

# МОДЕЛЬ ПОЛНОСВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ ДЕРЕВЬЕВ ПИХТЫ НА ИЗОБРАЖЕНИЯХ С БПЛА

Мачука М.К.Р.

Томский политехнический университет, ИШИТР, аспирант, гр. А2-36,

e-mail: kristianrodrigo1@tpu.ru

## Аннотация

В данной работе представлена модифицированная модель нейронной сети на основе U-Net для классификации деревьев *A. Sibirica*. Модель включает в себя пакетную нормализацию, сохранение размеров карт признаков и использование остаточных блоков. Доказано, что использование более крупных фрагментов изображения улучшает способность нейронной сети к обобщению, обеспечивая высокую эффективность семантической сегментации.

**Ключевые слова:** семантическая сегментация изображений, полносверточная нейронная сеть, деревья сибирской пихты, беспилотный летательный аппарат.

## Введение

Семантическая сегментация — важная задача в области компьютерного зрения, которая позволяет понимать содержимое изображений на уровне пикселей (попиксельная классификация). Размер входных изображений является одним из ключевых факторов, определяющих эффективность таких сетей. Поэтому сегодня актуальной является задача исследования влияния размера входных изображений на точность семантической сегментации хвойных деревьев на снимках с помощью полносверточных нейронных сетей.

В данной работе рассматривается задача обучения и исследования предложенной модели полносверточной нейронной сети, основанной на архитектуре классической модели U-Net [1]. Эта модель применяется для семантической сегментации фрагментов двух размеров, полученных из исходных снимков с беспилотных летательных аппаратов (БПЛА) деревьев сибирской пихты *Abies sibirica* (далее – *A. Sibirica*), поврежденных уссурийским полиграфом *Polygraphus proximus* (Blandford) [2].

## Исходные данные для исследования

В качестве исходных данных использовались пять RGB-снимков, сделанных на штатную фотокамеру БПЛА DJI Phantom 3 Standard. При анализе снимков с БПЛА деревьев *A. sibirica*, поврежденных уссурийским полиграфом, экспертами выделялись четыре класса их жизненного состояния [2]: «Живые», «Отмирающие» (сильно ослабленные, усыхающие), «Свежий сухостой» и «Старый сухостой» (сухостой прошлых лет). Пятым классом являлся «Фон» – деревья других пород и иные объекты земной поверхности. Каждое из деревьев экспертом путем визуального дешифрирования относилось к одному из классов. Кроме разметки деревьев по классам осуществлялось оконтуривание их крон. В итоге получены эталонные карты сегментации.

Для формирования первого датасета (обучающей и валидационной выборки), четыре исходных снимка деревьев *A. sibirica* были нарезаны на фрагменты размером 256 x 256 пикселей путем сдвига предыдущего фрагмента на 128 пикселей. Для увеличения объема данных было применено несколько поворотов. Всего было получено 2004 обучающих и 672 валидационных фрагментов. Тестовая выборка, созданная по снимку тестового участка, составила 196 фрагментов размером 256 x 256 пикселей.

Для формирования второго датасета, такие же четыре исходных снимка деревьев *A. sibirica* были нарезаны на фрагменты размером 480 x 480 пикселей путем сдвига предыдущего фрагмента на 240 пикселей. Всего было получено 240 обучающих и 40 валидационных фрагментов. Тестовая выборка, созданная по снимку тестового участка, составила 72 фрагмента размером 480 x 480 пикселей.

## Модель полносверточной нейронной сети

Для решения задачи мультиклассификации деревьев *A. sibirica* была разработана новая модель полносверточной нейронной сети, основанная на известной модели полносверточной сети U-Net, которая обычно используется для семантической сегментации (попиксельной классификации)

изображений. Ее архитектура приведена на рис. 1. Внесенные изменения в исходную модель U-Net включают в себя:

- Изменение размеров входных и выходных тензоров.
- Применение пакетной нормализации (BN) после каждой нелинейности.
- Сохранение размеров карт признаков при применении слоев свертки. Это изменение позволяет картам признаков сохранять свои пространственные размеры, предотвращая потерю информации во время субдискретизации, что крайне важно для сохранения деталей на изображениях леса.
- В отличие от исходной архитектуры U-Net, модифицированная версия не обрезает карт признаков для соединений пропуска. Это позволяет сети сохранять пространственный контекст, сохранять исходный размер карт признаков с сокращающегося пути и обеспечивать более полное и богатое представление признаков для реконструкции.
- Расширенный путь в модифицированной архитектуре U-Net включает интерполяцию карт признаков с последующей сверткой 3x3. Эта операция помогает восстановить пространственные размеры карт признаков, уменьшая при этом количество каналов признаков
- В выходном слое модифицированной сети U-Net свертки 1x1 и функция активации Softmax используются, чтобы гарантировать, что выходные данные представляют собой вектор с вероятностями каждого возможного результата, представляющий классы  $C$ .

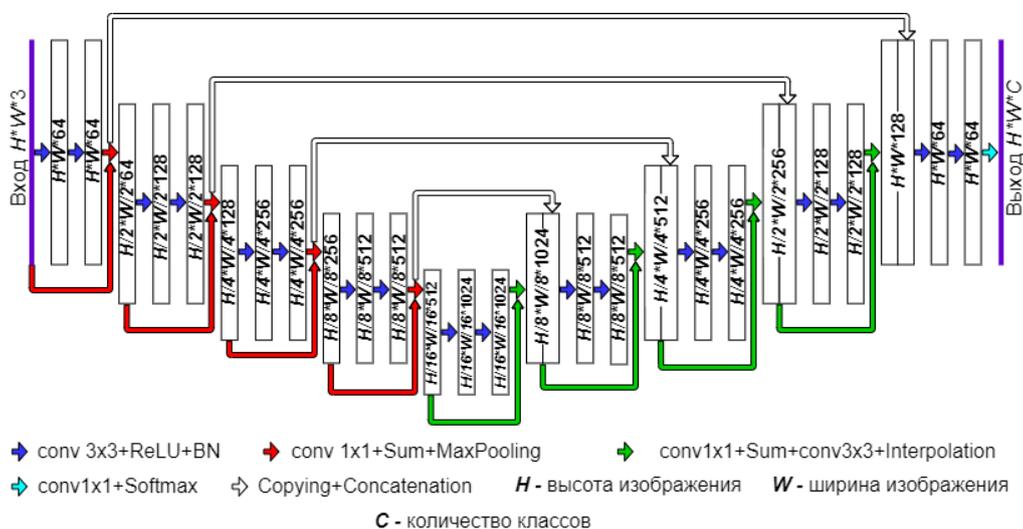


Рис. 1. Архитектура модифицированной модели U-Net

В классической модели U-Net выходные данные одного слоя служат входными данными для следующего слоя. Во время обучения нейронной сети градиенты распространяются по сети обратно, и они могут уменьшаться по величине при прохождении через несколько слоев, особенно в очень глубоких сетях. Это явление может затруднить процесс обучения, что приведет к замедлению сходимости или даже к невозможности эффективного обучения сети [3]. Для решения этой проблемы в модель были добавлены остаточные блоки [4]. Такие блоки позволяют обучать глубокие нейронные сети эффективнее, позволяя обучать остаточные признаки (изменения внесенные каждым слоем), что упрощает обучение и сокращает время до достижения оптимальных результатов.

### Обучение и валидация предлагаемой модели сети

При обучении предлагаемой модели полносверточной сети применялась функция потерь Focal Loss - стандартной функции потерь для устранения крайнего дисбаланса классов [5]. Функция Focal Loss включает в себя модулирующие параметры, которые позволяют сокращать потери от простых примеров и увеличивать значение корректирующих примеров в случаях неправильной классификации пикселей деревьев *A. sibirica*. При обучении использовался известный алгоритм оптимизации Adam. Более того, при обучении моделей также применялась техника аугментации "на лету", позволяющая улучшить разнообразие и насыщенность обучающих фрагментов за счет синтезированных фрагментов непосредственно во время обучения.

Чтобы оптимизировать эффективность модели, использовался метод байесовской оптимизации для настройки гиперпараметров. Мы выделили три ключевых параметра, которые существенно влияют на эффективность модели: аугментация данных, скорость обучения и параметры функции потерь. Для аугментации данных, яркости, контрастности, резкости и насыщенности мы стремимся оптимизировать связанные с ними диапазоны коэффициентов для изменения исходного изображения. Определение подходящего диапазона позволяет модели эффективно изучать как локальные, так и глобальные признаки данных. Скорость обучения играет важную роль в определении размера шага, с которым модель обновляет свои внутренние параметры во время обучения. Более высокая скорость обучения способствует более быстрой сходимости, но может привести к риску выйти за рамки оптимального решения. И наоборот, более низкая скорость обучения приводит к более медленной сходимости. Кроме того, были оптимизированы параметры Focal Loss, которые уравнивают положительные и отрицательные примеры и регулируют скорость снижения веса простых примеров во время обучения.

Всего для обучения модели для каждого из наборов данных было использовано 100 уникальных комбинаций гиперпараметров. Была построена вероятностная модель, отражающая взаимосвязь между значениями гиперпараметров и эффективностью моделей на валидационных выборках, измеряемую метрикой Intersection over Union (mIoU). Процесс поиска гиперпараметров управлялся функцией сбора данных, которая определяла следующий набор гиперпараметров для оценки на основе текущей вероятностной модели. После оценки каждого набора гиперпараметров модель обновлялась, и новая модель обучалась с использованием уточненных гиперпараметров. Затем эффективность новой модели была измерена и использована для дальнейшего обновления вероятностной модели. Этот итерационный процесс повторялся 100 раз, при этом на каждой итерации использовалась обновленная модель для определения наилучшей комбинации гиперпараметров, оценки их производительности и продолжения совершенствования модели.

## Результаты и их обсуждение

Из визуального анализа результатов обучения и валидации модели, приведенных в качестве примера на рис. 2, следует, что предложенная модель для каждого из датасетов успешно воспроизводит границы между деревьями и сегментирует большую долю кроны.

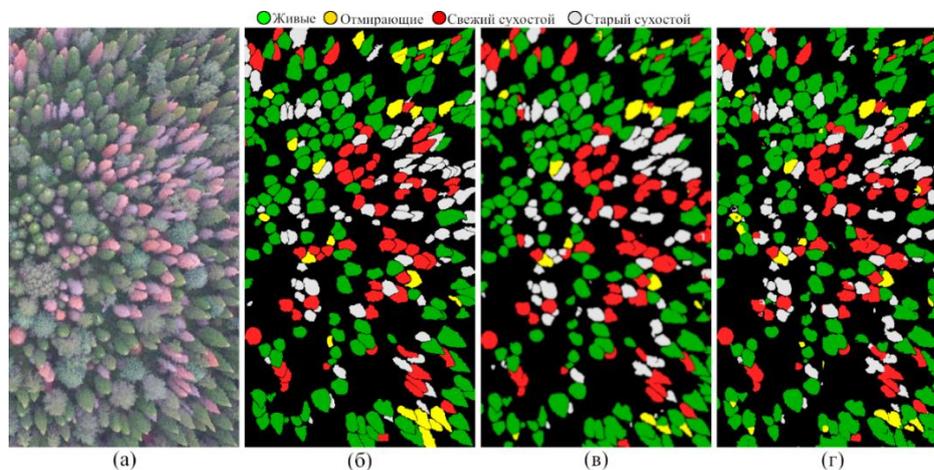


Рис. 2. Пример сегментации деревьев *A. Sibirica* на снимке:

- a* – исходный снимок; *б* – эталонная карта сегментации этого снимка, полученная экспертами;
- в* – результат для случая фрагментов 256\*256 пикселей;
- г* – результат для случая фрагментов 480\*480 пикселей

Из таблицы 1, где приведены результаты сегментации на тестовой выборке, видно, что модель полносверточной сети для каждой из датасетов продемонстрировала высокое качество классификации пикселей всех (IoUc > 0.5). Однако при обучении модели с использованием датасета фрагментов размером 256\*256 пикселей результаты ниже, чем при использовании датасета фрагментов размером 480\*480 пикселей.

Метрики качества сегментации деревьев *A. Sibirica* на тестовой выборке

Размер фрагмента, пикселей	IoU					mIoU
	Фон	Живые	Отмирающие	Свежий сухой	Старый сухой	
256*256	0,85	0,72	0,51	0,75	0,66	0,70
480*480	<b>0,86</b>	<b>0,76</b>	<b>0,57</b>	<b>0,80</b>	<b>0,69</b>	<b>0,74</b>

Есть несколько причин, по которым фрагменты большего размера могут привести к лучшим результатам, несмотря на их меньшее количество:

- Фрагменты большего размера предоставляют модели более широкий контекст, позволяющий понять пространственные отношения между различными частями изображения. Эта дополнительная контекстная информация может помочь модели делать более точные прогнозы.
- Благодаря более крупным фрагментам модель имеет доступ к более разнообразным и подробным признакам, что может помочь в обучении более богатых представлений данных. Это может привести к лучшему обобщению и эффективности новых данных.
- Фрагменты большего размера могут помочь уменьшить переобучение, предоставляя более разнообразные примеры для модели, на которых можно учиться. Это особенно важно при работе со сложными наборами данных и модель имеет большое количество параметров.
- Фрагменты большего размера содержат больше различительных признаков, которые могут помочь модели более эффективно различать разные классы. Это может привести к лучшим результатам сегментации, особенно при работе с визуально похожими классами.
- Большие фрагменты могут повысить устойчивость модели к шуму и изменчивости данных, предоставляя модели больше информации для обучения. Это может помочь модели лучше обобщать невидимые данные и повысить ее производительность в реальных сценариях.

В целом, хотя фрагменты большего размера могут потребовать больше вычислительных ресурсов и времени для их анализа, они обеспечивают преимущества с точки зрения эффективности модели и обобщения, что приводит к лучшим результатам, несмотря на их меньшее количество.

## Заключение

Рассмотрено обучение предложенной модели полносверточной нейронной сети для классификации деревьев *A. sibirica*, поврежденных уссурийским полиграфом *Polygraphus proximus*, на снимках с БПЛА. Исследования показали, что изменения, внесенные в классическую модель U-Net, такие как пакетная нормализация, сохранение размеров карт признаков и использование остаточных блоков, значительно повышают ее эффективность и точность в задаче семантической сегментации изображений лесного покрова. В результате тщательного анализа было установлено, что использование крупных фрагментов изображения приводит к заметному улучшению понимания контекста сети и обогащению признаков, что, в свою очередь, способствует повышению качества сегментации. Таким образом, разработанная модель представляет собой важный шаг в развитии методов анализа лесных экосистем с использованием нейронных сетей и изображений с БПЛА.

## Список использованных источников

1. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. [Электронный ресурс]. – URL: <https://arxiv.org/pdf/1505.04597.pdf> (дата обращения 12.11.2023).
2. Кривец С.А., Бисирова Э.М., Волкова Е.С., Дебков Н.М., Керчев И.А., Мельник М.А., Никифоров А.Н., Чернова Н.А. Технология мониторинга пихтовых лесов в зоне инвазии уссурийского полиграфа в Сибири: метод. пособие. Томск: УМИУМ. – 2018. – Т. 74. – С. 3.
3. Noh S.-H. Analysis of Gradient Vanishing of RNNs and Performance Comparison // Information. – 2021. – Vol. 12. – P 442.
4. He K., Xiangyu Z., Shaoqing R., Jian S. Deep residual learning for image recognition // In Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 770-778.
5. Lin T., Goyal P., Girshick R., He K., Dollár P. Focal Loss for Dense Object Detection // Facebook AI Research (FAIR). [Электронный ресурс]. – URL: <https://arxiv.org/pdf/1708.02002.pdf> (дата обращения 28.12.2023).