

УДК 004.93'1

РАСПОЗНАВАНИЕ СЕЛЬСКОХОЗЯЙСТВЕННОЙ РАСТИТЕЛЬНОСТИ НА ИЗОБРАЖЕНИЯХ ЗЕМНОЙ ПОВЕРХНОСТИ НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ U-NET



В.В. Ганченко
Старший научный
сотрудник ОИПИ НАН
Беларуси, кандидат
технических наук



А.А. Дудкин
Заведующий лабораторией
идентификации систем
ОИПИ НАН Беларуси, доктор
технических наук, профессор



С.В. Шелег
Инженер-
программист
ООО «БСВТ – новые
технологии»,
магистрант БГУИР

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь.
Объединенный институт проблем информатики НАН Беларуси, Республика Беларусь.
ООО «БСВТ – новые технологии».
E-mail: ganchenko@lsi.bas-net.by, doudkin@newman.bas-net.by, sergey.sheleg@hotmail.com

С.В. Шелег

Окончил Белорусский Государственный Университет Информатики и Радиоэлектроники, факультет компьютерных систем и сетей. Магистрант БГУИР КСиС. Работает в ООО «БСВТ – новые технологии» в должности инженера-программиста. Проводит научные исследования в области цифровой обработки изображений.

А.А. Дудкин

Заведующий лабораторией идентификации систем Объединенного института проблем информатики НАН Беларуси, д.т.н., проф. Проводит научные исследования в области цифровой обработки изображений, распознавания образов, систем технического зрения.

В.В. Ганченко

Старший научный сотрудник лаборатории идентификации систем Объединенного института проблем информатики НАН Беларуси, д.т.н., проф. Проводит научные исследования в области цифровой обработки изображений, распознавания.

Аннотация. В данной работе рассматривается задача сегментации картофельных полей по данным аэрофотосъемки различного пространственного разрешения. В качестве основы алгоритма распознавания используется сверточная нейронная сеть с архитектурой U-Net. Предложенный алгоритм позволяет выполнить разделение пикселей мультиспектрального изображения на два класса: «растительность» и «почва». Произведён сравнительный анализ предложенного алгоритма с другими нейросетевыми алгоритмами.

Ключевые слова: сверточная нейронная сеть, семантическая сегментация, мультиспектральное изображение, сельскохозяйственная растительность, распознавание.

Введение.

Семантическая сегментация изображений – задача разделения изображения на части по некоторому заданному предикату. Пусть $f(x, y)$ – дискретная функция, описывающая анализируемое изображение; X – конечное множество точек плоскости, на котором определена функция $f(x, y)$; $S = \{S_1, S_2, \dots, S_k\}$ – разбиение X на k непустых связных областей S_i , где $i = 1, 2, \dots, k$; L_p – предикат, заданный на множестве S и принимающий истинные значения тогда и только тогда, когда любая пара точек (x, y) из каждого подмножества S_i удовлетворяет некоторому

критерию однородности этого подмножества.

Сегментацией изображения $f(x, y)$ по предикату L_p называется разбиение $S = \{S_1, S_2, \dots, S_k\}$, удовлетворяющее условиям:

- а) $\bigcup_{i=1}^k S_i = X$ тривиально;
- б) $S_i \cap S_j = \emptyset$ для любых $i \neq j$;
- в) $L_p(S_i)$ принимает истинные значения для любого i ;
- г) $L_p(S_i \cap S_j)$ принимает ложные значения для любых $i \neq j$.

Отметим, что k областей, полученных в результате сегментации часто различных классов, группируются в m классов, где $2 \leq m \leq k$.

Для решения задачи семантической сегментации изображений, наряду с «классическими» методами (областно-ориентированные, когда непосредственно строятся S_i и гранично-ориентированные, когда определяются границы между S_i) [1], широкое применение находят искусственные нейронные сети. Последние не требуют явного задания предиката L_p , он автоматически строится в процессе обучения нейронной сети. На данный момент наибольшую эффективность при решении задач сегментации показывают свёрточные нейронные сети (СНС) [2]. Высокое качество сегментации обеспечивает технология глубокого обучения [3], применяемая в разработке СНС.

Семантическая сегментация является составной частью решения задач распознавания сельскохозяйственной растительности различного типа [4-13]. Выбор той или иной архитектуры СНС зависит от характеристик исходных изображений (размерности и разрешающей способности, в первую очередь) и сложности задачи семантической сегментации (типа предиката и количества классов разбиений). По результатам анализа следует отметить перспективность применения для двухклассового распознавания СНС U-Net. Она имеет простую архитектуру и, как следствие, малое потребление ресурсов в работе и при обучении – даже при небольшой обучающей выборке, достигаются приемлемые по качеству результаты.

Описание проблемы.

Материалами для исследований явились фотографии экспериментального участка поля, засаженного картофелем, выполненные с высоты 5, 15, 50 и 100 метров (827 снимков разрешением 3474×2314 пикселей). Съёмка растений проводилась ежедневно в 8, 10, 12, 14 и 16 часов на протяжении 8 дней июля [14]. Наблюдению подвергались 3 группы растений: зараженные инфекционной болезнью *alternaria* и бактериальной болезнью *erwinia*, а также здоровые растения. В результате развития указанных заболеваний происходит разрушение хлорофила в листьях картофеля, что приводит к изменению цвета растения. Анализ цветовых характеристик различных типов растительности на базе их гистограмм показывает заметное их различие для почвы и растительности (для всех цветовых каналов) независимо от того, здоровое и больное растение. Данная информация может быть учтена при использовании СНС U-Net для семантической сегментации.

Анализ состояния использования архитектуры U-Net для сегментации показывает преимущества сети архитектуры [15]. Вместе с тем, наибольшее количество ошибок возникало на участках, соответствующих границе здоровой растительности и почвы (в особенности в местах, где небольшие участки почвы окружены растительностью, создающей на этом участке почвы тень). Заметим, что использование комбинированных алгоритмов (комбинирование архитектур в разного типа в рамках обгной нейронной сети или использование ансамблей нейронных сетей) позволяет снизить искажение цветовых характеристик изображений окрестностей пикселей при попадании в него растительности и почвы.

В настоящей статье предлагается вариант реализации сегментатора на основе U-Net, который позволяет повысить точность сегментации в целом в основном за счет обучения. Различие

в архитектуре обусловлены разными обучающими выборками и различными способами обучения (используются разные функции потерь).

Нейронная сеть, описанная в [15], принимает на вход матрицу размером $256 \times 256 \times 3$, выходом является матрица $256 \times 256 \times 4$, где каждая из 4 подматриц размером 256×256 показывает степень принадлежности пикселя к одному из искомых классов. Сама сеть содержит 11 сверточных слоев помимо выходного, при этом ресемплирование осуществляется вплоть до разрешения 32×32 . Максимальное количество фильтров в сверточном слое – 256.

Обучающая выборка была получена путем «нарезки» имеющихся аэрофотоснимков с размеченными участками. При этом нарезались участки размером 32, 64 и 128 пикселей с перекрытием и добавлением поворотов на углы, кратные 90° . Для обучения формируются пары (растительность – почва), (здоровая растительность- больная растительность). Размер обучающей выборки: 110528 изображений. Размер валидационной выборки: 27632 изображений (20 % общей базы). Реализация на основе библиотеки Keras.

Архитектура предлагаемой СНС.

Сеть была настроена на входной тензор размерностью $512 \times 512 \times 3$. Выходом такой сети является тензор $512 \times 512 \times 1$. Архитектура U-Net представляет собой энкодер-декодер. В основе как энкодера так и декодера лежит следующий алгоритм: последовательное применение свёртки, нормализации и функции активации. Данные операции проводятся с целью изменения разрешения карты признаков. Так, каждый уровень энкодера понижает ширину и высоту карты (с помощью слоя MaxPool2d), но увеличивает глубину. В свою очередь, декодер увеличивает ширину и высоту карты, но уменьшает глубину (за счёт предварительного применением слоя ConvTranspose2d и объединения карты с соответствующим уровнем энкодера). Описание параметров слоев сети приведено в таблице 1. Через BaseBlock обозначены последовательные применения свёртки, нормализации и функции активации. Общее число обучаемых параметров: 7763041. Общее расчётное потребление памяти: 1648,61 МБ.

Таблица 1. Параметры СНС

Название слоя	Размер тензора на выходе слоя	Количество обучаемых параметров слоя
BaseBlock(encoder1)	[32, 512, 512]	10208
MaxPool2d	[32, 256, 256]	0
BaseBlock(encoder2)	[64, 256, 256]	55552
MaxPool2d	[64, 128, 128]	0
BaseBlock(encoder3)	[128, 128, 128]	221696
MaxPool2d	[128, 64, 64]	0
BaseBlock(encoder4)	[256, 64, 64]	885760
MaxPool2d	[256, 32, 32]	0
BaseBlock(bottleneck)	[512, 32, 32]	3540992
ConvTranspose2d	[256, 64, 64]	524544
BaseBlock(deocder4)	[256, 64, 64]	1770496
ConvTranspose2d	[128, 128, 128]	131200
BaseBlock(deocder3)	[128, 128, 128]	442880
ConvTranspose2d	[64, 256, 256]	32832
BaseBlock(deocder2)	[64, 256, 256]	110848
ConvTranspose2d	[32, 512, 512]	8224
BaseBlock(deocder1)	[32, 512, 512]	27776
Conv2d	[1, 512, 512]	33

В качестве loss-функции используется DiceLoss:

$$\text{Dice}(X, Y) = \frac{2|X \cap Y| + \text{smooth}}{|X| + |Y| + \text{smooth}},$$

где X является предсказанием,
 Y - правильно размеченной маской на текущем объекте,
 $|\bullet|$ означает мощность множества (количество элементов в этом множестве),
 \cap - операция пересечения множеств,
 smooth – плавный коэффициент.

Например, предположим, что предсказание на пикселе (0, 0) равно 0,567, а цель равна 1, получаем $0,567 \times 1 = 0,567$. Если цель равна 0, мы получаем 0 в этой позиции пикселя.

Также использован плавный коэффициент 1 для обратного распространения. Если предсказание является жестким порогом, равным 0 и 1, трудно обратно распространять функцию ошибки.

В качестве метода оптимизации используется Адам [16]. Алгоритм оптимизации Адам является расширением стохастического градиентного спуска, который в последнее время получил широкое распространение для приложений глубокого обучения в области компьютерного зрения и обработки естественного языка. Алгоритм используется со следующими параметрами:

$\text{lr} = 0.0001$ – коэффициент скорости обучения;

$\text{beta1} = 0.9$, $\text{beta2} = 0.999$ – коэффициенты, используемые для расчёта скользящего среднего градиентов и их квадратов;

$\text{eps} = 1e-8$ – слагаемое, добавляемое к знаменателю для улучшения числовой стабильности;

$\text{weight_decay} = 0$ – L2-регуляризация.

Нейронная сеть была реализована на языке Python с использованием фреймворка PyTorch.

Входными данными для обучения являются пары, состоящие из оригинальных изображений и масок, где белым закрашена область, которую необходимо научиться сегментировать. Пример входных данных показан на рисунке.

При обучении датасет был разбит на 2 множества: для обучения (90% от общего числа изображений) и валидации (10%). Так же в процессе обучения на каждой итерации использовались следующие преобразования изображений с целью аугментации датасета:

– случайное изменение масштаба от 90% до 110%;

– случайный поворот изображения от -75 до +75 градусов;

– случайное отражение изображения по горизонтали.

Аппаратное обеспечение, на котором производилось обучение позволило использовать максимальный размер батча, равный 4, для одной итерации. Длительность обучения составила 51 эпоху (2958 итераций). Таким образом можно сказать, что сеть обучалась на не более чем 11832 изображениях (если допустить, что в процессе случайной аугментации не было повторов).

Оценка точности алгоритма.

В качестве метрики для оценки точности работы алгоритма был выбран Intersection-Over-Union (IoU), рассчитываемый по формуле:

$$\text{IoU} = \frac{S_i}{S_u},$$

где S_i – площадь пересечения результата сегментации и истинной области,

S_u – площадь объединения результата сегментации и истинной области.

Точность при обучении(валидации) составила 98.3%. Точность при тестировании – 97.2%. Тестирование проводилось на общей базе имеющихся оригинальных изображений.

Как видно, прирост точности второй реализации на валидации составил 1,4% (98,3% против 96,9%) несмотря на то, что вторая реализация обучалась на меньшем общем количестве изображений.

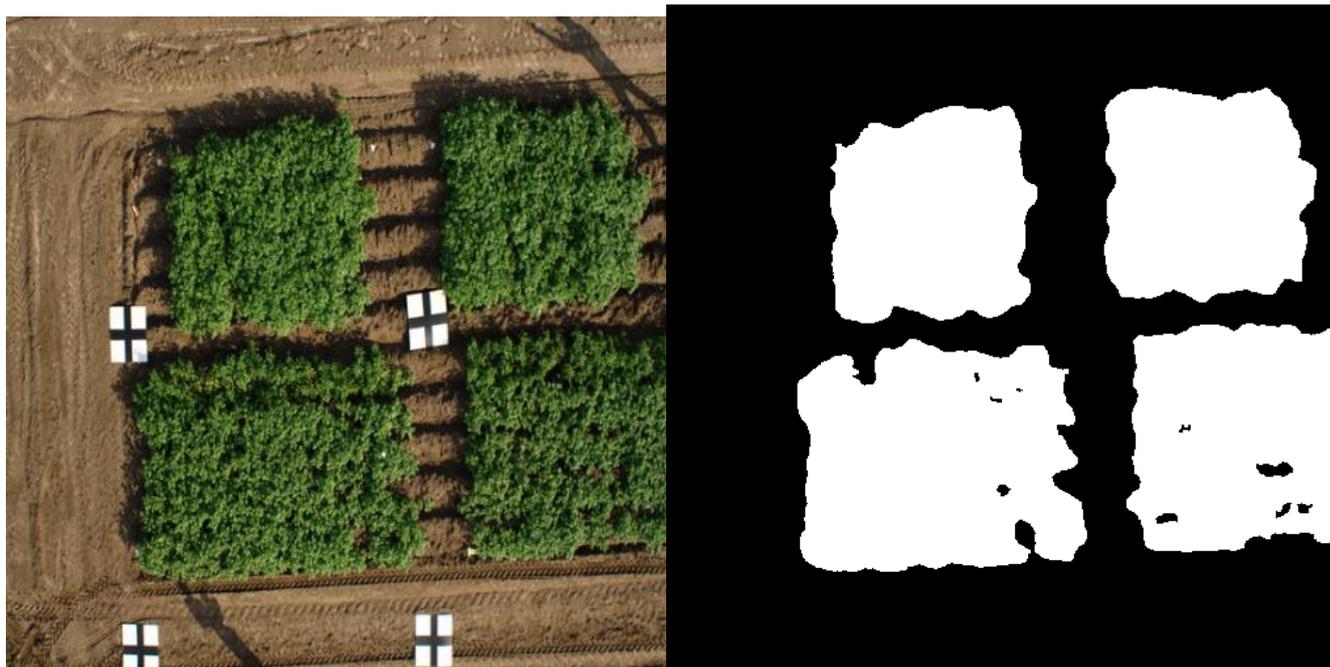


Рисунок 1. Пример данных для обучения нейронной сети: слева – оригинальное изображение, справа – маска сегментации.

Ключевыми отличиями, повлиявшие на результат, являются: размер входного слоя (512×512 в предлагаемой сети против 256×256 у сети [15]) и глубина нейронной сети (4 пары «энкодер-декодер» против 3 пар у [15]). Они прямо влияют на количество обучаемых параметров и, как следствие, качество сегментации. Из негативных последствий таких изменений можно отметить большее потребление ресурсов.

Так же на результат могли повлиять следующие отличия в процессе обучения (не исследовались):

- функция потерь;
- метод оптимизации;
- способ аугментации датасета;
- размер обучающей выборки;
- инициализация начальных значений параметров.

Сравнение разработанного с алгоритма с другими алгоритмами дано в таблице 2

Таблица 2. Сравнение методов сегментации

Метод	Точность
Custom [17]	96.76%
SegNet [18]	96.54%
DenseNet [19]	96.89%
FCDenseNet [20]	97.23%
Предложенная архитектура	98.3%

Заключение.

Предложенная нейронная сеть на основе архитектуры U-Net для семантической сегментации изображений позволяет выполнить разделение изображений картофельных полей на два класса (растительность – почва). Ключевыми отличиями от сети-прототипа [15], позволившими повысить точность идентификации почвы, являются увеличенный размер входного слоя и глубина нейронной сети, которые непосредственно влияют на количество обучаемых параметров и, как следствие, качество сегментации.

Результат данной работы могут иметь несколько полезных применений для повышения

эффективности распознавания сельскохозяйственной растительности за счет построения новых комбинированных нейросетевых алгоритмов классификации.

Во-первых, это разработка уникальных архитектур нейронных сетей. В этом случае разработанная сеть может реализовывать постобработку (корректировку) результата основной сети. В качестве последней, в частности, может быть и сеть [15]. Либо результаты работы предложенной в докладе сети можно использовать как дополнительные входные данные для создания новой сети, решающей задачу многоклассового распознавания.

Во-вторых, предложенная сеть может быть компонентой ансамбля нейронных сетей, в которых решение принимается на основании усредненных результатов работы составляющих ансамбль сетей.

Работа выполнена при поддержке БРФФИ (проекты Ф20КИ-017 и Ф20РА-014).

Список литературы

- [1] Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М.: Техносфера, 2005. – 1072 с.
- [2] Сикорский, О.С. Обзор свёрточных нейронных сетей для задачи классификации изображений / О. С. Сикорский // Новые информационные технологии в автоматизированных системах – Москва, 2017. – № 20. – С. 37–42.
- [3] Николенко С., Кадури А., Архангельская Е. Глубокое обучение – СПб.: Питер, 2018. – 480 с.
- [4] R. R. Atole, D. Park, A Multiclass Deep Convolutional Neural Network Classifier for Detection of Common Rice Plant Anomalies / (IJACSA) International Journal of Advanced Computer Science and Applications. – Vol. 9, No. 1. – 2018. – P. 67-70.
- [5] R.Rajmohan, M.Pajany, R.Rajesh, D.Raghu Raman, U. Prabu, Smart paddy crop disease identification and management using deep convolution neural network and SVM classifier / International Journal of Pure and Applied Mathematics. – Volume 118 No. 15. – 2018. – P. 255-264.
- [6] G. Athanikar1, P. Badar, Potato Leaf Diseases Detection and Classification System / International Journal of Computer Science and Mobile Computing. – Vol.5 Issue.2. – 2016. – P. 76-88.
- [7] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification / Computational Intelligence and Neuroscience. – 2016. – 11 p.
- [8] Abdullahi, H.S. Advances of image processing in Precision Agriculture: Using deep learning convolution neural network for soil nutrient classification / Halimatu Sadiyah. Abdullahi, Oba Mustpha Zubair // Journal of Multidisciplinary Engineering Science and Technology (JMEST) – Vol. 4 Issue 8, August – 2017. – P 7981-7987.
- [9] Huang H., Deng J., Lan Y., Yang A., Deng X., Zhang L. A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery / PLoS ONE 13(4): e0196302. – 2018.
- [10] Inkyu Sa, weedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming / Inkyu Sa, Zetao Chen, Marija Popovic, Raghav Khanna, Frank Liebisch, Juan Nieto, Roland Siegwa // IEEE Robotics and Automation Letters. – 2018 – Vol. 3(1). – P. 588-595.
- [11] Potena C., Nardi D., Pretto A. Fast and Accurate Crop and Weed Identification with Summarized Train Sets for Precision Agriculture / IAS 2016: Intelligent Autonomous Systems 14. – 2017. – P 105-121
- [12] Yao, C., Zhang, Y., Zhang, Y., and Liu, H.: Application of convolutional neural network in classification of high resolution agricultural remote sensing images, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W7. – 2017. – P. 989-992.
- [13] R. R. Atole, D. Park, A Multiclass Deep Convolutional Neural Network Classifier for Detection of Common Rice Plant Anomalies / (IJACSA) International Journal of Advanced Computer Science and Applications. – Vol. 9, No. 1. – 2018. – P. 67-70.
- [14] Монография 14 года
- [15] Ganchenko, V. Image Semantic Segmentation Based on Convolutional Neural Networks for Monitoring Agricultural Vegetation / V. Ganchenko, A. Doudkin // Communications in Computer and Information Science, Springer, 2019. – Ch. 5. – V. 1055. – 2019. – P. 52-63.
- [16] Adam: A Method for Stochastic Optimization [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1412.6980>
- [17] Fig Plant Segmentation from Aerial Images Using a Deep Convolutional Encoder-Decoder Network [Электронный ресурс]. – Режим доступа:

https://www.researchgate.net/publication/333131201_Fig_Plant_Segmentation_from_Aerial_Images_Using_a_Deep_Convolutional_Encoder-Decoder_Network

[18] SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [Электронный ресурс]. – Режим доступа: <https://ieeexplore.ieee.org/document/7803544/>

[19] Densely Connected Convolutional Networks [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1608.06993.pdf>

[20] Pyramid Scene Parsing Network [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1612.01105.pdf>

AGRICULTURAL VEGETATION RECOGNITION ON EARTH'S SURFACE IMAGES USING CONVOLUTIONAL NEURAL NETWORK U-NET

V.V GANCHENKO

*Senior researcher of of the
laboratory of System
Identification*

A.A. DOUDKIN

*Doctor of Computer
Sciences, Professor,
Head of the laboratory of
System Identification*

S.V. SHELEG

*Master's degree student of
the BSUIR, software engineer
in LLC «BSVT – New
Technologies»*

*United Institute of Informatics Problems the National Academy of Sciences of Belarus
Belarusian State University of Informatics and Radioelectronics, Republic of Belarus
LLC «BSVT – New Technologies»*

E-mail: ganchenko@lsi.bas-net.by, doudkin@newman.bas-net.by, sergey.sheleg@hotmail.com

Abstract. This paper considers the task of segmenting potato fields based on aerial photography of different spatial resolution. A convolutional neural network with U-Net architecture is used as the basis of the recognition algorithm. The proposed algorithm allows separating the pixels of the image into two classes: "vegetation" and "soil." A comparative analysis of the proposed algorithm has been made with other neural network algorithms.

Keywords: convolutional neural network, semantic segmentation, multispectral image, agricultural vegetation, recognition