

## Семантическая сегментация в машинном обучении

Демин И.С.<sup>1</sup>, Рыбкин С.В.<sup>1</sup>

*<sup>1</sup>Калужский филиал ФГБОУ ВО «Московский государственный технический университет им. Н.Э. Баумана (национальный исследовательский университет)», Калуга, e-mail: fn1-kf@mail.ru;*

---

В данной статье рассматривается задача семантической сегментации в машинном обучении и ее применение в анализе данных спутниковых снимков земли. Задача семантической сегментации является высокоуровневой задачей обработки изображений, относящейся к группе задач машинного распознавания. Она является более сложной, чем задача классификации изображений и поиска объектов, что обусловлено не только необходимостью определения классов объектов, но и выявления их структуры, правильного выделения частей объектов на изображении. Целью семантической сегментации является точная классификация и сегментация каждого объекта, представленного на изображении. Семантическая классификация состоит из трех основных шагов: выделение всех классов, сегментация и создание метки для каждого пикселя. Эти метки указывают на принадлежность определенного пикселя к сущности класса. В широком смысле архитектура семантической сегментации может рассматриваться как кодирующая сеть, за которой следует декодирующая сеть. Существует три основных метода семантической сегментации: областная семантическая сегментация, семантическая сегментация на основе полностью сверточной сети, семантическая сегментация со слабым контролем. В рамках статьи рассматривается реализация архитектур FCN-8 с использованием библиотеки машинного обучения tensorflow.

---

Ключевые слова: семантическая сегментация, R-CNN, FCN, FCN-8, машинное обучение

## Semantic segmentation in machine learning

Demin I.S.<sup>1</sup>, Rybkin S.V.<sup>1</sup>

*<sup>1</sup>Moscow State Technical University n.a. Bauman (National Research University), Kaluga Branch, Kaluga, e-mail: fn1-kf@mail.ru;*

---

In this article are discusses the problem of semantic segmentation in machine learning and its application in the analysis of satellite images of the earth. The task of semantic segmentation is a high-level image processing task belonging to the group of machine recognition tasks. It is more complex than the task of classifying images and searching for objects, which is caused not only by the need to determine the classes of objects, but also to identify their structure and correctly select parts of objects in the image. The purpose of semantic segmentation is the exact classification and segmentation of each object represented in the image. The semantic classification consists of three main steps: the selection of all classes, segmentation and the creation of labels for each pixel. These labels indicate that a particular pixel belongs to a class entity. In a broad sense, semantic segmentation architecture can be considered as a coding network, followed by a decoding network. There are three main methods of semantic segmentation: regional semantic segmentation, semantic segmentation on the basis of a fully convolutional network, semantic segmentation with weak control. The article discusses the implementation of FCN-8 architectures using the tensorflow machine learning library.

---

Keywords: semantic segmentation, R-CNN, FCN, FCN-8, machine learning

### Введение

В настоящее время семантическая сегментация является одной из ключевых задач в области компьютерного зрения. Обработывая все объекты на изображении, семантическая сегментация позволяет программе полностью интерпретировать обрабатываемую сцену. Важность программной интерпретации изображения подчеркивается тем фактом, что все большее число приложений переводит компьютерные данные в графический формат, который понятен человеку. К таким приложениям относятся: системы управления

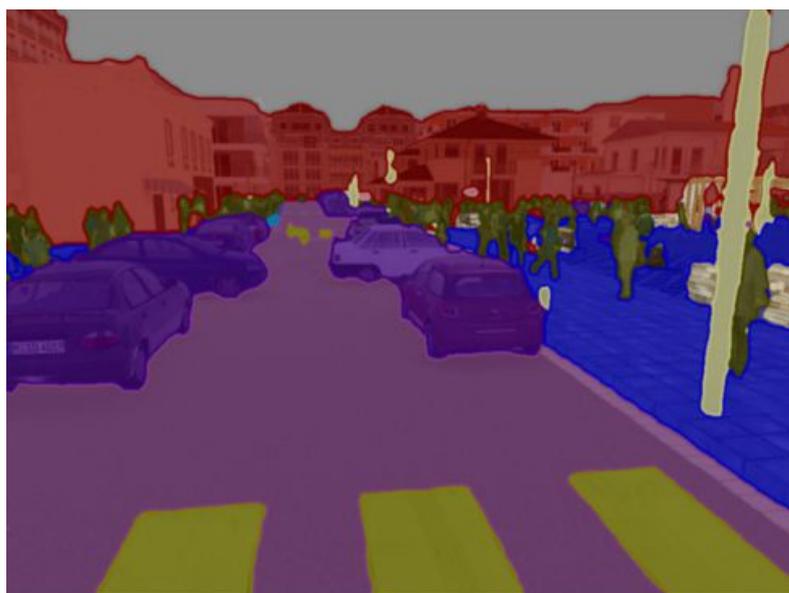
транспортным средством для автономного вождения, продвинутые способы взаимодействия человека с компьютером, виртуальная реальность и т.д. Из-за роста популярности машинного обучения в последние годы многие задачи семантической сегментации решаются путем использования архитектур нейронных сетей, чаще всего сверточных, которые превосходят другие подходы с большим отрывом, с точки зрения точности и эффективности. [1][2]

### **Обзор задачи семантической сегментации**

Целью семантической сегментации является точная классификация и сегментация каждого объекта, представленного на изображении. Семантическая классификация состоит из трех основных шагов: [3]

1. Первоначально происходит выделение всех классов, присутствующих на обрабатываемом изображении.
2. Следующим шагом является сегментация / обнаружение, которые предоставляют не только сущности классов, но и информацию относительно пространственного расположения этих классов.
3. На последнем шаге достигается точный вывод. Создаются метки для каждого пикселя. Эти метки указывают на принадлежность определенного пикселя к сущности класса. [4][5]

Результат задачи семантической сегментации, интерпретированный в изображение, понятное человеку, показан на рис. 1.



**Рис. 1.** Результат задачи семантической сегментации.

Далее представлены примеры некоторых стандартных архитектур глубоких нейронных сетей, которые внесли значительный вклад в область компьютерного зрения, поскольку они часто используются в качестве основы систем семантической сегментации:

1. AlexNet: архитектура CNN созданная в Торонто, победившая в конкурсе ImageNet 2012 года с точностью 84,6%. Она состоит из пяти слоев: 3 сверточных слоев с использованием активационной функции ReLU для нелинейности, одного слоя субдискретизации и одного слоя сброса.
2. VGG-16: эта модель нейронной сети была разработана в Оксфордском университете и победила в конкурсе ImageNet 2013 года с точностью 92,7%. Она использует стек слоев свертки с маленькими рецептивными полями в первых слоях вместо нескольких слоев с большими рецептивными полями.
3. GoogLeNet. Эта сеть от компании Google стала победителем конкурса ImageNet 2014 года с точностью 93,3%. Она состоит из 22 слоев и нового строительного блока нейронных сетей, называемого начальным модулем. Начальный модуль состоит из: слоя соединения сетей, слоя субдискретизации, уровня свертки большого размера и уровня свертки малого размера.
4. ResNet. Эта модель от компании Microsoft, которая выиграла конкурс ImageNet 2016 года с точностью 96,4%. Она хорошо известна благодаря своей глубине (152 слоя) и введению остаточных блоков. Остаточные блоки решают проблему обучения действительно глубокой архитектуры путем введения соединений с пропуском идентификаторов, чтобы слои могли копировать свои входные данные на следующий уровень.

### **Подходы семантической сегментации**

В широком смысле архитектура семантической сегментации может рассматриваться как кодирующая сеть, за которой следует декодирующая сеть:

1. Кодирующая сеть обычно представляет собой предварительно обученную классификационную нейронную сеть, такую как VGG / ResNet, за которой следует декодирующая сеть.
2. Задача декодирующей сети состоит в том, чтобы семантически проецировать распознаваемые признаки (более низкое разрешение), изученные кодирующей сетью, на пространство пикселей (более высокое разрешение), чтобы получить плотную классификацию. [6]-[8]

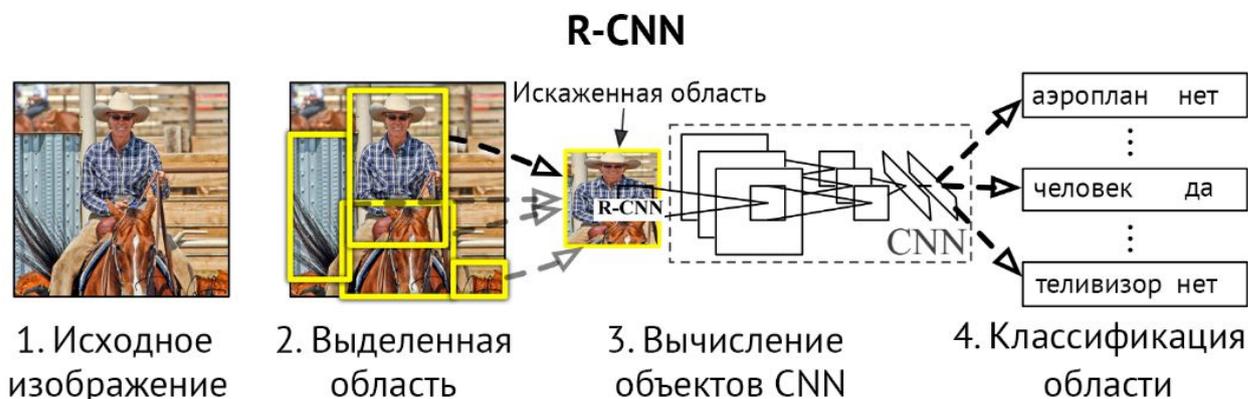
В отличие от обычной задачи классификации, где конечный результат глубокой сети является единственной важной вещью, семантическая сегментация требует не только

различия на уровне пикселей, но также и механизма для проецирования распознающих признаков, изученных на разных этапах кодирующей сети, в пространство пикселей. Различные подходы используют разные механизмы, являющиеся частью декодирующей сети. Три основных подхода семантической сегментации:

1. Областная семантическая сегментация
2. Семантическая сегментация на основе полностью сверточной сети
3. Семантическая сегментация со слабым контролем

*Областная семантическая сегментация* обычно использует в качестве источника информации данные, полученные с помощью «сегментации с использованием распознавания». Данный метод сначала извлекает области свободной формы из изображения и описывает их, а затем классифицирует их. Во время тестирования предсказания данного метода преобразуются в пиксели. Обычно это происходит путем маркировки пикселя в соответствии с областью наивысшей оценки, которая его содержит.

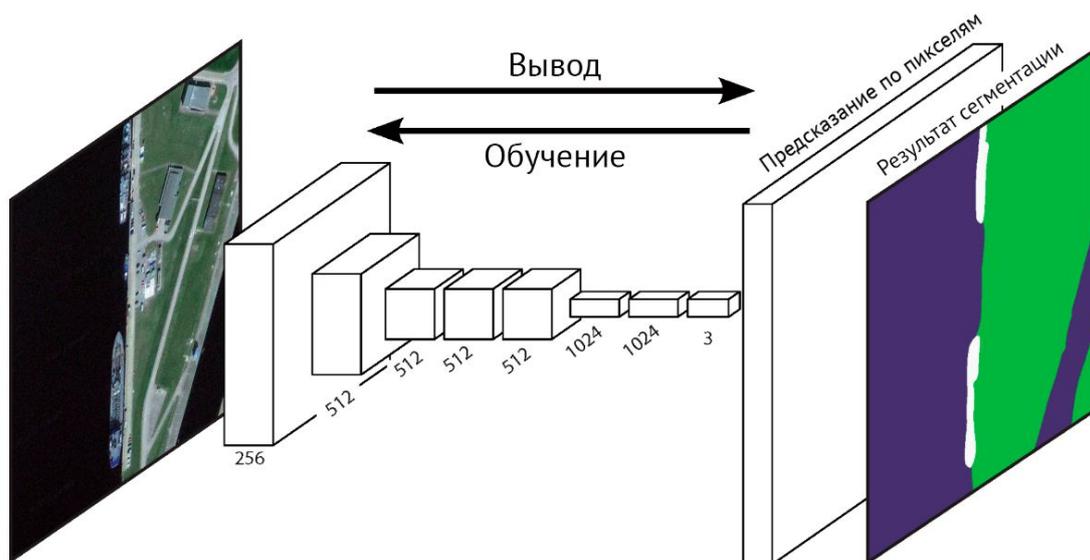
R-CNN - репрезентативная сеть для использования методов, основанных на областных выделениях. Она выполняет семантическую сегментацию на основе результатов обнаружения объекта. Изначально R-CNN использует выборочный поиск для извлечения большого количества предполагаемых объектов, а затем вычисляет функции CNN для каждого из них. В итоге, она классифицирует каждую область изображения, используя классовые линейные SVM. По сравнению с традиционными структурами CNN, которые в основном предназначены для классификации изображений, R-CNN может решать более сложные задачи, такие как обнаружение объектов и сегментация изображений. Она может послужить основой для обеих задач. Кроме того, R-CNN может быть построена поверх любых эталонных структур CNN, таких как AlexNet, VGG, GoogLeNet и ResNet.



**Рис. 2.** Архитектура R-CNN

Для задачи сегментации изображения R-CNN выделила 2 типа объектов для каждой области: объект полного региона и объект переднего плана и обнаружила, что это может привести к повышению производительности при объединении этих областей вместе. R-CNN позволяет значительно улучшить производительность благодаря использованию сильно различающихся функций CNN. Тем не менее, она также страдает от нескольких недостатков для задачи сегментации.

Семантическая сегментация на основе полностью сверточной сети (FCN) обрабатывает изображение от пикселя к пикселю без извлечения объектов по определенной области. Сетевой конвейер FCN является продолжением классического конвейера CNN. Основная идея заключается в том, чтобы заставить классическую архитектуру CNN принимать в качестве входных данных изображения произвольного размера. Ограничение CNN принимать и производить разметку только для входов определенного размера исходит от полностью связанных слоев, которые являются фиксированными. В отличие от них, FCN имеет сверточный и объединяющий слои, которые дают им возможность делать прогнозы на входах произвольного размера.



**Рис. 3.** Архитектура FCN

Одна из проблем данной архитектуры заключается в том, что при распространении через несколько чередующихся сверточных и объединяющих слоев разрешение выходных карт объектов снижается. Поэтому прямые предсказания FCN обычно имеют низкое разрешение, что приводит к относительно нечетким границам объектов. Для решения этой проблемы было предложено множество более продвинутых подходов, основанных на FCN, включая SegNet, DeepLab-CRF и Dilated Convolutions.

*Семантическая сегментация со слабым контролем.* Большинство соответствующих методов в семантической сегментации полагаются на большое количество изображений с пиксельными масками. Однако ручное выделение этих масок является довольно трудоемким, утомительным и коммерчески дорогим процессом. Поэтому в последнее время были предложены некоторые слабо контролируемые методы, предназначенные для выполнения семантической сегментации с использованием аннотированных ограничивающих рамок.



**Рис. 4.** Обучение Voxsup

Например, Voxsup использовал аннотирование ограничивающей области в качестве контроля для обучения сети и итеративного улучшения оценочных масок для семантической сегментации. Классификация на уровне пикселей позволила интерпретировать задачу сегментации в структуре обучения с несколькими экземплярами и позволила добавить дополнительный слой, чтобы ограничить модель для присвоения большего веса всем важным пикселям для классификации на уровне изображения.

### **Семантическая сегментация с использованием полностью сверточной сети**

Вот ключевые особенности архитектуры FCN:

1. FCN передает знания из VGG16 для выполнения семантической сегментации.
2. Полностью соединенные слои VGG16 преобразуются в полностью сверточные слои с использованием свертки 1x1. Этот процесс создает тепловую карту присутствия класса в низком разрешении.
3. Повышение дискретизации карт семантических признаков низкого разрешения выполняется с использованием транспонированных сверток (инициализированных с помощью билинейных интерполяционных фильтров).
4. На каждом этапе процесс повышения дискретизации дополнительно совершенствуется путем добавления функций из более неточных, но с более высоким разрешением карт объектов из более низких уровней в VGG16.

5. Пропускное соединение вводится после каждого блока свертки, чтобы позволить последующему блоку извлекать более абстрактные, характерные для класса функции из ранее объединенных функций.

Далее будет рассматриваться пошаговая реализацию самой популярной архитектуры семантической сегментации - полностью сверточной сети (FCN-8). Для реализации используется библиотека TensorFlow (Python 3), а также другие зависимости, такие как NumPy и Scipy.

Пошаговое внедрение архитектуры FCN-8:

1. Кодировочная сеть: предварительно обученная сеть VGG16 используется в качестве кодировщика. Декодировочная сеть начинается со слоя 7 VGG16.
2. Слой FCN-8: последний полностью подключенный слой VGG16 заменяется сверткой  $1 \times 1$ .
3. Слой FCN-9: слой FCN-8 увеличивается в 2 раза, чтобы соответствовать размерам слоя 4 VGG16, с использованием транспонированной свертки с параметрами: (kernel=(4,4), stride=(2,2), padding='same'). После этого необходимо добавить пропущенное соединение между слоем 4 VGG16 и слоем 9 FCN.
4. Слой FCN-10: слой FCN-9 увеличивается в 2 раза для соответствия размерам со слоем 3 VGG16 с использованием транспонированной свертки с параметрами: (kernel=(4,4), stride=(2,2), padding='same'). После этого было добавлено пропущенное соединение между слоем 3 VGG16 и слоем 10 FCN.
5. Слой FCN-11: слой FCN Layer-10 увеличивается в 4 раза для согласования размеров с размером входного изображения, поэтому мы получаем реальное изображение обратно, а глубина равна количеству классов, используя транспонированную свертку с параметрами: (kernel = (16,16) , stepde = (8,8), padding = 'same').

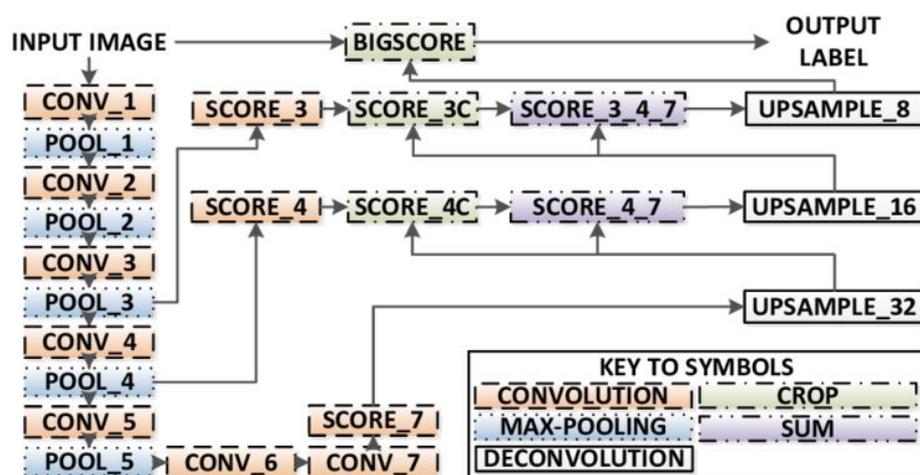


Рис. 5. Архитектура FCN-8

После реализации архитектуры FCN-8, необходимо подготовить обучающие данные, или взять их из свободных ресурсов и обучить сеть. Результат распознавания должен получиться как на рис. 6.



**Рис. 6.** Результат распознавания нейронной сети FCN-8

### **Заключение**

Задача семантической сегментации является высокоуровневой задачей обработки изображений, относящейся к группе задач машинного распознавания. Она является более сложной, чем задача классификации изображений и поиска объектов, что обусловлено не только необходимостью определения классов объектов, но и выявления их структуры, правильного выделения частей объектов на изображении.

Существует три основных метода семантической сегментации: областная семантическая сегментация, семантическая сегментация на основе полностью сверточной

сети, семантическая сегментация со слабым контролем. Данные методы включают в себя разные архитектуры нейронных сетей и постоянно совершенствуются.

В рамках статьи была описана реализация архитектур FCN-8 с использованием библиотеки машинного обучения tensorflow.

### Список литературы

1. A. Dadhich Practical COMPUTER VISION: EXTRACT INSIGHTFUL INFORMATION FROM IMAGES USING TENSORFLOW, KERAS, AND OPENCV // Packt Publishing Ltd. 2018. с. 57-67
2. J. C. Russ THE IMAGE PROCESSING HANDBOOK // CRC Press. 2016. с. 43-57
3. AGGREGATED RESIDUAL TRANSFORMATIONS FOR DEEP NEURAL NETWORKS // Электронный ресурс – Режим доступа: <https://arxiv.org/pdf/1611.05431.pdf> (дата обращения 18.12.2017).
4. Рыбкин С.В., Нестеров А.Ю. ПОДХОДЫ В РАСПОЗНАВАНИИ ДОРОЖНЫХ ЗНАКОВ // В сборнике: ПРОБЛЕМЫ ФОРМИРОВАНИЯ ЕДИНОГО НАУЧНОГО ПРОСТРАНСТВА. Сборник статей Международной научно-практической конференции. В 4-х частях. 2017. с. 16-19
5. Вершинин В.Е., Гришунов С.С., Логинова М.Б. МОДЕЛИРОВАНИЕ ПРОЦЕССОВ РАСПОЗНАВАНИЯ И КЛАССИФИКАЦИИ МНОГОМЕРНЫХ ОБЪЕКТОВ ПЕРЕСЕКАЮЩИХСЯ КЛАССОВ НА ОСНОВЕ ПРЕДСТАВЛЕНИЙ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ // Электронный журнал: наука, техника и образование. 2016. № 1 (5). С. 120-133.
6. Бобылева Е.А., Драч В.Е., Родионов А.В., Чухраев И.В. РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ СИМВОЛОВ С ПОМОЩЬЮ КАСКАДНЫХ КЛАССИФИКАТОРОВ // Электромагнитные волны и электронные системы. 2017. Т. 22. 3. С. 61-66.
7. Нестеров А.Ю., Бурмистров А.В., Белов Ю.С. МЕТОД ОПРЕДЕЛЕНИЯ ПОЛОЖЕНИЯ ОБЪЕКТА В ЗАДАЧАХ РАСПОЗНАВАНИЯ ОБРАЗОВ // Электронный журнал: наука, техника и образование. 2016. № 2 (6). С. 82-89.
8. Гуркина Е.Д., Белов Ю.С. ОБНАРУЖЕНИЕ ЧАСТИЧНО РАЗМЫТЫХ ИЗОБРАЖЕНИЙ И ИХ КЛАССИФИКАЦИЯ // Научное обозрение. Технические науки. 2017. № 2. С. 38-41.