:

$$\mathcal{L}_{Dice} = 1 - \text{Dice}$$

Обучение модели





- Достигнутая точность по пикселям (Accuracy): $\sim 98.6\%$
- Коэффициент перекрытия масок (IoU): $\sim 91.6\%$
- Графики потерь и метрик показывают стабильное обучение без признаков переобучения
- Модель успешно сегментирует изображения рыб и корректно выделяет границы объектов
- Возможности для улучшения:
 - добавить Dropout и Batch Normalization
 - увеличить размер обучающей выборки

- Удалось разработать и обучить модель для задачи семантической сегментации изображений рыб
- Подтверждена эффективность сверточных нейронных сетей (U-Net) в задачах сегментации объектов на изображениях
- Предложены направления дальнейшей работы

Спасибо за внимание!