

УДК 004.032.26

ИССЛЕДОВАНИЕ АППАРАТНО-РЕАЛИЗОВАННЫХ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ КЛАССА U-NET

Зоев Иван Владимирович¹,
ivz3@tpu.ru

Маслов Константин Андреевич²,
kocheg@mail.ru; orpheus.const@outlook.com

Марков Николай Григорьевич²,
orloff234@gmail.ru; markovng@tpu.ru

Мыцко Евгений Алексеевич²,
orloff234@gmail.ru; evgenvt@tpu.ru

¹ АО «ИнфоТекС»,
Россия, 634041, г. Томск, пр. Кирова, 51.

² Национальный исследовательский Томский политехнический университет,
Россия, 634050, г. Томск, пр. Ленина, 30.

Разработаны и программно-реализованы две сверточные нейронные сети класса U-Net: модификация классической U-Net и U-Net с дилатационными свертками. Для обучения и тестирования сверточных нейронных сетей использованы датасеты на основе снимков с беспилотного летательного аппарата деревьев пихты, поврежденных уссурийским полиграфом. В зависимости от степени повреждения на снимках присутствуют деревья четырех классов и фон. Полученные при обучении весовые коэффициенты для каждой из сверточных нейронных сетей затем использовались при аппаратной реализации сверточных нейронных сетей на программируемой логической интегральной схеме системы на кристалле Zynq 7000 (Kintex FPGA) компании Xilinx. Приведены результаты исследования точности сегментации и производительности каждой аппаратно-реализованной на программируемой логической интегральной схеме сверточной нейронной сети.

Ключевые слова: сверточные нейронные сети, программируемая логическая интегральная схема, аппаратно-реализованные сверточные нейронные сети, U-Net.

Введение

В настоящее время в мире актуальными являются задачи распознавания различных объектов (лица людей, автомобили, объекты земной поверхности и т. д.) и мониторинга их состояния. Для решения таких задач успешно используются сверточные нейронные сети (СНС) различных классов [1, 2]. Однако при решении некоторых из таких задач возникает потребность в использовании беспилотных летательных аппаратов (БПЛА) с установленным на них специальным оборудованием для мониторинга труднодоступных объектов (опасные технологические объекты на промыслах, лесные массивы в горных районах и т. п.). В таких случаях для мониторинга требуется оснащать каждый БПЛА интеллектуальной системой компьютерного зрения (СКЗ), которая будет включать: видеокамеру, тепловизор и вычислительное устройство (ВУ) с аппаратно-реализованной СНС, позволяющее прямо на борту беспилотного аппарата решать задачи распознавания объектов различной физической природы (автомобили, технологические объекты, люди и т. д.). При проектировании таких мобильных систем мониторинга с интеллектуальными СКЗ необходимо соблюдать баланс между быстродействием используемого ВУ, точностью распознавания объектов земной поверхности с помощью СНС, массой и энергопотреблением этого устройства с целью увеличения времени полета БПЛА.

Данная работа посвящена исследованию эффективности аппаратно-реализованных моделей СНС класса U-net на программируемых логических интегральных схемах (ПЛИС) современных систем на кристалле (СнК). Для обучения, верификации и исследования таких моделей СНС применялся датасет, созданный с использованием снимков с БПЛА деревьев пихты сибирской, поврежденных уссурийским полиграфом. В зависимости от степени повреждения вредителем на снимках присутствуют деревья четырех классов пихты и фон. Результаты исследования аппаратной реализации моделей СНС класса U-net в составе ВУ сравниваются по критериям быстродействия, точности сегментации изображений деревьев, а также энергопотребления с результатами исследования программных реализаций этих же моделей СНС.

Аппаратно-реализованные на ПЛИС модели СНС

Кратко проанализируем результаты проведенных ранее исследований в области аппаратной реализации на ПЛИС моделей СНС, предназначенных для решения задач семантической сегментации изображений. К таким моделям СНС обычно относят полносверточные сети U-Net и SegNet и их модификации. В последние годы эта область развивается весьма интенсивно, причем наибольшее число исследований проведено для реализованных на ПЛИС моделей СНС класса U-Net [3–5]. Важные результаты в случае моделей SegNet приведены в

[6, 7], где описаны масштабные исследования производительности и энергопотребления акселераторов, реализованных на ПЛИС системы на кристалле (СнК) Intel Arria-10 GX1150. Однако только в [7] исследована точность (качество) сегментации изображений с помощью реализованной в акселераторе оптимизированной модели SegNet. Отметим, что для создания акселераторов используются в основном ПЛИС современных СнК, например, Intel Arria-10 GX1150 в работах [6, 7] и Xilinx Zynq ZC706 в [8]. Однако в [3] применена устаревшая ПЛИС Altera Cyclone V со скромными ресурсами, но обладающая минимальным энергопотреблением.

В большинстве исследований для повышения производительности акселераторов на ПЛИС предлагается уменьшать сложность архитектуры базовых моделей СНС U-Net и SegNet. Например, применяются несложные энкодер и декодер с разделенными по глубине свертками [3, 4], а в [5] для эффективной реализации свертки и деконволюции используется один модуль умножения векторов и применяются различные уровни параллельных вычислений моделей СНС. Для распознавания лесных пожаров в [9] предложена и исследована несложная модификация архитектуры легковесной СНС Squeeze U-Net. Кроме того, для повышения производительности акселераторов на ПЛИС используются операции над числами 8-16-32-bit fixed-point [5, 7, 8]. Результаты исследования производительности акселератора при работе с числами 32-bit float приведены только в [4], а для случая использования чисел 16-bit float исследования отсутствуют. Если исследования производительности акселераторов при использовании новых архитектур в этих работах проведены масштабно, то исследования потерь точности сегментации для новых архитектур не являются полноценными. Иногда приведены бездоказательные тезисы о том, что точность для той или

иной новой модели СНС несколько уменьшается по сравнению с точностью сегментации, даваемой базовой моделью U-Net или SegNet. Все это указывает на актуальность проведения масштабных исследований точности сегментации при трансформациях базовых моделей СНС U-Net и SegNet, в том числе при использовании операций над числами с плавающей запятой.

Из всех анализируемых акселераторов только рассмотренные в [3, 5, 9] акселераторы созданы с учетом особенностей решения задач сегментации изображений дистанционного зондирования. Более того, в [3, 9] рассматриваются простые задачи бинарной сегментации изображений с помощью моделей СНС на ПЛИС, а не решаются задачи распознавания объектов многих классов. Это еще раз подчеркивает актуальность исследований производительности и точности сегментации изображений с объектами нескольких классов с помощью аппаратно-реализованных на ПЛИС моделей класса U-Net.

Модели СНС класса U-Net

Ставилась задача семантической сегментации снимков с БПЛА деревьев пихты сибирской, поврежденных уссурийским полиграфом. В зависимости от степени повреждения вредителем на снимках присутствуют деревья четырех классов («живые», «отмирающие», «свежий сухостой» и «старый сухостой»), а также фон. Предлагаемые для решения этой задачи модели СНС класса U-Net разработаны на основе классической полносверточной сети U-Net [1]. Классическая СНС этого класса состоит из энкодера и декодера. Характерной особенностью моделей СНС класса U-Net является наличие операций конкатенации, соединяющих карты признаков из энкодера с картами признаков в декодере с целью повышения детальности результирующих карт классификации.

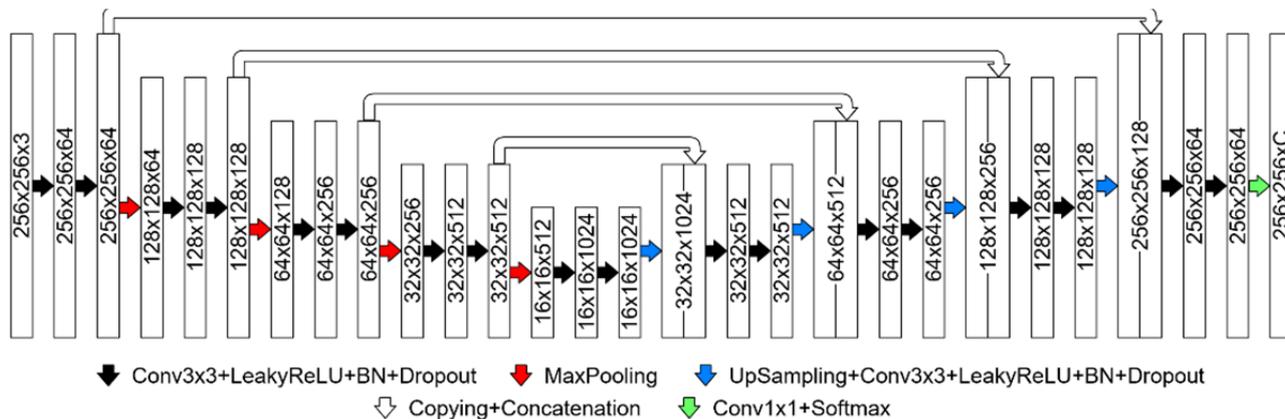


Рис. 1. Архитектура модели СНС, разработанной на основе классической U-Net
 Fig. 1. Architecture of the CNN model developed on the basis of the classical U-Net

Классическая модель СНС U-Net была модифицирована с учетом необходимости классификации пикселей, соответствующих четырем классам деревьев пихты и фону. На рис. 1 представлена архитектура разработанной модели СНС, где каждому прямоугольнику соответствует тензор – многомерный массив, представляющий собой набор карт признаков, числами показано число компонент тензоров. Стрелками показаны соответствующие операции: свертка (Conv3x3, Conv1x1), вычисление

функции активации leakyReLU (использовались также функции ReLU и eLU), пакетная нормализация (BN), уменьшение масштаба карт признаков путем выбора максимального значения в окрестности 2x2 в картах признаков (MaxPooling), операция увеличения масштаба карт признаков методом ближайшего соседа (UpSampling), копирование тензора и его конкатенация с другим (Copying+Concatenation) и исключение случайных сигналов слоя путем приравнивания их значений к

нулю (Dropout). Категориальное распределение на выходе декодера моделируется для каждого пикселя путем применения многомерной логистической функции Softmax. В отличие от классической U-Net эта модель СНС обладает следующими особенностями: входное изображение сети представлено тензором с числом компонентов $256 \times 256 \times 3$, что соответствует фрагменту трехканального RGB снимка; операции сверток не уменьшают размер выходных карт признаков; обрезка карт признаков не используется для соединений проброса; после каждого вычисления функции leakyReLU следует операция пакетной нормализации; входной тензор формируется с помощью сверток с ядрами размером 1×1 , позволяя тем самым сразу классифицировать пиксели С классов деревьев пихты и фон.

На основе этой разработанной модели СНС была создана другая модель СНС с дилатационными свертками (англ. dilated convolution) [10]. Дилатационная свертка отличается от классической свертки добавлением нулевых коэффициентов в сверточное ядро. Вводится параметр «фактор разреженности», который показывает расстояние между соседними ненулевыми коэффициентами ядра. В новой модели СНС каждые две свертки с ядрами размером 3×3 в модели на рис. 1 заменены на одну дилатационную свертку с ядром размером 5×5 (фактор разреженности равен 2). Это приводит к значительному упрощению модели СНС и, как следствие, должно позволить ускорить процесс выполнения СНС.

Для оценки эффективности предложенных моделей СНС при проведении их исследований использовалась метрика intersection over union (IoU) [11], считающаяся общепринятой метрикой эффективности при решении задачи семантической сегментации цифровых изображений. Известно, что значения IoU, превышающие 0,5, соответствуют приемлемому качеству сегментации. В качестве совокупного показателя эффективности модели использовалась метрика mean intersection over union (mIoU), рассчитываемая как среднее значение IoU по всем классам деревьев пихты и фону.

Обучение разработанных моделей СНС

Для программной реализации разработанных моделей СНС был использован язык программирования Python и библиотека Keras [12]. Для обучения программно-реализованных моделей СНС, их верификации и исследования использовался датасет, созданный нами по снимкам с БПЛА деревьев пихты, пораженных уссурийским полиграфом [13]. Распознаванию на снимках из датасета подлежали четыре класса деревьев пихты с разной степенью поражения и фон. С целью формирования обучающих и валидационных примеров для используемых моделей СНС снимки из обучающей выборки были нарезаны на фрагменты 256 на 256 пикселей. Всего было получено 2004 обучающих и 673 валидационных примеров. С целью улучшения обобщающей способности модели СНС была использована техника аугментации «на лету» (англ. online augmentation).

В качестве критерия для определения оптимальных весовых коэффициентов моделей СНС была использована максимизация метрики mIoU и алгоритм оптимизации Adam, основанный на вычислении градиента целевой функции. Оптимизация mIoU позволяет частично решить

выявленную проблему несбалансированности выборки по классам. Для поиска лучших наборов гиперпараметров для предложенных моделей СНС использовался байесовский подход.

Полученные таким образом при обучении каждой разработанной модели СНС оптимальные весовые коэффициенты модели переносились в СнК при аппаратной реализации моделей СНС.

Особенности аппаратной реализации предложенных моделей СНС

Известно, что современные мобильные СКЗ, устанавливаемые на автономных транспортных средствах, в том числе на БПЛА, должны иметь малое энергопотребление. Например, СКЗ на БПЛА должны потреблять не более 10–12 Вт. Поэтому ВУ таких СКЗ следует создавать на основе современных СнК с ПЛИС, обладающих малым энергопотреблением и высокой производительностью. Нами использовалась СнК Zynq 7000 (Kintex FPGA) компании Xilinx [14].

Архитектура большинства СнК позволяет организовать прямой доступ аппаратно-реализованной на ПЛИС модели СНС к внешней памяти. Такая организация взаимодействия ПЛИС с внешней памятью позволяет выполнять некоторые операции, отличные от процедур свертки и подвыборки СНС, не на ПЛИС, а на процессоре (процессорной системе) СнК. В работе [15] предложен оригинальный метод выполнения вычислений в аппаратной СНС на ПЛИС, отличающийся от известных методов использованием унифицированных вычислительных блоков свертки и подвыборки. Унификация блоков свертки/подвыборки достигается путем извлечения параметров блоков, обычно задающихся на этапе их синтеза, и размещения их в отдельную изменяемую область памяти ПЛИС, называемую конфигурационной областью памяти CONFIG space. Это позволяет использовать блоки в слоях СНС, имеющих разные архитектурные параметры. Более того, в слоях свертки и подвыборки аппаратной СНС используются блоки только одного типа – универсальные блоки. Такие блоки содержат как субблоки свертки, так и субблоки подвыборки [16]. Это позволяет более гибко организовать вычисления в аппаратной СНС. Реализация метода предполагает, что число задействованных в аппаратной СНС универсальных вычислительных блоков может быть переменным и определяется только ресурсами ПЛИС.

При реализации метода унификации кроме выделения конфигурационной области памяти CONFIG space необходимо было организовать доступ к этой области памяти процессору СнК, для чего было создано соответствующее программное обеспечение.

Унификация универсальных вычислительных блоков аппаратной СНС и предложенный в [17] способ организации вычислений в СнК с учетом необходимости взаимодействия ПЛИС с внешней памятью позволили реализовать ВУ на основе отладочной платы Avnet Mini-ITX Board с СнК Zynq 7000 (Kintex FPGA). Укрупненная функциональная схема этого ВУ представлена на рис. 2.

В устройстве (рис. 2) два основных функциональных блока.

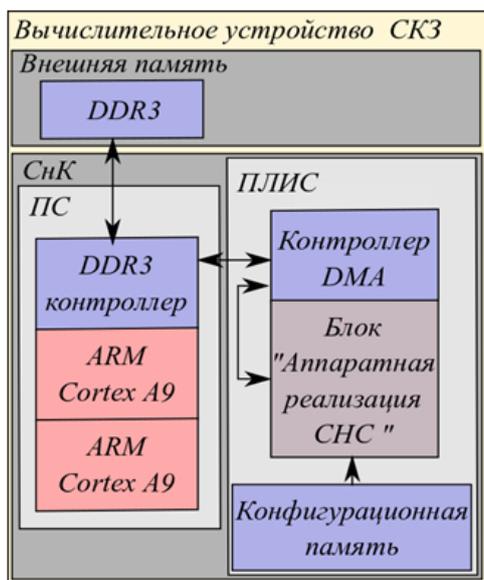


Рис. 2. Укрупненная функциональная схема ВУ
 Fig. 2. Enlarged functional diagram of the WU

Первый из них – СнК, он содержит процессорную систему (ПС) и ПЛИС. В состав ПС, в свою очередь, входят контроллер внешней памяти DDR3-типа и два процессорных ядра ARM Cortex A9 с тактовой частотой 800 МГц. На ПЛИС созданы контроллер прямого доступа к внешней памяти (Контроллер DMA), блок «Аппаратная реализация СНС» и конфигурационная память CONFIG sрасе, входящая в этот блок (на рис. 2 конфигурационная память для наглядности представления схемы ВУ вынесена за пределы блока). В блоке «Аппаратная реализация СНС» имеется также нейровычислительное устройство (НВУ), которое и выполняет вычисление процедур свертки и подвыборки для моделей СНС класса U-Net. НВУ содержит 64 универсальных вычислительных блока, которые одновременно работают над каждым слоем СНС. Вторым функциональным блоком ВУ является внешняя память DDR3 с эффективной частотой 800 МГц.

Предложен и реализован алгоритм работы НВУ, в соответствии с которым ведутся параллельные вычисления внутри каждого слоя модели СНС. Идея алгоритма заключается в том, что каждый универсальный вычислительный блок НВУ производит последовательное вычисление элементов назначенной ему выходной карты признаков в текущем слое. Получается, что каждая выходная карта признаков одновременно формируется отдельным таким блоком. Если выходных карт признаков слоя больше количества вычислительных блоков в НВУ, то один блок производит поочередное формирование нескольких выходных карт признаков слоя.

Для обеспечения взаимодействия ПЛИС, ПС и внешней памяти разработано программное обеспечение в виде многофункционального драйвера под ОС с ядром Linux. Драйвер написан на языке Си.

На языке C++ разработана библиотека для взаимодействия пользователя с ВУ при его настройках.

Для проведения быстрого реконфигурирования НВУ на ПЛИС и с целью упрощения аппаратной реализации различных моделей СНС был разработан программный инструмент на языке Python с процедурами (вставками) на языке Cython. Этот инструмент также дает возможность выбора варианта реализации модели СНС на процессорной системе (x86 или ARM) или на ПЛИС.

Результаты исследований моделей СНС и их обсуждение

Исследования программно-реализованных моделей СНС

Обученные модели СНС были применены для семантической сегментации изображения тестового участка пихтового древостоя. Совместный визуальный анализ изображения тестового участка (рис. 3, a), эталонной карты сегментации (рис. 3, b) и результирующих выходных карт сегментации этих моделей СНС (рис. 3, c, d) показал, что обе модели СНС способны воспроизводить границы между деревьями и успешно классифицировать значительную часть крон деревьев пихты. При этом основную сложность для обеих моделей СНС при сегментации составил класс деревьев «Отмирающие».

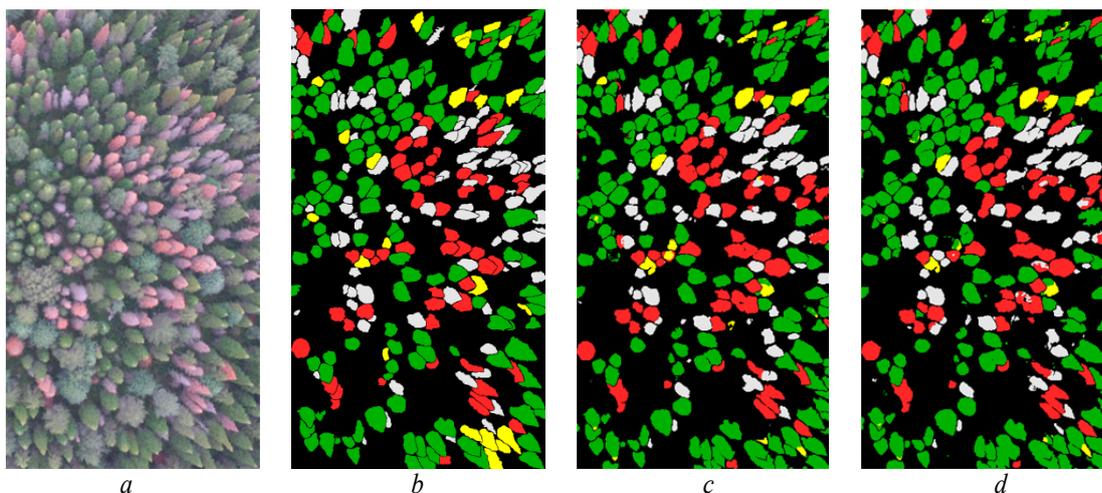


Рис. 3. Результаты сегментации деревьев пихты на изображении тестового участка: a) изображение тестового участка; b) эталонная карта сегментации; c) результат сегментации моделью U-Net; d) результат сегментации моделью U-Net с дилатационными свертками

Fig. 3. Results of segmentation of fir trees on the image of the test plot: a) image of the test plot; b) reference segmentation map; c) result of segmentation by the U-Net model; d) result of segmentation by the U-Net model with dilated convolutions

В табл. 1 приведены значения IoU для каждой из моделей СНС (в случае использования функции активации leakyReLU) и классов деревьев, а также значения $mIoU$ для оценки совокупного качества моделей СНС при сегментации изображения тестового участка. Обе модели СНС успешно сегментируют ($IoU > 0,5$) все классы деревьев пихты, кроме класса «Отмирающие». Низкое качество сегментации деревьев этого класса может быть объяснено его малой представленностью в обучающей выборке и визуальной схожестью с деревьями класса «Живые» и деревьями ели, входящими в класс «Фон».

Таблица 1. Точность сегментации изображений с помощью моделей СНС U-Net и U-Net с дилатационными свертками (U-Net*) для разных классов деревьев

Table 1. Accuracy of image segmentation using U-Net and U-Net CNN models with dilated convolutions (U-Net*) for different classes of trees

Модель Model	IoU					$mIoU$
	Фон Background	Живые Live	Отмирающие Dying	Свежий сухой Fresh deadwood	Старый сухой Old deadwood	
U-Net	0,87	0,77	0,39	0,75	0,62	0,68
U-Net*	0,86	0,76	0,37	0,75	0,62	0,67

Также было проведено сравнение вычислительной производительности программно-реализованных моделей СНС. Для этого оценивалось время выполнения каждой модели СНС при работе с фрагментом $256 \times 256 \times 3$. Результаты приведены в табл. 2, использование медианы и MAD обусловлено несимметричностью распределения времени обработки фрагмента. В таком эксперименте использовались реализации моделей СНС на языке программирования Python с применением библиотек tensorflow и keras [12]. Эксперимент проводился на персональном компьютере с процессором Intel Core i7-8700, 31 Гб ОЗУ и графическим ускорителем NVIDIA GeForce GTX 1080 Ti (3584 CUDA-ядер с частотой 1582 МГц, память GDDR5X с частотой 1372 МГц и шириной шины данных 352 бит). Энергопотребление графического ускорителя составляет 250 Вт.

Таблица 2. Скорость вычисления моделей СНС U-Net и U-Net с дилатационными свертками (U-Net*)

Table 2. Calculation speed of U-Net and U-Net CNN models with dilated convolutions (U-Net*)

Модель Model	Время обработки фрагмента, мс Fragment processing time, ms	
	Медиана/Median	MAD
U-Net	30,78	0,16
U-Net*	27,67	0,16

Сравнение результатов из табл. 2 показывает, что модель СНС U-Net с дилатационными свертками позволяет получить преимущество в скорости обработки одного фрагмента около 3 мс на персональном компьютере, давая при этом сопоставимое качество сегментации (табл. 1).

Исследование аппаратно-реализованных моделей СНС

Для исследования эффективности аппаратно-реализованных моделей СНС в составе ВУ проведены

эксперименты по семантической сегментации фрагментов изображения с БПЛА того же тестового участка деревьев пихты, что рассмотрено выше. Фрагменты изображения размером $256 \times 256 \times 3$ подавались на вход аппаратно-реализованных в ВУ моделей СНС U-Net и U-Net с дилатационными свертками.

Эксперименты с моделью СНС U-Net проводились с использованием 16-разрядных (float16) и 32-разрядных (float32) чисел с плавающей запятой и с разными функциями активации: eLU, ReLU и leakyReLU. Результаты для этой модели приведены в табл. 3.

Таблица 3. Значения точности семантической сегментации IoU фрагментов, полученные с помощью модели СНС U-Net

Table 3. Accuracy values of the semantic segmentation of IoU fragments obtained using the U-Net CNN model

Функция активации Function activation	Точность семантической сегментации IoU для различных классов деревьев пихты и фона IoU semantic segmentation accuracy for different classes of fir trees and backgrounds					$mIoU$
	Фон Background	Живые Live	Отмирающие Dying	Свежий сухой Fresh deadwood	Старый сухой Old deadwood	
float16						
eLU	0,84	0,72	0,39	0,74	0,63	0,66
ReLU	0,82	0,68	0,28	0,76	0,66	0,64
leakyReLU	0,84	0,72	0,36	0,77	0,65	0,67
float32						
eLU	0,84	0,72	0,39	0,74	0,63	0,66
ReLU	0,83	0,68	0,29	0,76	0,66	0,64
leakyReLU	0,84	0,72	0,36	0,77	0,65	0,67

Видим, что лучшие результаты по точности сегментации фрагментов получены при использовании модели СНС U-Net с функцией активации leakyReLU. Для дальнейших исследований эффективности аппаратно-реализованных моделей СНС использовались результаты обучения в виде весовых коэффициентов программно-реализованных моделей СНС только с этой функцией активации.

В табл. 4 приведены значения точности семантической сегментации фрагментов, полученные с помощью модели СНС U-Net с дилатационными свертками и функцией активации leakyReLU. Из табл. 3, 4 видно, что результаты по точности сегментации фрагментов при использовании в вычислениях 16-разрядных чисел с плавающей запятой не отличаются от результатов по точности сегментации при использовании 32-разрядных чисел с плавающей запятой.

Применение дилатационных сверток приводит к незначительной потере точности сегментации (примерно на 0,02 для усредненной метрики $mIoU$) относительно модели U-Net. Однако по сравнению с программно-реализованными моделями СНС, результаты для которых представлены в табл. 1, наблюдается уменьшение точности на 0,02–0,04 по метрике $mIoU$ в зависимости от используемой в модели функции активации. Отметим, что падение точности сегментации еще выше для некоторых классов деревьев пихты. Так, например, для класса «Отмирающие» разница составляет 0,03–0,06 по метрике IoU .

Таблица 4. Значения точности семантической сегментации IoU фрагментов, полученные с помощью модели СНС U-net с дилатационными свертками

Table 4. Accuracy values of the semantic segmentation of IoU fragments obtained using the U-net CNN model with dilated convolutions

Точность семантической сегментации IoU для различных классов деревьев пихты и фона IoU semantic segmentation accuracy for various classes of fir trees and background					
Фон Background	Живые Live	Отмирающие Dying	Свежий сухой Fresh deadwood	Старый сухой Old deadwood	mIoU
float16					
0,83	0,69	0,33	0,78	0,62	0,65
float32					
0,83	0,69	0,33	0,78	0,62	0,65

Исследование скорости вычисления моделей СНС при их аппаратной реализации на ПЛИС заключалось в изменении времени, затраченного на семантическую сегментацию одного фрагмента размером $256 \times 256 \times 3$ на ПЛИС. СНК позволяла работать ПЛИС с тактовой частотой 100 МГц и организовать использование в НВУ 64 универсальных вычислительных блоков. Результаты по среднему времени обработки по всем используемым в экспериментах фрагментам для случаев float32 и float16 приведены в табл. 5. Из нее следует, что модель СНС U-Net с дилатационными свертками позволяет анализировать на ПЛИС фрагмент изображения по времени более чем на 25 % быстрее, чем модель U-Net для каждого из случаев float32 и float16. Использование чисел float16 («IEEE float16») приводит к падению производительности ВУ из-за того, что слои upsample и последний слой CNN обрабатываются на ПС, которая не поддерживает числа с плавающей запятой половинной точности. Преобразование чисел из формата float32 в float16 требует дополнительных затрат, что является узким местом ВУ.

Скорость семантической сегментации фрагментов с помощью программно-реализованных моделей CNN (табл. 2) на персональном компьютере с процессором Intel Core i7-8700, 31 ГБ ОЗУ и графическим процессором NVIDIA GeForce GTX 1080 Ti практически в тысячу раз выше, чем скорость вычислений этих моделей CNN на ПЛИС. Это связано со значительно более высокой тактовой частотой процессора (3200 МГц) используемого компьютера, чем тактовая частота ПЛИС (100 МГц) и наличием более совершенной оперативной памяти, а также с наличием графического ускорителя.

Таблица 5. Скорость вычисления моделей СНС на ПЛИС

Table 5. Calculation speed of SNS models on FPGA

CNN model	Среднее время обработки фрагмента, с Average fragment processing time, sec			
	float32		float16	
	Медиана Median	MAD	Медиана Median	MAD
U-Net	31,12	0,07	38,03	0,12
U-Net*	23,47	0,03	27,58	0,05

В табл. 6 приведены результаты измерения энергопотребления СНК и требующихся ресурсов ПЛИС в виде количества различных ячеек при аппаратной реализации

каждой из моделей СНС U-Net в случаях использования чисел в форматах float32 и float16. Анализируя эти результаты, можно сказать, что применение чисел с плавающей запятой float16 позволяет уменьшить по сравнению со случаем чисел float32 энергопотребление СНК и требуемые ресурсы ПЛИС для хранения входных данных и весовых коэффициентов модели СНС. Энергопотребление СНК с ПЛИС составляет немного более 4–5 Вт, что в десятки раз меньше, чем у того же графического ускорителя NVIDIA GeForce GTX 1080 Ti, имеющего энергопотребление 250 Вт.

Таблица 6. Энергопотребление СНК и ресурсы ПЛИС

Table 6. Power consumption of SoC and FPGA resources

Энергопотребление и ресурсы ПЛИС Power consumption and FPGA resources				
Тип Type	Энергопотребление, Вт Energy consumption, W	LUT, шт/рс	LUTRAM, шт/рс	BRAM, шт/рс
float16	4,32	138917	2074	225,5
float32	5,33	165278	871	328,5

Все эти результаты еще раз указывают на то, что основным преимуществом устройств на основе современных СНК с ПЛИС является низкое энергопотребление при довольно высокой производительности. Более того, современные СНК обладают весьма малой массой по сравнению с теми же графическими ускорителями.

Отметим, что все полученные результаты исследования аппаратно-реализованных на ПЛИС двух моделей СНС класса U-Net являются весьма важными для разработчиков интеллектуальных СКЗ в составе мобильных систем мониторинга на основе БПЛА. Разработчики таких СКЗ должны соблюдать баланс между быстродействием используемого ВУ, точностью распознавания объектов земной поверхности с помощью СНС, массой и энергопотреблением этого устройства с целью увеличения времени полёта БПЛА. При поиске баланса им достаточно проанализировать приведенные результаты исследований и принять соответствующие проектные решения. Например, если создается СКЗ с относительно невысокой точностью сегментации изображений объектов мониторинга, то можно использовать модель СНС U-Net с дилатационными свертками и проводить вычисления этой модели на ПЛИС при работе с 16-разрядными числами с плавающей запятой. Это позволит значительно снизить энергопотребление и требования к ресурсам вычисляемой ПЛИС.

Заключение

Сегодня актуальными являются задачи создания интеллектуальных СКЗ в составе мобильных систем мониторинга объектов земной поверхности на основе БПЛА. В данной работе с целью последующей разработки такой СКЗ предложены, программно-реализованы и обучены две новые модели СНС класса U-net. Затем эти модели СНС аппаратно-реализованы на ПЛИС современной СНК Zynq 7000 (Kintex FPGA). Для обучения, верификации и исследования моделей СНС использовался датасет, созданный по снимкам с БПЛА деревьев пихты сибирской, поврежденных уссурийским полиграфом.

Проведены исследования этих программно-реализованных моделей СНС и ВУ, созданного на основе этой СНК. Выяв-

лено, что по сравнению с программно-реализованными моделями СНС для аппаратно-реализованных моделей СНС на ПЛИС наблюдается уменьшение точности сегментации изображений на 0,02–0,04 по метрике mIoU. Падение точности сегментации еще выше для некоторых классов деревьев пихты; так, для класса «Отмирающие» разница составляет 0,03–0,06 по метрике IoU. Применение дилатационных сверток в модели СНС приводит к незначительной потере точности сегментации (примерно на 0,02 для метрики mIoU). Результаты по точности сегментации при использовании в вычислениях на ПЛИС 16-разрядных чисел с плавающей запятой не отличаются от результатов по точности сегментации при использовании 32-разрядных чисел с плавающей запятой.

При исследовании производительности ВУ выявлено, что модель СНС U-Net с дилатационными свертками позволяет анализировать на ПЛИС фрагмент изображения размером 256×256×3 более чем на 25 % быстрее, чем исходная модель U-Net. Однако скорость вычисления этих моделей СНС в ВУ почти на три порядка меньше по сравнению со скоростью выполнения программно-реализованных моделей СНС на персональном компьютере с процессором Intel Core i7-8700, 31 Гб ОЗУ и с графическим ускорителем NVIDIA GeForce GTX 1080 Ti.

СПИСОК ЛИТЕРАТУРЫ

1. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation // arXiv preprint arXiv: 1505.04597v1. – 2015. URL: <https://doi.org/10.48550/arXiv.1505.04597> (дата обращения: 10.03.2023).
2. Bochkovskiy A., Wang C.Y., Mark Liao H.Y. YOLOv4: optimal speed and accuracy of object detection // arXiv preprint arXiv:2004.10934v1. – 2020. URL: <https://doi.org/10.48550/arXiv.2004.10934> (дата обращения: 10.03.2023).
3. Low-power neural networks for semantic segmentation of satellite images / G. Bahl, L. Daniel, M. Moretti, F. Lafarge // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). – Seoul, Korea (South), 2019. – P. 2469–2476. DOI: 10.1109/ICCVW.2019.00302
4. An FPGA-based CNN accelerator integrating depthwise separable convolution / L. Bing, Z. Danyin, F. Lei, F. Shou, F. Ping, L. Junbao // Electronics. – 2019. – V. 8. – № 3. – P. 281. URL: <https://doi.org/10.3390/electronics8030281> (дата обращения: 10.03.2023).
5. Liu S., Luk W. Towards an efficient accelerator for DNN-based remote sensing image segmentation on FPGAs // 29th International Conference on Field Programmable Logic and Applications (FPL). – Barcelona, Spain, 2019. – P. 187–193. DOI: 10.1109/FPL.2019.00037
6. EDSSA: an encoder-decoder semantic segmentation networks accelerator on OpenCL-based FPGA platform / H. Hongzhi, W. Yakun, Y. Mengqi, S. Xuesong, Q. Fei, L. Li, Qi Wei, L. Xinjun // Sensors. – 2020. – V. 20. – № 14. – P. 3969. URL: <https://doi.org/10.3390/s20143969> (дата обращения: 10.03.2023).
7. Optimizing FPGA-based convolutional encoder-decoder architecture for semantic segmentation / Yu. Mengqi, H. Hongzhi, L. Hong, H. Shuyi, Q. Fei, L. Li, X. Fugui, L. Xin-Jun, Y. Huazhong // 2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). – Suzhou, China, 2019. – P. 1436–1440. DOI: 10.1109/CYBER46603.2019.9066759
8. Optimizing CNN-based segmentation with deeply customized convolutional and deconvolutional architectures on FPGA / S. Liu, H. Fan, X. Niu, H.-C. Ng, Y. Chu, W. Luk // ACM Transactions on Reconfigurable Technology and Systems. – 2018. – V. 11. – Iss. 3. – Article no. 19. – P. 1–22. URL: <https://doi.org/10.1145/3242900> (дата обращения: 10.03.2023).

Полученные результаты измерения энергопотребления СНК и требующихся ресурсов ПЛИС при аппаратной реализации моделей СНС класса U-Net указывают на то, что применение 16-разрядных чисел с плавающей запятой при проведении вычислений моделей СНС позволяет уменьшить энергопотребление СНК и требуемые ресурсы ПЛИС по сравнению со случаем применения 32-разрядных чисел. Энергопотребление СНК с ПЛИС составляет немногим более 5 Вт, что в десятки раз меньше, чем у того же графического ускорителя NVIDIA GeForce GTX 1080 Ti, имеющего энергопотребление 250 Вт.

Можно считать, что полученные результаты исследования предложенных и программно- и аппаратно-реализованных на ПЛИС моделей СНС класса U-Net имеют важное научное и практическое значение для разработчиков интеллектуальных СКЗ в составе мобильных систем мониторинга объектов земной поверхности на основе БПЛА. Использование этих результатов позволит им принимать обоснованные проектные решения при создании интеллектуальных СКЗ для решения различных прикладных задач.

Исследование выполнено при финансовой поддержке РФФИ и Томской области, проект № 18-47-700010.

9. ATT squeeze U-Net: a lightweight network for forest fire detection and recognition / J. Zhang, H. Zhu, P. Wang, X. Ling // IEEE Access. – 2021. – V. 9. – P. 10858–10870. DOI: 10.1109/ACCESS.2021.3050628
10. Fisher Yu., Koltun V. Multi scale context aggregation by dilated convolutions // arXiv preprint arXiv: 1511.07122v3. – 2016. URL: <https://doi.org/10.48550/arXiv.1511.07122> (дата обращения: 10.03.2023).
11. Taha A.A., Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool // BMC Medical Imaging. – 2015. – V. 15. – Article number 29. URL: <https://doi.org/10.1186/s12880-015-0068-x> (дата обращения: 10.03.2023).
12. Keras. URL: <https://keras.io/> (дата обращения 05.10.2021).
13. Семантическая сегментация поврежденных деревьев пихты на снимках с беспилотных летательных аппаратов / И.А. Керчев, К.А. Маслов, Н.Г. Марков, О.С. Токарева // Современные проблемы дистанционного зондирования Земли из космоса. – 2021. – Т. 18. – № 1. – С. 116–126. DOI: 10.21046/2070-7401-2021-18-1-116-126. EDN: FXPNTX
14. Xilinx. URL: xilinx.com/products/silicon-devices/soc/zynq-7000.html (дата обращения 05.10.2021).
15. Аппаратная реализация на ПЛИС сверточных нейронных сетей для распознавания объектов на изображениях / И.В. Зоев, Н.Г. Марков, А.П. Береснев, Т.А. Ягунов // ГрафиКон 2018: 28-я Международная конференция по компьютерной графике и машинному зрению. – Томск, 2018. – Т. 2. – С. 200–203. URL: <https://www.graphicon.ru/html/2018/papers/200-203.pdf> (дата обращения 05.10.2021).
16. Зоев И.В., Марков Н.Г., Рыжова С.Е. Интеллектуальная система компьютерного зрения беспилотных летательных аппаратов для мониторинга технологических объектов предприятий нефтегазовой отрасли // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2019. – Т. 330. – № 11. – С. 34–49. DOI: 10.18799/24131830/2019/11/2346
17. Zoev I.V., Markov N.G., Ryzhova S.E. The monitoring system of hazardous technological objects based unmanned aerial vehicles // E3S Web of Conferences. – 2020. – V. 223. – № 03005. DOI: 10.1051/e3sconf/202022303005

*Поступила 10.03.2023 г.
Прошла рецензирование 25.04.2023*

Информация об авторах

Зоев И.В., программист АО «ИнфоТеКС».

Маслов К.А., программист отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

Марков Н.Г., доктор технических наук, профессор отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

Мыцко Е.А., кандидат технических наук, доцент отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

UDC 004.032.26

STUDY OF HARDWARE-IMPLEMENTED CONVOLUTIONAL NEURAL NETWORKS OF THE U-NET CLASS

Ivan V. Zoev¹,
ivz3@tpu.ru

Konstantin A. Maslov²,
koheg@mail.ru; orpheus.const@outlook.com

Nikolay G. Markov²,
orloff234@gmail.ru; markovng@tpu.ru

Evgeny A. Mytsko²,
orloff234@gmail.ru; evgenvt@tpu.ru

¹ Joint-Stock Company «InfoTeCS»,
51, Kirov avenue, Tomsk, 634041, Russia.

² National Research Tomsk Polytechnic University,
30, Lenin avenue, Tomsk, 634050, Russia.

The authors have developed and implemented two convolutional neural networks of the U-Net class: a modification of the classical U-Net and a U-Net with dilated convolutions. For training and testing convolutional neural networks, data sets were used based on images from an unmanned aerial vehicle of fir trees damaged by the Ussuri polygraph. Depending on the degree of damage, the images contain trees of four classes and a background. The weights obtained during training for each of convolutional neural network were then used in the hardware implementation of the convolutional neural networks on a programmable logic integrated circuit of the Xilinx Zynq 7000 (Kintex FPGA) system-on-chip. The paper introduces the results of the study of segmentation accuracy and performance of each convolutional neural network implemented on programmable logic integrated circuits.

Key words: convolutional neural networks, programmable logic integrated circuit, hardware-implemented convolutional neural networks, U-Net.

The research was financially supported by the RFBR and Tomsk region, project no. 18-47-700010.

REFERENCES

- Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. *arXiv preprint arXiv: 1505.04597v1*. 2015. Available at: <https://doi.org/10.48550/arXiv.1505.04597> (accessed: 10 March 2023).
- Bochkovskiy A., Wang C.Y., Mark Liao H.Y. YOLOv4: optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934v1*. 2020. Available at: <https://doi.org/10.48550/arXiv.2004.10934> (accessed: 10 March 2023).
- Bahl G., Daniel L., Moretti M., Lafarge F. Low-power neural networks for semantic segmentation of satellite images. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), 2019. pp. 2469–2476. DOI: 10.1109/ICCVW.2019.00302
- Bing L., Danyin Z., Lei F., Shou F., Ping F., Junbao L. An FPGA-based CNN accelerator integrating depthwise separable convolution. *Electronics*, 2019, vol. 8, no. 3, article no. 281. Available at: <https://doi.org/10.3390/electronics8030281> (accessed: 10 March 2023).
- Liu S., Luk W. Towards an efficient accelerator for DNN-based remote sensing image segmentation on FPGAs. *29th International Conference on Field Programmable Logic and Applications (FPL)*. Barcelona, Spain, 2019. pp. 187–193. DOI: 10.1109/FPL.2019.00037
- Hongzhi H., Yakun W., Mengqi Y., Xuesong S., Fei Q., Li L., Qi Wei, Xinjun L. EDSSA: an encoder-decoder semantic segmentation networks accelerator on OpenCL-based FPGA platform. *Sensors*, 2020, vol. 20, no. 14, article no. 3969. Available at: <https://doi.org/10.3390/s20143969> (accessed: 10 March 2023).
- Mengqi Yu., Hongzhi H., Hong L., Shuyi H., Fei Q., Li L., Fugui X., Xin-Jun L., Huazhong Y. Optimizing FPGA-based convolutional encoder-decoder architecture for semantic segmentation. *2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*. Suzhou, China. 2019. pp. 1436–1440. DOI: 10.1109/CYBER46603.2019.9066759
- Liu S., Fan H., Niu X., Ng H-C., Chu Y., Luk W. Optimizing CNN-based segmentation with deeply customized convolutional and deconvolutional architectures on FPGA. *ACM Transactions on Reconfigurable Technology and Systems*, 2018, vol. 11, Iss. 3, article no. 19, pp. 1–22. Available at: <https://doi.org/10.1145/3242900> (accessed: 10 March 2023).
- Zhang J., Zhu H., Wang P., Ling X. ATT squeeze U-Net: a lightweight network for forest fire detection and recognition. *IEEE Access*, 2021, vol. 9, pp. 10858–10870. DOI: 10.1109/ACCESS.2021.3050628
- Fisher Yu., Koltun V. Multi scale context aggregation by dilated convolutions. *arXiv preprint arXiv: 1511.07122v3*. 2016. Available at: <https://doi.org/10.48550/arXiv.1511.07122> (accessed: 10 March 2023).
- Taha A.A., Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 2015, vol. 15, Article no. 29. Available at: <https://doi.org/10.1186/s12880-015-0068-x> (accessed: 10 March 2023).
- Keras. Available at: <https://keras.io/> (accessed 5 October 2021).
- Kerchev I.A., Maslov K.A., Maslov N.G., Tokareva O.S. Semantic segmentation of damaged fir trees in unmanned aerial vehicle images. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2021, vol. 18, no. 1, pp. 116–126. In Rus. DOI: 10.21046/2070-7401-2021-18-1-116-126. EDN: FXPNTX.
- Xilinx. Available at: [xilinx.com/products/silicon-devices/soc/zynq-7000.html](https://www.xilinx.com/products/silicon-devices/soc/zynq-7000.html) (accessed 5 October 2021).
- Zoev I.V., Markov N.G., Beresnev A.P., Yagunov T.A. FPGA-based hardware implementation of convolutional neural networks for images recognition. *GraphiCon 2018. 28th International Conference on Computer Graphics and Vision on Proceedings*. 2018, vol. 2, pp. 200–203.
- Zoev I.V., Markov N.G., Ryzhova S.E. Intelligent computer vision system for unmanned aerial vehicles for monitoring technological objects of oil and gas industry. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2019, vol. 330, no. 11, pp. 34–49. In Rus. DOI: 10.18799/24131830/2019/11/2346
- Zoev I.V., Markov N.G., Ryzhova S.E. The monitoring system of hazardous technological objects based unmanned aerial vehicles. *E3S Web of Conferences*, 2020, vol. 223, no. 03005. DOI: 10.1051/e3sconf/202022303005.

Received: 10 March 2023.

Reviewed 25 April 2023.

Information about the authors

Ivan V. Zoev, programmer, Joint-Stock Company «InfoTeCS».

Konstantin A. Maslov, programmer, National Research Tomsk Polytechnic University.

Nikolay G. Markov, Dr. Sc., professor, National Research Tomsk Polytechnic University.

Evgeny A. Mytsko, Cand. Sc., associate professor, National Research Tomsk Polytechnic University.