

УДК 004.932

DOI: 10.15587/1729-4061.2021.248390

Вдосконалення моделі нейронної мережі для семантичної сегментації зображень об'єктів моніторингу на аерофотознімках

В. І. Слюсар, М. М. Проценко, А. А. Чернуха, В. В. Мелькін, О. І. Петрова, М. М. Кравцов, С. В. Вельма, Н. В. Косенко, О. В. Сидоренко, М. О. Соболев

Розглянуто модель нейронної мережі для семантичної сегментації зображень об'єктів моніторингу на аерофотознімках. Безпілотні літальні апарати здійснюють моніторинг об'єктів шляхом аналізу (обробки) аерофотознімків та відеопотоку. Результати аерофотозйомки обробляються оператором у ручному режимі, але є об'єктивні труднощі, пов'язані з обробкою оператором великої кількості аерофотознімків, тому доцільно цей процес автоматизувати. Аналіз моделей показав, що для виконання завдання семантичної сегментації зображень об'єктів моніторингу на аерофотознімках в якості базової моделі найбільш підходить модель U-Net (Німеччина), яка є нейронною мережею. Дану модель вдосконалено шляхом використання вейвлет шару та застосування оптимальних значень параметрів навчання моделі: швидкість (крок) – 0,001, число epoch – 60, алгоритм оптимізації – Adam. Навчання проводилось набором сегментованих зображень, які підготовлені з аерофотознімків (роздільною здатністю 6000×4000 пікселів) програмою Image Labeler середовища математичного моделювання MATLAB R2020b (США). В результаті отримано нову модель семантичної сегментації зображень об'єктів моніторингу на аерофотознімках із запропонованою назвою U-NetWavelet.

Проведено дослідження ефективності вдосконаленої моделі на прикладі обробки 80 аерофотознімків. В якості показників ефективності моделі обрані: точність; чутливість; помилка сегментації. Використання модифікованого вейвлету шару дозволило адаптувати розмір аерофотознімка до параметрів вхідного шару нейронної мережі, підвищити оперативність сегментації зображень на аерофотознімках, а застосування згорткової нейронної мережі дозволило автоматизувати даний процес.

Ключові слова: семантична сегментація зображень, згорткова нейронна мережа, аерофотознімок, безпілотний літальний апарат.

1. Введение

Использование беспилотных летательных аппаратов (БПЛА) позволяет ускорять процесс мониторинга за объектами критической инфраструктуры [1]. К таким объектам относятся промышленные предприятия [2], объекты энергетики [3], химически опасные производства [4] и другие стратегические объекты [5]. Нарушение функционирования данных объектов может угрожать национальным интересам жизни людей [6, 7]. С помощью БПЛА проводят мониторинг объектов путем обработки (анализа) аэрофотоснимков и видеопотока. Одним из видов обработки изображения является его сегментация. Сегментация

аэрофотоснимков предполагает разбиение его на области по определенным критериям. Результатом сегментации является множество областей, которые покрывают весь аэрофотоснимок. Поэтому особую актуальность приобретает разработка новых и усовершенствование существующих моделей нейронной сети для сегментации изображений объектов мониторинга (ОМ) на аэрофотоснимках.

2. Анализ литературных данных и постановка проблемы

В работе [8] показано, что наблюдение за дорожным движением с использованием БПЛА приобрело большую популярность в гражданских приложениях и задачах дистанционного зондирования. Благодаря своей высокой мобильности и большому полю зрения, а также способности охватывать большие районы на разных высотах БПЛА в последние годы стали востребованным средством наблюдения. Предлагается вариант подсчета транспортных средств с устранением проблемы избыточного подсчета информации в последовательных кадрах видео с БПЛА. Однако в этой статье не рассматривались вопросы, связанные с сегментацией изображений ОМ.

В работе [9] предложены различные модели на основе сверточных нейронных сетей (convolutional neural networks – CNN) для сбора информации, полученной с помощью сети сегментации, была предложена генеративная состязательная сеть на основе Pixel2Pixel. Дискриминатор использовал CNN, чтобы различать результаты сегментации сгенерированной модели и эксперта. Результаты показали, что сетевая модель может обеспечить эффективную автоматическую сегментацию гиппокампа и имеет практическое значение для правильной диагностики заболеваний, таких как болезнь Альцгеймера. Недостатком данного метода является его высокая вычислительная сложность, неадаптированность его для сегментации изображений ОМ на аэрофотоснимках.

В работе [10] предложен быстрый алгоритм кластеризации на основе суперпикселей для сегментации изображений радара с синтезированной апертурой. Экспериментальные результаты двух реальных изображений радара с синтезированной апертурой показывают, что предложенный метод превосходит другие современные методы как с точки зрения точности сегментации, так и с точки зрения вычислительной эффективности. Недостатком данной модели является неадаптированность ее для сегментации изображений ОМ на аэрофотоснимках.

В работе [11] показано, что методы обнаружения вредоносного кода, основанные на глубоком обучении, в основном имеют высокую точность. Но при обнаружении семейств вредоносных кодов с высокой степенью сходства из-за отсутствия очевидных функций обучения точность обнаружения серьезно снижается. Для решения этой проблемы в данной статье предлагается метод обнаружения вредоносного кода, основанный на сегментации изображения и глубокой CNN. Недостаток модели: высокая вычислительная сложность и неадаптированность для сегментации изображений ОМ на аэрофотоснимках.

В работе [12] предлагается многомасштабная модель семантической сегментации в реальном времени. Экспериментальным путем показано, что предлагаемая модель может применяться для решения множества задач распознавания, обладает хорошей способностью декодирования. Несмотря на

это, вопросы автоматизации процесса сегментации изображений ОМ на аэрофотоснимках не рассматривались.

В работе [13] предлагается новая схема классификации для гиперспектральных изображения дистанционного зондирования земли. Предложенная модель способна увеличивать внутриклассовое сходство за счет локального подавления спектральных вариаций, одновременно способствуя межклассовой различимости в глобальном масштабе, что приводит к восстановлению с более различимыми пикселями. Экспериментальные результаты на трех тестовых наборах данных демонстрируют значительное превосходство предложенного метода над современными. Недостатком данной модели является неадаптированность ее для сегментации изображений ОМ на аэрофотоснимках.

В работе [14] рассматривается получение точной многомасштабной семантической информации из изображений для качественной семантической сегментации. Предложена модель под названием cross fusion net (CF-Net) для быстрого и эффективного извлечения многомасштабной семантической информации. Модель способна кодировать более точную семантическую информацию от мелкомасштабных объектов, и соответственно повышать точность сегментации мелкомасштабных объектов. Недостатком модели является ее вычислительная сложность.

Анализ литературы [8–14] показал, что недостатками известных моделей (методов) являются:

- вычислительная сложность сегментации изображений ОМ на аэрофотоснимках, полученных с БПЛА;
- отсутствие моделей нейронных сетей, которые решают задачу сегментации изображений ОМ на аэрофотоснимках.

Все это позволяет утверждать, что целесообразно проведение исследования по усовершенствованию модели нейронной сети для семантической сегментации изображений объектов мониторинга на аэрофотоснимках, что позволит значительно улучшить точность и оперативность сегментации изображений ОМ на аэрофотоснимках.

3. Цель и задачи исследования

Цель исследования заключается в усовершенствовании модели нейронной сети для сегментации изображений ОМ на аэрофотоснимках с выбором параметров ее обучения. Это даст возможность автоматизировать процесс анализа (обработки) аэрофотоснимков.

Для достижения цели были поставлены следующие задачи:

- исследовать эффективность сегментации изображений ОМ с использованием CNN;
- оценить эффективность сегментации изображений ОМ на аэрофотоснимках предложенной моделью U-NetWavelet.

4. Материалы и методы исследования

Предположим, что цифровая фотокамера установлена на борту БПЛА. При этом аэрофотоснимки передаются через канал связи на компьютер наземного пункта управления. Здесь они хранятся в цифровом виде в виде файла. Сегмен-

тация важна для задач анализа изображений объектов мониторинга на аэрофотоснимках. Семантическая сегментация (Semantic segmentation) описывает процесс соединения каждого пикселя изображения с меткой класса (цветом).

Математическая постановка задачи семантической сегментации изображений заключается в том, чтобы каждому пикселю изображения OM на аэрофотоснимке $S(x,y,z)$ поставить в соответствие метку (цвет) каждого пикселя класса (объекта) B_i :

$$P[\|S\|] = B_i, \quad (1)$$

где P – оператор, который характеризует работу CNN.

В предложенной модели на вход CNN подаётся RGB аэрофотоснимок; размерностью $6000 \times 4000 \times 3$; формат JPEG, на выходе – метка (цвет) каждого пикселя класса (объекта) табл. 1.

Таблица 1
Метка (цвет) каждого пикселя класса (объекта)

Класс	Название класса	Метка	Цвет каждого пикселя класса (объекта)
1	Вертолет	Helicopter	
2	Самолет	Airplane	
3	Танк	Tank	
4	Тягач техники	Vehicletractor	
5	Грузовой автомобиль	Truck	
6	Легковой автомобиль	Car	
7	Автобус	Bus	

Исследование распознавания объектов на аэрофотоснимках проводилось с использованием методов CNN в сочетании с выбором оптимальных параметров обучения.

Для автоматизации процесса семантической сегментации изображений OM на аэрофотоснимках предложено использовать модель U-Net как базовую, которая показала высокую эффективность при решении задач биомедицины.

Архитектура CNN U-Net рассмотрена в [15, 16] и показана на рис. 1. В CNN в операциях свертки используется матрица весов. Слой свертки производит суммирование результатов поэлементного произведения каждого фрагмента изображения на матрицу – ядро свертки.

U-Net состоит из сужающегося пути (левая сторона) и расширяющегося пути (правая сторона). Он состоит из применения двух сверток 3×3 (неполнотных сверток), за каждой из которых следует положительно линейная функция ReLU и операция максимального объединения (пулинга) 2×2 с шагом 2 для понижения дискретизации. На каждом этапе понижающей дискретизации удваивается количество функциональных каналов. Каждый шаг расширенного пути состоит из повышающей дискретизации карты признаков, за которой следует свертка 2×2 («свертка вверх»), которая вдвое уменьшает количество каналов признаков.

Каждый шаг сужающегося пути состоит из понижающей дискретизации карты признаков, за которой следует свертка 3×3 , за каждой из которых следует ReLU.

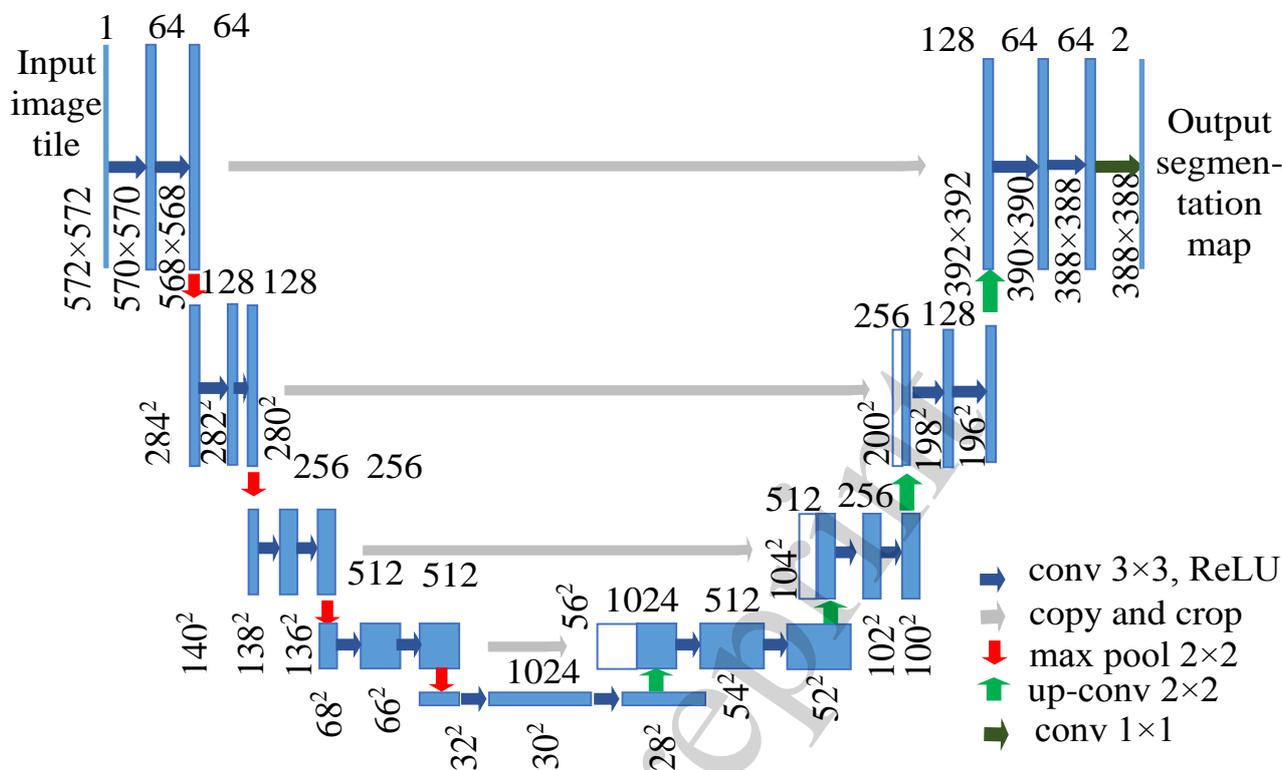


Рис. 1. Архитектура U-Net [15] (пример для 32×32 пикселей в самом низком разрешении). Каждое синее поле соответствует карте многоканальных функций. Количество каналов указано в верхней части окна. Размер x , y указан в нижнем левом углу поля. Белые прямоугольники представляют собой скопированные карты функций. Стрелки обозначают различные операции

Обрезка необходима из-за потери краевых пикселей при каждой свертке. На последнем слое свертка 1×1 используется для отображения каждого 64-компонентного вектора признаков на желаемое количество классов. Всего в сети 23 сверточных слоя.

Особенности функции активации ReLU ее математическое описание детально описано в [17, 18] а реализация операции максимального объединения (пулинга) в [17].

Обучение U-Net.

U-Net проходит обучение методом стохастического градиентного спуска на основе входных изображений и соответствующих им карт сегментации. Из-за свертки выходное изображение меньше входного сигнала на постоянную ширину границы. Применяемая попиксельно, функция Softmax, которая вычисляет энергию по окончательной карте признаков вместе с функцией кросс-энтропии. Функция Softmax определяется как [15]:

$$\rho_k(x) = \frac{\exp(a_k(x))}{\sum_{k=1}^{k=K} \exp(a_k(x))}, \quad (2)$$

где $p_k(x)$ – значение функции приближается к 1, когда k имеет максимальную активацию $a_k(x)$, который представляет канал функции активации k позиции пикселя ($x \in \Omega$) и $(\Omega \square \mathbb{Z}^2)$;

k – обозначает количество классов.

Перекрестная энтропия в каждой точке, показывает отклонение и определяется как [15]:

$$E = \sum_{x \in \Omega} w(x) \log(\rho_{\ell(x)}(x)), \quad (3)$$

где $\ell: \Omega \rightarrow \{1, \dots, K\}$ – истинная метка каждого пикселя;

$w: \Omega \rightarrow \mathbb{R}$ – карта веса, которая введена, чтобы придать некоторым пикселям большее значение в тренировке.

Граница разделения вычисляется с использованием морфологических операций. Вычисление карты весовых коэффициентов осуществляется по формуле [15]:

$$w(x) = w_c(x) + w_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right), \quad (4)$$

где $w_c: \Omega \rightarrow \mathbb{R}$ – карта весов для балансировки частот классов;

$d_1: \Omega \rightarrow \mathbb{R}$ – расстояние до границы ближайшей ячейки;

$d_2: \Omega \rightarrow \mathbb{R}$ – расстояние до границы второй ближайшей ячейки;

экспериментальным образом установлено $w_0=10$ и $\sigma=5$ пикселей [11].

Обоснование архитектуры и использованного математического аппарата для реализации, предложенной CNN.

Анализ литературы [15, 16] показал, что модель U-Net показывает высокую эффективность для семантической сегментации изображений объектов различной формы и положения.

Преимущества U-Net и нейронных сетей на ее основе:

– высокая эффективность для решения задач сегментации медицинских изображений [14, 15];

– информация с больших масштабов (верхние слои) позволяет модели лучше классифицировать;

– информация с меньших масштабов (глубокие слои) помогает модели лучше сегментировать;

– увеличение размерности за счет увеличения количества каналов признаков, позволяет распространять CNN контекстную информацию на слои большего разрешения;

– стратегия симметричной сети позволяет обрабатывать большие изображения (снимки) такие как аэрофотоснимки, гиперспектральные снимки, изображения для ортофотопланов;

– использование небольшого количества изображений [15] для обучения и получения хорошей точности.

Для решения задачи семантической сегментации изображений объектов мониторинга на аэрофотоснимках по 7 классам и повышения эффективности сегментации предлагается использовать в качестве входного слоя модифицированный вейвлет слой, а в качестве базовой модели – CNNU-Net. Обучение модели осуществлялось набором изображений, подготовленным с аэрофотоснимков.

Архитектура CNN (базовая U-Net) показана на рис. 2. Задача, решаемая CNN, – семантическая сегментация изображений объектов мониторинга по 7 классам.



Рис. 2. Архитектура предложенной CNN (базовая U-Net) в фреймворке Terra AI

Слои нейронной сети (рис. 2):

1. Входной – 1) Input Вход 1.
2. Сверточный – 3) Conv2D: Entry block (filters=16, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='relu').
3. Нормализации – 4) BatchNormalization.
4. Сверточный – 5) Conv2D: Layer 2 (filters=16, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='relu').

5. Нормализации – 6) BatchNormalization.
 6. MaxPool – 7) MaxPool2D: 1stDo...(pool_size=[2, 2], padding='same').
 7. Сверточный – 8) Conv2D: Layer 6 (filters=32, kernel_size=[3, 3],
 strides=[1, 1], padding='same', activation='relu').

· · ·
 8. Объединяющий – 30) Concatenate: Layer.

· · ·
 56. Выходной – 2) Conv2D: Выход2 (filters=2, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='sigmoid').

Фрагмент кода на языке Python с использованием библиотеки Keras для нейронной сети (рис. 2):

```
from tensorflow.keras.layers import Input
from tensorflow.keras.layers import Conv2D
from tensorflow.keras.layers import BatchNormalization
from tensorflow.keras.layers import MaxPool2D
from tensorflow.keras.layers import Conv2DTranspose
from tensorflow.keras.layers import Concatenate
from tensorflow.keras.models import Model
input_1 = Input(shape=(512, 512, 3), name='1')
x_3 = Conv2D(filters=16, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='relu', data_format='channels_last', dilation_rate=[1, 1], groups=1, use_bias=True, kernel_initializer='glorot_uniform', bias_initializer='zeros', kernel_regularizer=None, bias_regularizer=None, activity_regularizer=None, kernel_constraint=None, bias_constraint=None, name='Conv2D_3')(input_1)
x_4 = BatchNormalization(axis=-1, momentum=0.99, epsilon=0.001, center=True, scale=True, beta_initializer='zeros', gamma_initializer='ones', moving_mean_initializer='zeros', moving_variance_initializer='ones', beta_regularizer=None, gamma_regularizer=None, beta_constraint=None, gamma_constraint=None, name='BatchNormalization_4')(x_3)
· · ·
output_2 = Conv2D(filters=7, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='softmax', data_format='channels_last', dilation_rate=[1, 1], groups=1, use_bias=True, kernel_initializer='glorot_uniform', bias_initializer='zeros', kernel_regularizer=None, bias_regularizer=None, activity_regularizer=None, kernel_constraint=None, bias_constraint=None, name='2')(x_55)
model = Model([input_1], [output_2])
```

Входной слой – служит для загрузки входных данных(изображения).

Сверточный слой – основной слой свёрточной нейронной сети. Слой свёртки включает в себя для каждого канала свой фильтр, ядро свёртки которое

обрабатывает предыдущий слой по фрагментам (суммируя результаты поэлементного произведения для каждого фрагмента).

Слой нормализации – необходим, чтобы различные элементы в разных местах одной карты признаков (образ операции свёртки) были нормализованы одинаково.

MaxPool слой – необходим с целью ускорения процесса обучения и уменьшения используемых вычислительных ресурсов.

Объединяющий слой – объединяет выходы слоев нейронной сети.

Выходной слой – последний слой нейронной сети, выдает выходные данные (результат) работы нейронной сети.

Архитектура предложенной CNN аналогична U-Net, отличие состоит в размерности входного и выходного слоя сети.

В качестве показателей эффективности, которые характеризуют процесс обучения и оценки эффективности CNN выбрано [18]:

– точность (accuracy) – это отношение правильно сегментированных объектов к общему количеству предполагаемых и истинных объектов [18]:

$$\text{Accuracy}_{val} = \sum_{t=1}^{N_{val}} \frac{N_{TP_k}}{N_{TP_k} + N_{FP_k}} \cdot 100 \% , \quad (5)$$

где N_{TP} – количество правильно сегментированных объектов на аэрофотоснимке;

N_{FP} – количество ошибочно сегментированных объектов на аэрофотоснимке;

N_{val} – количество аэрофотоснимков в проверочной выборке;

t – текущий аэрофотоснимок.

– чувствительность (sensitivity) – это отношение правильно сегментированных объектов к общему количеству объектов на аэрофотоснимке [18]:

$$\text{Sensitivity}_{val} = \sum_{t=1}^{N_{val}} \frac{N_{TP_k}}{N_{TP_k} + N_{FN_k}} \cdot 100 \% , \quad (6)$$

где N_{FN} – количество ошибочно несегментированных объектов на аэрофотоснимке.

Тестирование моделей CNN проводилось на компьютере ACPI X64 (Китай), оснащённом видеокартой GPU Tesla 12 Гбайт и оперативной памятью объемом 8 Гбайт.

Для подготовки аэрофотоснимков тренировочной выборки использовалась программа Image Labeler среды математического моделирования MATLAB R2020b (США). Подготовка (разметка) аэрофотоснимка объектов “Truck”, “Car” показана на рис. 3.

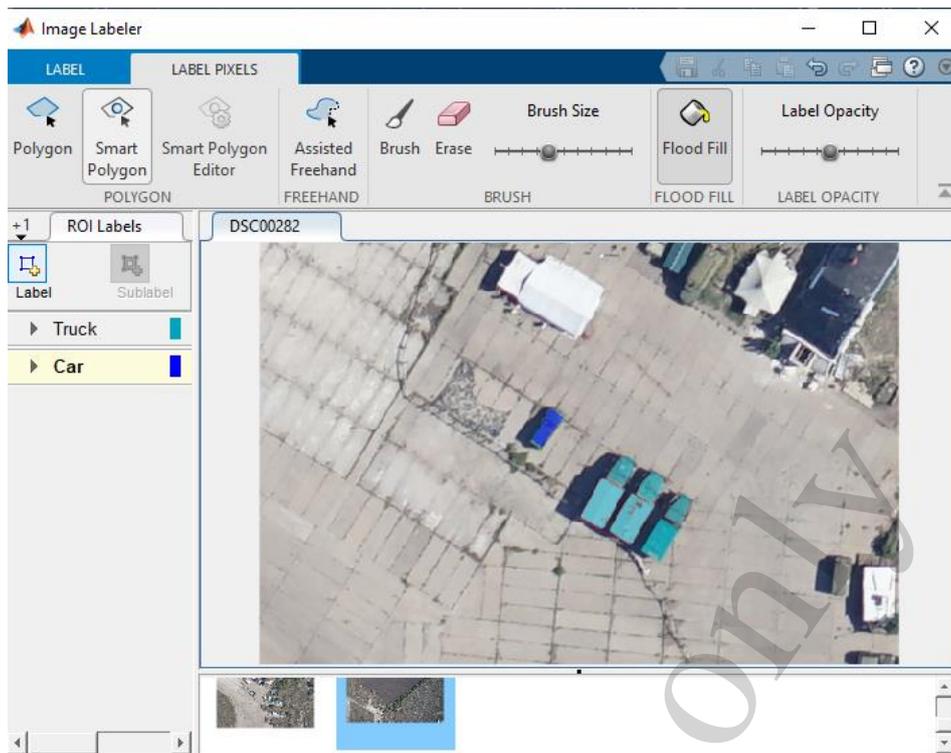


Рис. 3. Подготовка сегментированного изображения тренировочной выборки объектов “грузовой автомобиль, легковой автомобиль” из аэрофотоснимка программой Image Labeler среды математического моделирования MATLAB R2020b

Исследования выполнялись при следующих допущениях и ограничениях:

- цифровая фотокамера установлена на борту БПЛА и осуществляет съемку в видовом диапазоне в дневное время;
- аэрофотоснимок в цифровом виде передается через канал связи на наземный пункт управления;
- процесс семантической сегментации изображений объектов на аэрофотоснимке проводится на компьютере наземного пункта управления беспилотного авиационного комплекса.

5. Результаты исследования эффективности сегментации изображений объектов мониторинга на аэрофотоснимках с использованием CNN

5.1. Исследование эффективности сегментации изображений объектов мониторинга с использованием CNN

Исследована эффективность сегментации изображений ОМ с использованием CNN следующих моделей U-Net (рис. 3), PSPsmall (рис. 4), U-Netaverage (рис. 5). Для обучения и проверки моделей использовался набор изображений самолетов размерностью $128 \times 160 \times 3$, тип RGB, формат JPEG. Тренировочная выборка – 800 изображений, проверочная – 140 изображений.

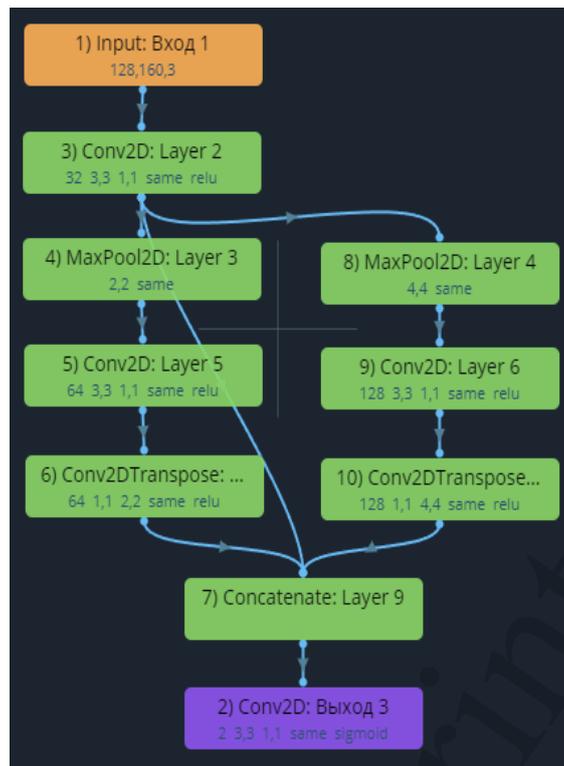


Рис. 4. Архитектура модели PSPsmall в фреймворке Terra AI

Слои нейронной сети (рис. 4):

1. Входной – 1) Input Вход 1 размер входного изображения 128, 160, 3 пикселей.

2. Сверточный – 3) Conv2D: Layer 2 (filters=32, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='relu').

3. MaxPool – 4) MaxPool2D: Layer 3 (pool_size=[2, 2], padding='same').

4. Сверточный – 5) Conv2D: Layer 5 (filters=64, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='relu').

5. Сверточный – 6) Conv2D: Layer 2 (filters=64, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='relu').

...

8. Выходной – 2) Conv2D Выход 3 (filters=2, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='sigmoid').

Слои нейронной сети (рис. 5):

1. Входной – 1) Input Вход 1 размер входного изображения 128, 160, 3 пикселей.

2. Сверточный – 3) Conv2D: Entry block (filters=64, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='relu').

3. Нормализации – 4) BatchNormalization.

4. Сверточный – 5) Conv2D: Layer 3 (filters=64, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='relu').

...

42. Выходной слой – 2) Conv2D Выход 2 (filters=2, kernel_size=[3, 3], strides=[1, 1], padding='same', activation='sigmoid').

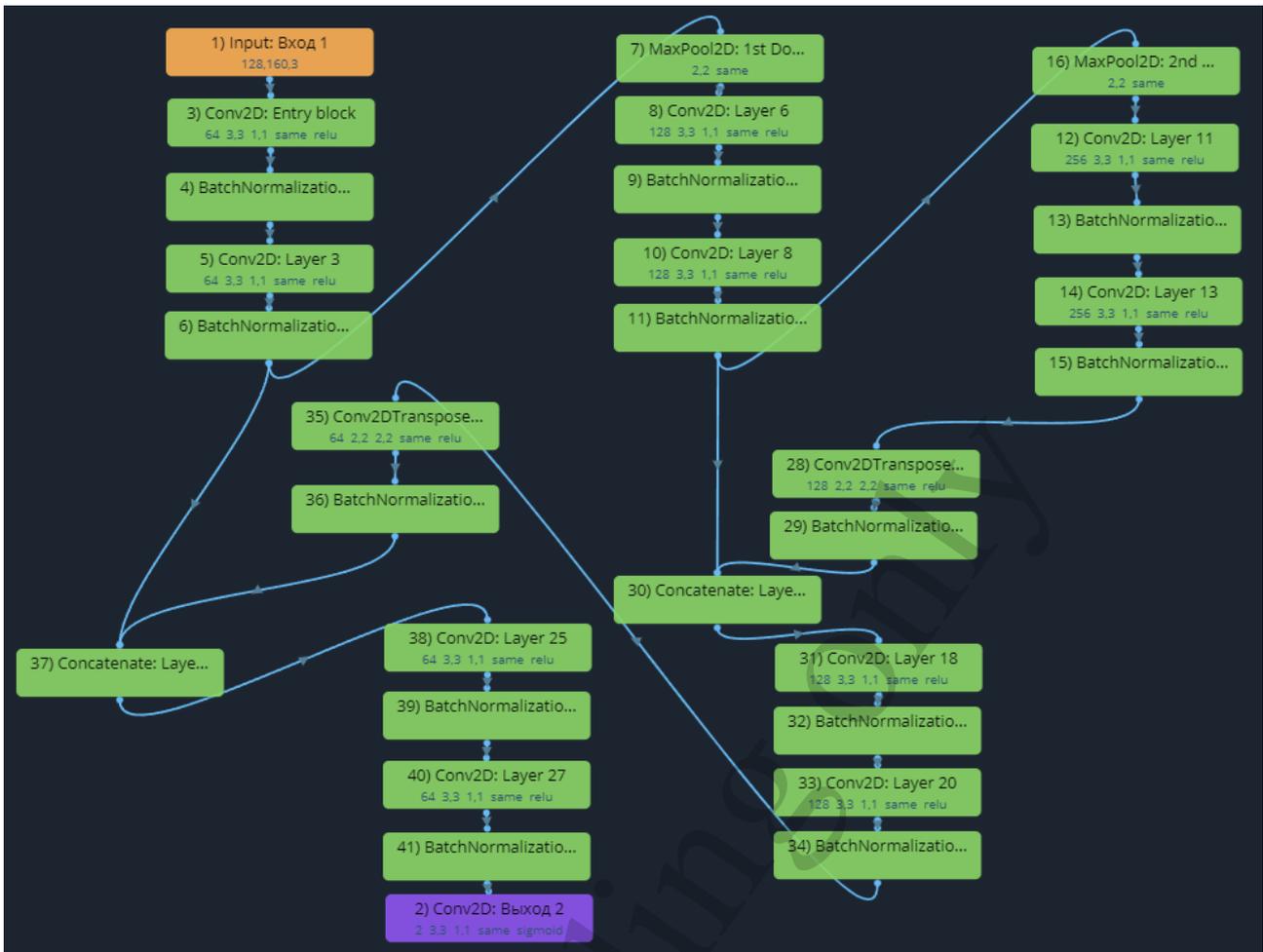


Рис. 5. Архитектура модели U-Netaverage в фреймворке Terra AI

Показателями эффективности семантической сегментации изображений объектов CNN было выбрано точность и чувствительность. Параметрами обучения CNN являются продолжительность обучения (количество эпох), алгоритм оптимизации, скорость обучения (шаг обучения). Физический смысл скорости обучения (шага обучения) CNN показано в [18].

Параметры обучения и тестирования моделей CNN для семантической сегментации изображений объектов показаны в табл. 2. При этом сравнение проводилось по трём типам нейросетей: PSPsmall, U-Netaverage и U-Net.

Таблица 2
Параметры обучения и проверки CNN

Модель	Число эпох	Время	Алгоритм оптимизации	Размер пакета (батча)	Скорость
PSPsmall	30	4 м 37 с	Adam	20	0,001
U-Netaverage	30	6 м 00 с	Adam	20	0,001
U-Net	30	28 м 15 с	Adam	20	0,001

Для моделирования использовался фреймворк Terra AI и среда математического моделирования MATLAB R2020b.

На рис. 6 показаны графики проверки точности на проверочной выборке моделей PSPsmall, U-Netaverage, U-Net.

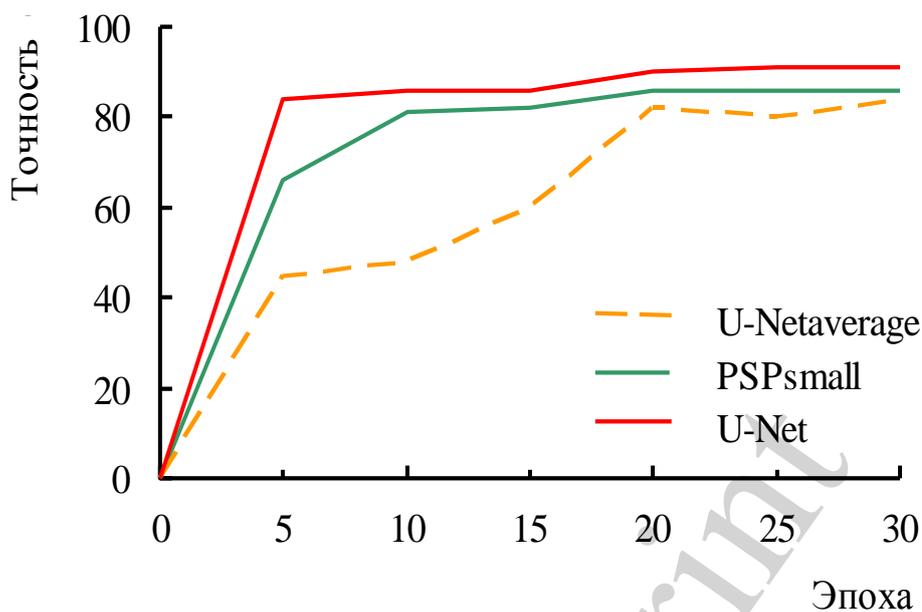


Рис. 6. Графики изменения точности на проверочной выборке в зависимости от эпохи для модели PSPsmall, U-Netaverage, U-Net

На рис. 6 видно, что на проверочной выборке лучшую точность (91 %) показывает модель U-Net. После 20 эпохи точность модели изменяется в диапазоне от 90 % до 91 %.

На рис. 7 показаны графики проверки чувствительности на проверочной выборке моделей PSPsmall, U-Netaverage, U-Net.

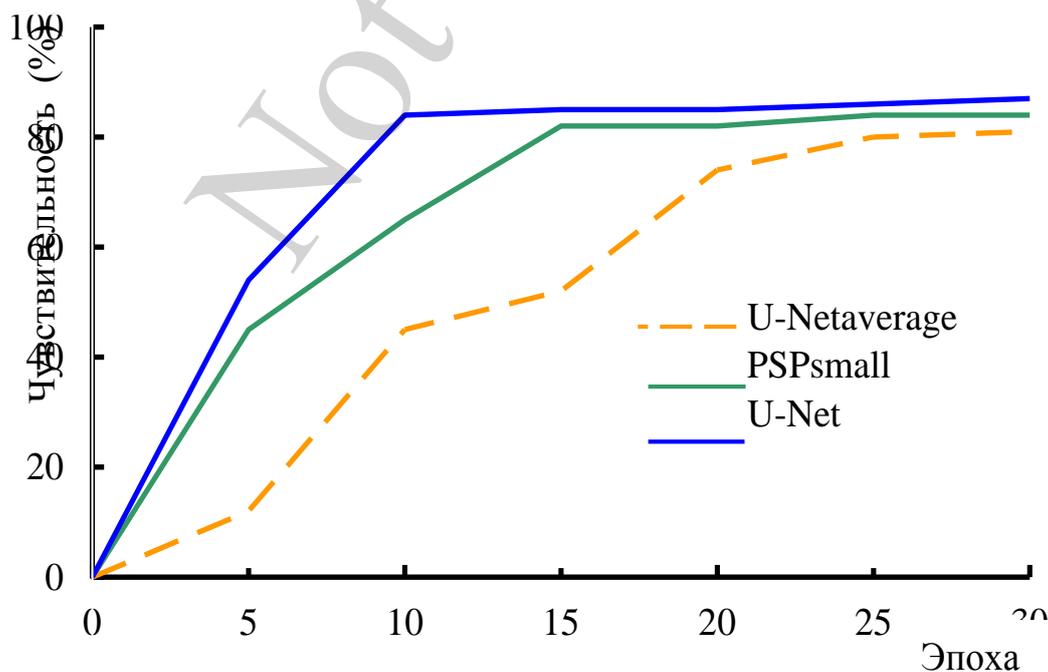


Рис. 7. Графики изменения чувствительности на проверочной выборке в зависимости от эпохи для модели PSPsmall, U-Netaverage, U-Net

Из рис. 7 видно, что на проверочной выборке модель U-Net показывает наилучшую чувствительность (87 %), которая после 10-й эпохи стабилизируется и изменяется в диапазоне от 84 % до 87 %.

На рис. 8 показано результат семантической сегментации изображений “самолет” моделью U-Net в фреймворке Terra AI. На рис. 8 при сегментации выделено 2 области: ■ “самолет”, ■ “небо”.



Рис. 8. Семантическая сегментация изображений “самолет” в фреймворке Terra AI с использованием модели U-Net

Анализ полученных результатов показывает, что наилучшие показатели эффективности показывает модель U-Net: точность (91 %), чувствительность (87 %), максимальное значение ошибки (0,232), минимальное значение ошибки (0,0132).

5. 2. Оценка эффективности сегментации изображений ОМ на аэрофотоснимках предложенной моделью U-NetWavelet

Для исследования эффективности сегментации изображений ОМ на аэрофотоснимках были подготовлены аэрофотоснимки тренировочной и проверочной выборки. В качестве тренировочной выборки использовалось 100 аэрофотоснимков. Общее количество классов для семантической сегментации равнялось 7 (вертолет, самолет, танк, тягач техники, грузовой автомобиль, легковой автомобиль, автобус). Тип тренировочной и проверочной выборки (одинаковый) – аэрофотоснимок 6000×4000 пикселей; формат JPEG. В качестве проверочной выборки использовалось 80 аэрофотоснимков.

Сегментация изображений объектов мониторинга на аэрофотоснимках с использованием CNN проводилось на наземном пункте управления. Для съёмки использовался БПЛА, который оснащённый камерой Sony ILCE-7M2. Даная камера производила аэрофотосъёмку в следующем режиме:

- выдержка 1/1600 с.;
- фокусное расстояние 55 мм.;

– размер аэрофотоснимка (пикселей): 6000×4000 (24М).

Аэрофотоснимок сделан цифровой камерой Sony ILCE-7M2K с БПЛА на высоте 1100 метров, показан на рис. 9.



Рис. 9. Аэрофотоснимок камерой Sony ILCE-7M2K

Методика исследования (моделирования) на примере предложенной модели U-NetWavelet:

Шаг 1. Загрузка аэрофотоснимков $6000 \times 4000 \times 3$ пикселей.

Шаг 2. Разбивка аэрофотоснимков с $6000 \times 4000 \times 3$ пикселей на снимки $1000 \times 1000 \times 3$, всего 24 для каждого аэрофотоснимка.

Шаг 3. Применение вейвлет слоя к снимку $1000 \times 1000 \times 3$ (реализованного на модифицированном преобразовании Хаара – значение соседних двух пикселей суммируется и делится на два) и адаптируется под размерность $512 \times 512 \times 3$.

Шаг 4. Разделение данных на наборы данных обучения и проверки.

Шаг 5. Обучение и проверка (валидация) сети.

Шаг 6. Сегментация снимков проверочной выборки.

Шаг 7. Оценка точности сегментации тестовой выборки.

Шаг 8. Оценка чувствительности модели на тестовой выборке.

Шаг 9. Сборка сегментированного аэрофотоснимка $3072 \times 2048 \times 3$.

Обучение предложенной модели U-NetWavelet проводилось с использованием оптимальных значения параметров, которые получены экспериментальным путем:

– скорость обучения – 0,001;

– длительность обучения (число эпох) – 60;

– размер пакета (батча) – 20;

– алгоритм оптимизации – Adam.

В результате получена новая модель с предложенным названием U-NetWavelet. Результаты проверки точности, чувствительности данной нейронной сети показано на рис. 10.

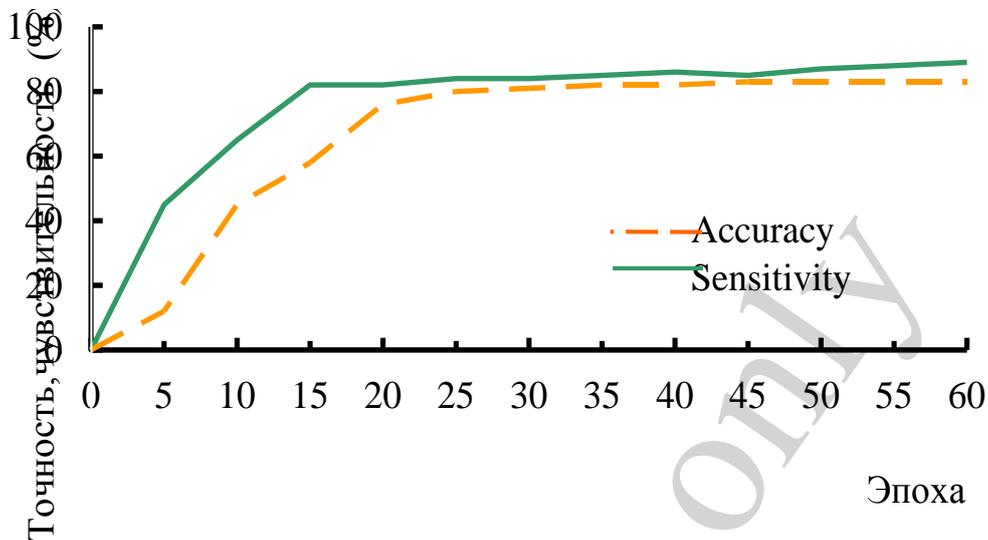


Рис. 10. Графики изменения точности (accuracy), чувствительности (sensitivity) на проверочной выборке в зависимости от эпохи для модели U-NetWavelet

Из рис. 10 видно, что для модели U-NetWavelet точность на проверочной выборке после 15-й эпохи стабилизируется, а после окончания 60-й эпохи точность достигает – 89 %, чувствительность – 83 %.

На рис. 11 показано фрагмент сегментированного аэрофотоснимка моделью U-NetWavelet.



Рис. 11. Фрагмент сегментированного аэрофотоснимка моделью U-NetWavelet

На рис. 11 при сегментации выделены объекты двух типов: ■ – “легковой автомобиль”, ■ – “грузовой автомобиль”.

Сравнение новой модели U-NetWavelet проводилось с моделями FCN, SegNet. Для проведения оценки модели U-NetWavelet на сходимость, адекватность и достоверность в качестве проверочной выборки использовалось 80 аэрофотоснимков.

Сходимость. CNN показывает сходимость при условии, что с каждой эпохой ошибка уменьшается. На сходимость модели CNN оказывает влияние три составляющие: полнота базы данных (аэрофотоснимков); правильность выбора архитектуры; подбор параметров обучения CNN.

На рис. 12 показана оценка сходимости U-NetWavelet на проверочной выборке.

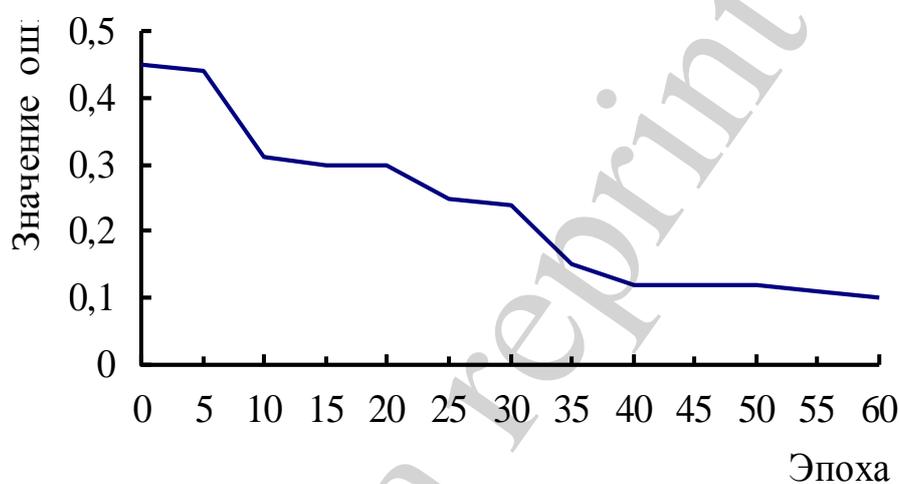


Рис. 12. Оценка сходимости предложенной модели U-NetWavelet

Таблица 3

Результаты оценки точности и чувствительности моделей

Модель	Точность (accuracy), %	Чувствительность (sensitivity), %	Максимальное значение ошибки	Минимальное значение ошибки
FCN	83	79	0,741	0,187
SegNet	85	82	0,536	0,124
Предложенная модель U-NetWavelet	89	83	0,451	0,102

Из анализа рис. 12 видно, что предложенная модель U-NetWavelet обладает сходимостью.

Адекватность. Нейронная сеть адекватна, если результаты обучения сходятся к близким значениям, – необходимое условие того, что между выходными и входными данными есть зависимость, которая реализуется CNN.

Наиболее рекомендуемым способом проверки модели CNN на адекватность является сравнение результатов с известными моделями.

Результаты проверки на проверочной выборке (80 аэрофотоснимков) показаны в табл. 3.

Из табл. 3 видно, что в сравнении с моделями FCN, SegNet предложенная модель U-NetWavelet имеет наилучшие показатели эффективности: точность (89 %), чувствительность (83 %), максимальное значение ошибки (0,451), минимальное значение ошибки (0,102).

6. Обсуждение результатов исследования семантической сегментации изображений объектов на аэрофотоснимках с использованием CNN

Предложено использовать CNN U-Net [15, 16] для сегментации изображений объектов на аэрофотоснимках. Для повышения эффективности нейронной сети выполнено обучение данной модели набором аэрофотоснимков (рис. 9) с подбором оптимальных параметров (скорость (шаг) обучения, число эпох, размер пакета (батча), алгоритм оптимизации). В результате получена новая модель с предложенным названием U-NetWavelet (рис. 2).

За счет использования модифицированного вейвлет слоя размер аэрофотоснимка адаптируется к параметрам входного слоя нейронной сети, повышается оперативность сегментации изображений на аэрофотоснимках. Применение CNN U-NetWavelet позволяет повысить быстродействие и автоматизировать процесс семантической сегментации изображений ОМ.

Использование предложенной модели позволяет решить проблему [8–14]:

- вычислительной сложности сегментации изображений ОМ на аэрофотоснимках, полученных с БПЛА;
- отсутствия моделей нейронных сетей, которые решают задачу сегментации изображений ОМ на аэрофотоснимках.

Ограничения предложенной модели:

- сегментация изображений ОМ на аэрофотоснимках осуществляется в пределах 7 классов (табл. 1);
- ориентация ОМ на изображениях не учитывается;
- разрешение аэрофотоснимков для классификации ОМ составляет 6000×4000 пикселей;
- трансляционная инвариантность CNN не учитывается;
- аэрофотосъемка осуществляется в видимом диапазоне в дневное время.

Ограничениями предложенной модели есть то, что она адаптирована для сегментации объектов на аэрофотоснимке по семи классам. Обучение CNN проводилось на аэрофотоснимках высокой контрастности, четкости (рис. 6). Съемка проводилась в дневное время, пора года – лето. Поэтому получены высокие значения точности и чувствительности сегментации изображений объектов (табл. 3). Для других видов изображений объектов (условий съемки) точность, чувствительность сегментации изображений ОМ по классам может меняться, что требует проведения дополнительных исследований.

Для развития предложенной модели планируется:

- увеличить базу размеченных (сегментированных) аэрофотоснимков для тренировочной выборки;

- исследовать предложенную и другие модели [19–21] (PSPNet, DenseNet, DeepLab, DilatedNet и др.) для разных условий аэрофотосъёмки;
- провести оптимизацию предложенной модели по вычислительной сложности, увеличить быстродействие;
- разработать метод подсчета количества объектов на аэрофотоснимках по классам;
- разработать метод обнаружения и идентификации объектов в видеопотоке, полученного видеокамерой БПЛА.

Данную модель предлагается использовать: на наземном пункте управления БПЛА при обработке аэрофотоснимков, ортофотопланов; в системах с искусственным интеллектом; в комплексах контроля ОМ; при создании роботов; в системах беспилотного транспорта.

7. Выводы

1. Исследованы показатели эффективности моделей PSPsmall, U-Netaverage, U-Net. Проверка эффективности данных моделей проводилась на основе изображений самолетов (800 изображений тренировочная выборка, 140 проверочная). Установлено, что наилучшие показатели показывает модель U-Net: точность (91 %), чувствительность (87 %), максимальное значение ошибки (0,232), минимальное значение ошибки (0,0132). Наименьшую точность (84 %) и чувствительность (81 %) показывает модель U-Netaverage.

2. Оценена эффективность предложенной модели U-NetWavelet (на основе изображений, подготовленных из аэрофотоснимков). Модель имеет наилучшие показатели эффективности в сравнении с моделями FCN, SegNet: точность (89 %), чувствительность (83 %), максимальное значение ошибки (0,451), минимальное значение ошибки (0,102). Полученные значения показателей эффективности модели U-NetWavelet позволяют утверждать о правильности выбора архитектуры CNN и подбора параметров ее обучения: скорость обучения – 0,001; длительность обучения (число эпох) – 60; алгоритм оптимизации – Adam.

Литература

1. Pospelov, B., Andronov, V., Rybka, E., Krainiukov, O., Maksymenko, N., Meleshchenko, R. et. al. (2020). Mathematical model of determining a risk to the human health along with the detection of hazardous states of urban atmosphere pollution based on measuring the current concentrations of pollutants. *Eastern-European Journal of Enterprise Technologies*, 4 (10 (106)), 37–44. doi: <https://doi.org/10.15587/1729-4061.2020.210059>
2. Semko, A. N., Beskrovnaya, M. V., Vinogradov, S. A., Hritsina, I. N., Yagudina, N. I. (2014). The usage of high speed impulse liquid jets for putting out gas blowouts. *Journal of Theoretical and Applied Mechanics*, 52 (3), 655–664.
3. Chernukha, A., Teslenko, A., Kovalov, P., Bezuglov, O. (2020). Mathematical Modeling of Fire-Proof Efficiency of Coatings Based on Silicate Composition. *Materials Science Forum*, 1006, 70–75. doi: <https://doi.org/10.4028/www.scientific.net/msf.1006.70>

4. Vambol, S., Vambol, V., Kondratenko, O., Suchikova, Y., Hurenko, O. (2017). Assessment of improvement of ecological safety of power plants by arranging the system of pollutant neutralization. *Eastern-European Journal of Enterprise Technologies*, 3 (10 (87)), 63–73. doi: <https://doi.org/10.15587/1729-4061.2017.102314>
5. Vambol, S., Vambol, V., Sobyna, V., Koloskov, V., Poberezhna, L. (2018). Investigation of the energy efficiency of waste utilization technology, with considering the use of low-temperature separation of the resulting gas mixtures. *Energetika*, 64 (4), 186–195. doi: <https://doi.org/10.6001/energetika.v64i4.3893>
6. Pospelov, B., Rybka, E., Meleshchenko, R., Borodych, P., Gornostal, S. (2019). Development of the method for rapid detection of hazardous atmospheric pollution of cities with the help of recurrence measures. *Eastern-European Journal of Enterprise Technologies*, 1 (10 (97)), 29–35. doi: <https://doi.org/10.15587/1729-4061.2019.155027>
7. Dadashov, I., Loboichenko, V., Kireev, A. (2018). Analysis of the ecological characteristics of environment friendly fire fighting chemicals used in extinguishing oil products. *Pollution Research*, 37 (1), 63–77. URL: <http://repositsc.nuczu.edu.ua/handle/123456789/6849>
8. Holla, A., Pai, M., Verma, U., Pai, R. M. (2020). Efficient Vehicle Counting by Eliminating Identical Vehicles in UAV aerial videos. 2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), 246–251. doi: <https://doi.org/10.1109/discover50404.2020.9278095>
9. Deng, H., Zhang, Y., Li, R., Hu, C., Feng, Z., Li, H. (2022). Combining residual attention mechanisms and generative adversarial networks for hippocampus segmentation. *Tsinghua Science and Technology*, 27 (1), 68–78. doi: <https://doi.org/10.26599/tst.2020.9010056>
10. Jing, W., Jin, T., Xiang, D. (2021). Fast Superpixel-Based Clustering Algorithm for SAR Image Segmentation. *IEEE Geoscience and Remote Sensing Letters*, 1–1. doi: <https://doi.org/10.1109/lgrs.2021.3124071>
11. Xin, L., Chao, L., He, L. (2021). Malicious code detection method based on image segmentation and deep residual network RESNET. 2021 International Conference on Computer Engineering and Application (ICCEA), 473–480. doi: <https://doi.org/10.1109/ICCEA53728.2021.00099>
12. Xie, B., Yang, Z., Yang, L., Luo, R., Wei, A., Weng, X., Li, B. (2021). Multi-Scale Fusion With Matching Attention Model: A Novel Decoding Network Cooperated With NAS for Real-Time Semantic Segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 1–11. doi: <https://doi.org/10.1109/tits.2021.3115705>
13. Yang, S., Hou, J., Jia, Y., Mei, S., Du, Q. (2021). Superpixel-Guided Discriminative Low-Rank Representation of Hyperspectral Images for Classification. *IEEE Transactions on Image Processing*, 30, 8823–8835. doi: <https://doi.org/10.1109/tip.2021.3120675>
14. Peng, C., Zhang, K., Ma, Y., Ma, J. (2021). Cross Fusion Net: A Fast Semantic Segmentation Network for Small-Scale Semantic Information Capturing in Aerial Scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–13. doi: <https://doi.org/10.1109/tgrs.2021.3053062>

15. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241. doi: https://doi.org/10.1007/978-3-319-24574-4_28
16. Jwaid, W. M., Al-Husseini, Z. S. M., Sabry, A. H. (2021). Development of brain tumor segmentation of magnetic resonance imaging (MRI) using U-Net deep learning. *Eastern-European Journal of Enterprise Technologies*, 4 (9 (112)), 23–31. doi: <https://doi.org/10.15587/1729-4061.2021.238957>
17. Slyusar, V., Protsenko, M., Chernukha, A., Gornostal, S., Rudakov, S., Shevchenko, S. et. al. (2021). Construction of an advanced method for recognizing monitored objects by a convolutional neural network using a discrete wavelet transform. *Eastern-European Journal of Enterprise Technologies*, 4 (9 (112)), 65–77. doi: <https://doi.org/10.15587/1729-4061.2021.238601>
18. Slyusar, V., Protsenko, M., Chernukha, A., Kovalov, P., Borodych, P., Shevchenko, S. et. al. (2021). Improvement of the model of object recognition in aero photographs using deep convolutional neural networks. *Eastern-European Journal of Enterprise Technologies*, 5 (2 (113)), 6–21. doi: <https://doi.org/10.15587/1729-4061.2021.243094>
19. Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: <https://doi.org/10.1109/cvpr.2015.7298965>
20. Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (12), 2481–2495. doi: <https://doi.org/10.1109/tpami.2016.2644615>
21. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. (2017). Pyramid Scene Parsing Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: <https://doi.org/10.1109/cvpr.2017.660>