

РЕШЕНИЕ ЗАДАЧИ СЕГМЕНТАЦИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

Павловский Д.А., Фролов О.Ю.

*Национальный исследовательский Томский государственный университет,
Томск*

Ключевые слова: сегментация, свёрточная нейронная сеть, обучение с учителем.

Аннотация. В статье дан обзор проблемы понимания композиции сцены компьютером и рассмотрены подходы к решению этой проблемы. Отдельно разобраны задача сегментации и её подразделения. Подробно рассмотрен способ решения задачи сегментации с использованием свёрточных нейронных сетей. Детально разобрана архитектура сети U-net – как один из распространённых методов решения задачи сегментации. Приведены результаты обучения нейронной сети, построенной на основе архитектуры U-net.

SOLVING SEGMENTATION TASK USING DEEP LEARNING METHODS

Pavlovsky D.A., Frolov O. Yu.

National Research Tomsk State University, Tomsk

Keywords: segmentation, convolutional neural network, supervised learning.

Abstract. This article offers an overview of scene understanding problem and of various approaches to solving this problem. Segmentation task and its different types were addressed separately. Use of convolutional neural networks to solve segmentation problem was explored in depth. U-net architecture – a common solution method for segmentation task, was described thoroughly. Training results of U-net based neural network were presented in the article.

Понимание композиции сцены – задача в области машинного зрения, решения которой ищутся множество лет. Подобная задача выполняется человеком без усилий, однако создание алгоритма с подобным уровнем восприятия – комплексная задача.

К решению этой задачи подходят по-разному. Можно получить неплохое понимание сцены определением объектов, находящихся на изображении. Такой подход называется классификацией изображений. Однако в некоторых задачах только информации о наличии объекта недостаточно, необходимо также определить положение объекта на изображении.

Обнаружение объектов предполагает наложение ограничивающей рамки на каждый объект известного класса, найденный на изображении. Обнаружение объектов позволяет получить информацию так же о расположении объектов, но из-за своего способа выделения объекта имеет ограниченную точность.

Сегментация стремится к более точному определению расположения объектов и заключается в попиксельном определении границ объектов, создавая маску изображения. Иными словами задача сегментации состоит в классификации каждого пикселя изображения, как принадлежащего объекту известного класса. Результатом сегментации будет являться набор сегментов, покрывающих всё исходное изображение.

Сегментация в свою очередь подразделяется на два вида: семантическую и объектную. Семантическая сегментация предполагает выделение маской расположения известных классов объектов, то есть все объекты одного класса будут выделены одним сегментом. Объектная сегментация выделяет отдельной маской каждый объект каждого класса, то есть каждому объекту на изображении соответствует уникальный сегмент [1-2].

Практическое применение сегментации позволяет компьютеру извлекать информацию из реальных ситуаций и использовать её для решения множества задач, которые включают в себя:

- определение окружения самоуправляемого транспорта, требующее попиксельной точности;
- создание систем ориентации в пространстве, пригодных для использования в робототехнике;
- анализ медицинских изображений путем выделения интересующих клеток, тканей или органов;
- использование сегментации для упрощения работы в фото- и видеоредакторах.

Существуют разные подходы к решению задачи сегментации, однако использование свёрточных нейронных сетей показывает себя наиболее эффективным и в настоящее время активнее всего развивается.

Целью данной работы стало написание свёрточной нейронной сети, способной решать задачу сегментации. В качестве основы нейросети бралась архитектура U-net, представленная на рисунке 1.

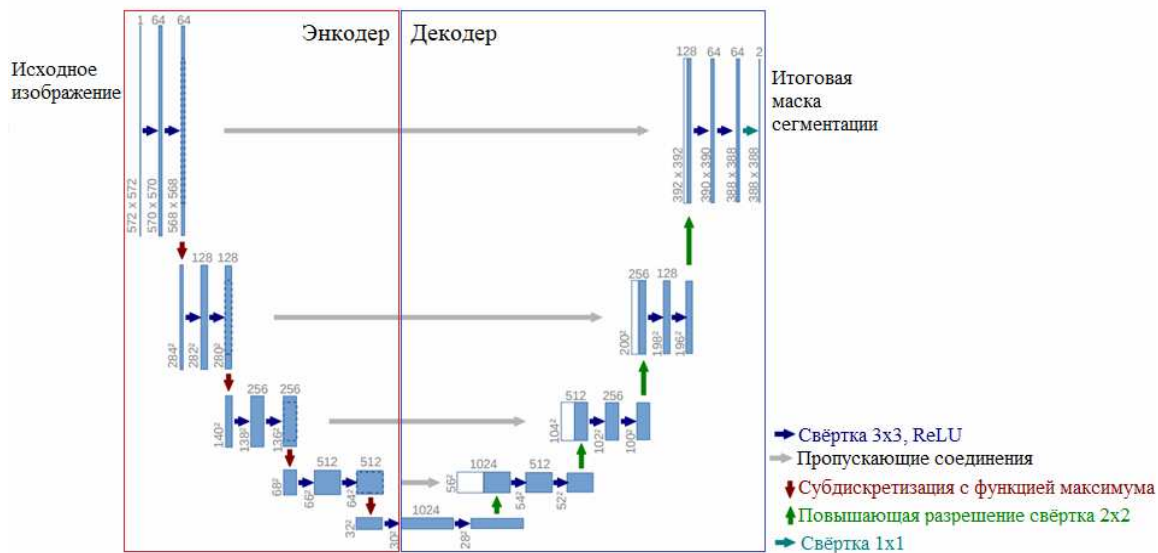


Рис. 1. Схема архитектуры сети U-net

Сеть U-net состоит из двух частей: энкодера и декодера.

Энкодер представляет из себя обычную свёрточную сеть, которая создаёт карту характеристик – многоканальное изображение, используемое для передачи информации внутри сети. Операция субдискретизации постепенно уменьшает количество информации исходного изображения. Потеря информации необходима для того, чтобы сеть извлекала из изображения общие признаки, а не просто запоминала входные данные.

Декодер – это симметричная энкодеру свёрточная сеть, которая с помощью повышающей разрешение свёртки увеличивает карту характеристик и на её основе создаёт сегментированное изображение.

Ключевой особенностью архитектуры U-net является использование пропускающих соединений [3]. После каждого повышения разрешения с картой характеристик, полученной повышением разрешения, сцепляется карта характеристик, скопированная с соответствующего уровня энкодера. Таким образом сеть использует как обобщённую абстрактную информацию, так и информацию о локальных характеристиках исходного изображения, которая была потеряна при проходе через энкодер, что позволяет повысить точность определения границ сегментов.

Работа велась в среде google colab на языке программирования python. Тренировка сети проходила в формате обучения с учителем, то есть имелся размеченный набор данных, где каждому изображению соответствовала собственная маска сегментации. В качестве набора данных использовался The Oxford-III Pet Dataset [4], содержащий 37 классов объектов примерно по 200 изображений для каждого класса. Общее число изображений в наборе данных – 7349. График роста точности работы сети с числом эпох представлен на рисунке 2.

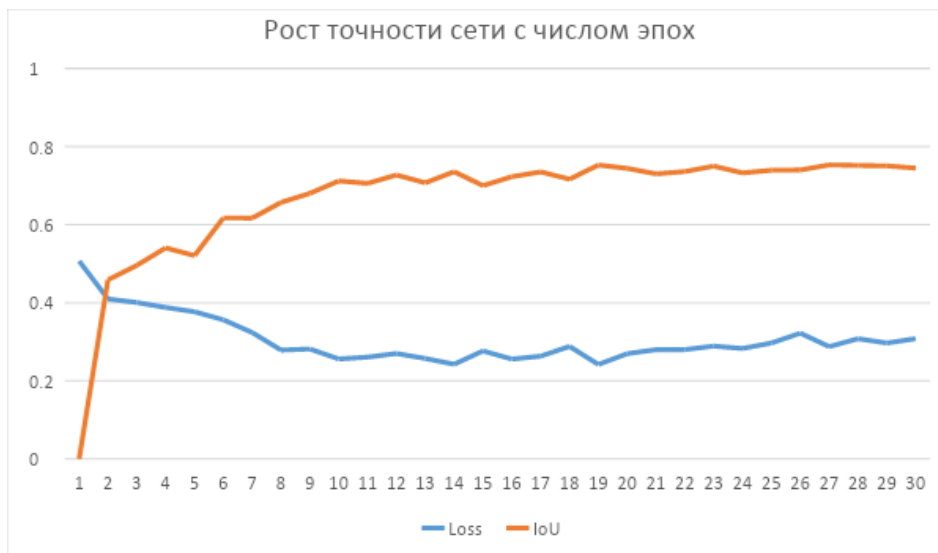


Рис. 2. Оценки точности работы сети с числом эпох: синяя кривая – значение функции потерь, оранжевая кривая – значение метрики точности

В качестве функции потерь использовалась перекрёстная энтропия (1), а в качестве метрики точности – отношение пересечения полученной и действительной масок к их объединению (2):

$$\text{Loss}(y_p, y_t) = - \sum_x y_t(x) \log y_p(x), \quad (1)$$

$$\text{IoU} = (Y_t \cap Y_p) / (Y_t \cup Y_p), \quad (2)$$

где y_t – действительное значение пикселя, y_p – полученное значение пикселя, x – пиксель изображения, Y_t – действительная маска сегментации, Y_p – полученная маска сегментации.

На рисунке 3 показаны примеры сегментации известных классов объектов обученной сетью.

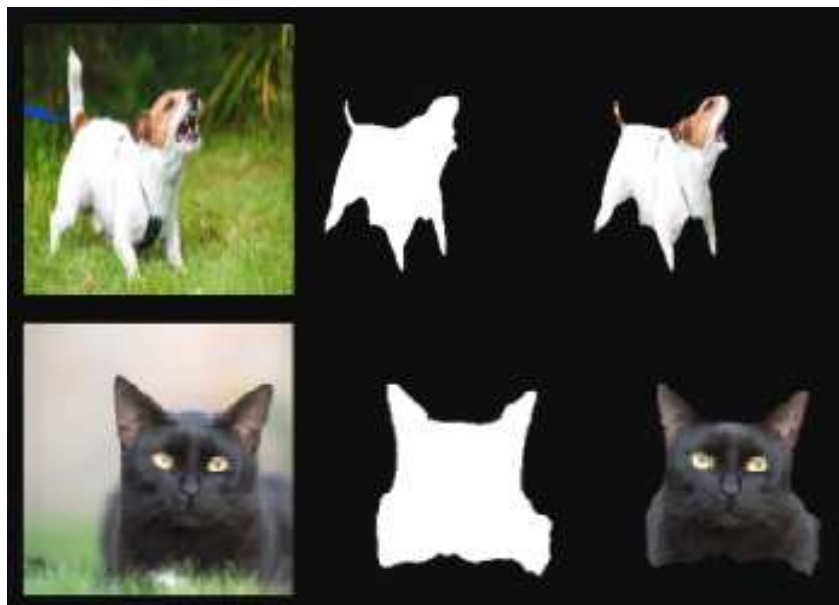


Рис. 3. Примеры работы сети

Реализованная сеть показала приемлемую точность, однако показатели могут быть улучшены дальнейшей настройкой гиперпараметров сети. Результат обучения сети сильно зависит от качества выбранного набора данных.

Список литературы

1. Conditional Random Fields Meet Deep Neural Networks for Semantic Segmentation: Combining Probabilistic Graphical Models with Deep Learning for Structured Prediction / A. Arnab, S. Zheng, S. Jayasumana [et al.] // IEEE Signal Processing Magazine. – 2018. – Vol. 35, is. 1. – P.37-52.
2. Liu X. Recent progress in semantic image segmentation / X. Liu, Z. Deng, Y. Yang // Artificial Intelligence Review. – 2019. – Vol. 52. – P. 1089-1106.
3. U-Net: Convolutional Networks for Biomedical Image Segmentation [Электронный ресурс]. – URL.: <https://arxiv.org/abs/1505.04597>
4. The Oxford-IIIT Pet Dataset [Электронный ресурс]. – URL.: <https://www.robots.ox.ac.uk/~vgg/data/pets>

Сведения об авторах:

Павловский Дмитрий Андреевич – студент, НИ ТГУ, Томск;

Фролов Олег Юрьевич – к.ф.-м.н., доцент, НИ ТГУ, Томск.