

ИССЛЕДОВАНИЕ АРХИТЕКТУР СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ СНИМКОВ ДЕРЕВЬЕВ СИБИРСКОЙ ПИХТЫ

Малкин А.Ю.¹, Мыцко Е.А.²

¹ Томский политехнический университет, ОИТ ИШИТР, магистрант гр. 8ВМ22, e-mail: mal-kin@tpu.ru

² Томский политехнический университет, ОИТ ИШИТР, доцент, e-mail: evgenvt@tpu.ru

Введение

Одной из проблем экологии является болезненное состояние деревьев в лесных массивах. Это происходит на фоне потепления климата и активизации вредителей (различные паразиты, такие как уссурийский полиграф). Для своевременного принятия мер, необходимо как можно скорее обнаружить поврежденное дерево. Поскольку лесные массивы распространяются на большие площади, используется технология сканирования леса с применением беспилотных летательных аппаратов (БПЛА), оснащённых фото- или видеокамерой [1]. При решении задач автоматизированного распознавания состояния деревьев для своевременной передачи этой информации специалистам необходимо использовать современные технологии, такие как искусственные нейронные сети [2].

Таким образом, актуальность данной работы заключается в применении современных технологий на основе свёрточных нейронных сетей для своевременного обнаружения сухостоев среди деревьев сибирской пихты. Целью данной работы является исследование архитектур свёрточных нейронных сетей для семантической сегментации снимков деревьев сибирской пихты.

Описание алгоритма

В рамках данной работы проводится сравнение эффективности работы мобильной системы компьютерного зрения на основе трёх архитектур свёрточных нейронных сетей: U-Net [3], ENet [4], SegNet [5].

Для обучения свёрточной нейронной сети применяется подготовленная выборка фрагментов снимков сибирской пихты с БПЛА [6]. Над выборкой проведена операция data augmentation (увеличение и уменьшение экспозиции изображения, добавления цифрового шума). Для этих фрагментов проведена семантическая сегментация по состоянию деревьев и получена маска со следующей цветовой схемой: синий – живое дерево, бирюзовый – отмирающее, фиолетовый – свежий сухостой, жёлтый – старый сухостой, чёрный – фон. Пример выборки представлен на рисунке 1.

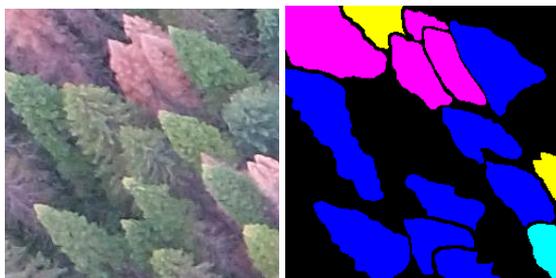


Рис. 1. Исходный снимок (слева) и эталонная маска (справа) изображения из датасета для сегментации пихты сибирской пихты.

В работе проведено обучение свёрточных нейронных сетей, построенных по архитектурам U-Net, ENet, SegNet с применением библиотеки Keras на языке программирования Python для фрагментов изображений размером 256x256 пикселей с применением графического ускорителя Nvidia Tesla T4 с 16 ГБ видеопамяти. Рассчитанные значения потерь (*loss*) и точности (*accuracy*) для обучающей и валидационной выборок представлены в таблице 1. Экспериментальным путём установлено, что архитектура ENet оказалась наиболее «легковесной», что позволило использовать большее значение *batch_size* (размерность вектора изображений, подаваемых в память графического процессора за один раз) и уменьшить время на обучение одной эпохи.

Эффективность работы различных архитектур будем рассматривать с помощью F-меры (1).

$$F = 2 * \frac{Precision * Recall}{(Precision + Recall)}, \quad (1)$$

где *Precision* выражается точностью в сегментации (те пиксели, обозначающие какую-либо из категорий, не являющейся ею, уменьшают этот параметр), а *Recall* чувствителен к тем пикселям, которые не были обозначены системой как данная категория.

Таблица 1

Результаты обучения архитектур свёрточных нейронных сетей

Архитектура, число эпох	Training loss	Training accuracy	Validation loss	Validation training	Batch_size	Seconds for epoch
U-Net, 25	0,027	0,989	0,383	0,948	40	98
U-Net, 50	0,010	0,996	0,715	0,944		
ENet, 25	0,034	0,986	0,412	0,941	45	17
ENet, 50	0,020	0,992	0,589	0,937		
SegNet, 25	0,198	0,924	0,286	0,899	25	126
SegNet, 50	0,023	0,990	0,804	0,908		

Для каждой модели была проведена проверка на отдельной выборке, предназначенной для валидации. Использована F-мера для определения качества сегментации больных и здоровых сибирских пихт в лесах Томской области. В таблице 2 приведены значения *Recall* (полноты)/*Precision* (точность) каждой категории сегментации для валидационной выборки. В таблице 3 отображены значения F-меры сегментации.

Таблица 2

Значения Recall / Precision для валидационной выборки

Архитектура, число эпох	Живое	Отмирающее	Свежий сухостой	Старый сухостой	Фон
U-Net, 25	0,648 / 0,607	0,151 / 0,174	0,766 / 0,609	0,324 / 0,741	0,844 / 0,815
U-Net, 50	0,628 / 0,624	0,145 / 0,174	0,688 / 0,662	0,456 / 0,669	0,854 / 0,816
ENet, 25	0,641 / 0,606	0,100 / 0,280	0,629 / 0,671	0,456 / 0,554	0,854 / 0,817
ENet, 50	0,711 / 0,539	0,091 / 0,258	0,607 / 0,693	0,484 / 0,443	0,782 / 0,827
SegNet, 25	0,540 / 0,595	0,107 / 0,147	0,706 / 0,667	0,309 / 0,677	0,871 / 0,790
SegNet, 50	0,655 / 0,572	0,121 / 0,186	0,645 / 0,714	0,470 / 0,638	0,832 / 0,817

Таблица 3

Значения F-меры для валидационной выборки

Архитектура, число эпох	Живое	Отмирающее	Свежий сухостой	Старый сухостой	Фон	Среднее
U-Net, 25	0,627	0,162	0,679	0,451	0,830	0,549
U-Net, 50	0,626	0,158	0,675	0,543	0,835	0,567
ENet, 25	0,623	0,147	0,650	0,500	0,835	0,551
ENet, 50	0,613	0,135	0,647	0,463	0,804	0,532
SegNet, 25	0,566	0,124	0,686	0,424	0,829	0,526
SegNet, 50	0,611	0,146	0,678	0,542	0,824	0,560

По расчётам, представленным в таблице 3, можно сделать вывод, что архитектура U-Net является более точной при сегментации живых (62,6 %) и отмирающих (15,8 %) деревьев, фона (83,5 %). Одинаково с U-Net эффективна в определении старого сухостоя (54,3 %) архитектура SegNet (54,2 %), которая способна сегментировать свежий сухостой (67,8 %). Архитектура ENet имеет меньшее значение F-меры в среднем (53,2 %) в сравнении с SegNet (56,0 %) и U-Net (56,7 %).

В таблице 4 представлено время обработки обученной нейросетью валидационной выборки из 672 фрагментов 256x256. Эта информация показывает, насколько выбранная архитектура пригодна для реализации в системе реального времени. Замер времени проводился по десять раз для каждой архитектуры, подавалось одно изображение за раз (*batch_size = 1*).

Время обработки валидационной выборки различными архитектурами (672 фрагмента с разрешением 256x256 пикселей)

Архитектура	Минимум, с	Среднее, с	Максимум, с	Время на обработку одного фрагмента, мс
U-Net	20,027	20,834	22,279	28
ENet	4,063	4,720	5,437	4
SegNet	21,609	22,013	23,703	30

По таблице 4 можно сделать вывод, что архитектура ENet является наиболее быстродействующей и приспособленной для реализации в системах реального времени: время, требуемое ENet для обработки одного фрагмента 256x256, в 7 раз меньше, чем у U-Net, и 7,5 раза меньше, чем у SegNet.

Рассмотрим визуально сегментацию одного изображения размером 256x256. Как видим на рисунке 2, наибольшую трудность для всех архитектур составляют точная сегментация сухостоя (свежий или старый – фиолетовый и жёлтый соответственно) и определение отмирающих деревьев (бирюзовый). Для архитектур SegNet и ENet характерно ошибочное определение листовенного дерева как хвойного, что может быть критичным в некоторых задачах.

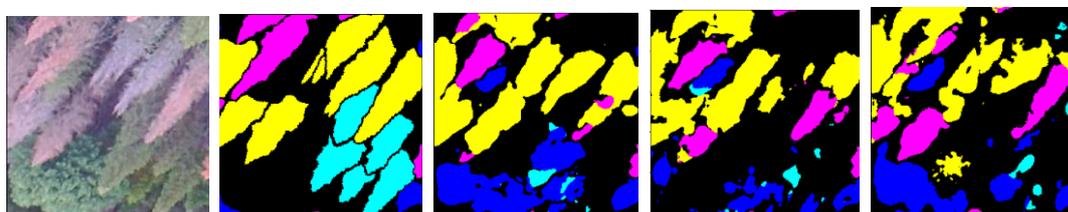


Рис. 2. Слева направо: исходное, эталонное изображения, обработка SegNet, U-Net, ENet

Заключение

В настоящее время проведена работа по обучению и исследованию свёрточных нейронных сетей различных архитектур для разработки мобильной системы компьютерного зрения. Изучена особенность работы архитектур U-Net, SegNet, ENet, рассчитана F-мера, как метрика точности сегментации изображений.

В результате установлено, что наиболее точно выполняет сегментацию фрагментов изображений сибирской пихты размером 256x256 пикселей архитектура U-Net, что демонстрирует более высокое значение F-меры относительно архитектур SegNet и ENet. При проектировании системы реального времени архитектура ENet является более предпочтительной, так как требует меньше времени для обработки одного фрагмента в сравнении с другими архитектурами.

Список использованных источников

1. Денисов С. А., Домрачев А. А., Елсуков А. С. Опыт применения квадрокоптера для мониторинга возобновления леса // Вестник Поволжского государственного технологического университета. Сер.: Лес. Экология. Природопользование. – 2016. – № 4 (32). – С. 34–46.
2. Shasha Lu: Examining the influencing factors of forest health, its implications on rural revitalization. – 2021. URL: <https://www.sciencedirect.com/science/article/pii/S026483772032603X> (дата обращения: 09.01.2023).
3. Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. – 2015. URL: <https://arxiv.org/abs/1505.04597> (дата обращения: 09.02.2023).
4. Adam Paszke, Abhishek Chaurasia, Sangpil Kim, Eugenio Culurciello. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. – 2015. URL: <https://arxiv.org/abs/1606.02147> (дата обращения: 09.02.2023).
5. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. – 2015. URL: <https://arxiv.org/abs/1511.00561> (дата обращения: 09.02.2023).
6. Керчев И. А., Маслов К. А., Марков Н. Г., Токарева О. С. Семантическая сегментация поврежденных деревьев пихты на снимках с беспилотных летательных аппаратов // Современные проблемы дистанционного зондирования Земли из космоса. – 2021. – Т. 18. – №. 1. – С. 116–126.