

ДУДКИН А. А., ГАНЧЕНКО В. В., ИНЮТИН А. В., МАРУШКО Е. Е.

ИДЕНТИФИКАЦИЯ И КЛАССИФИКАЦИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ, ПОЛУЧЕННЫХ С ПОМОЩЬЮ СЪЕМОЧНЫХ СРЕДСТВ БПЛА И ОРБИТАЛЬНОГО БАЗИРОВАНИЯ

*Объединенный институт проблем информатики НАН Беларуси,
г. Минск, Республика Беларусь*

Для идентификации и классификации объектов на изображениях, полученных с помощью съемочных средств БПЛА и орбитального базирования, предложена нейросетевая модель классификации, основанная на использовании автоэнкодера и построенная по архитектуре ансамбля многослойных перцептронов. При выделении информативных признаков дополнительно добавляется цветовая информация, инвариантная к масштабу и поворотам изображения и основанная на построении поканальных гистограмм. Модель реализована с использованием библиотеки Keras. Использование предложенной модели для классификации на четыре класса: «Пожар», «Задымление», «Растительность» и «Строения», позволяет достичь точности классификации выше 99 %.

Ключевые слова: автоэнкодер; ансамбль многослойных перцептронов; классификация.

Введение

Мониторинг земной поверхности с целью обнаружения стихийных бедствий и оценки их последствий, контроля природопользования и экологической безопасности является актуальной задачей. Ряд стихийных бедствий, таких как паводковые наводнения, лесные пожары и засухи, аварии на техногенных объектах, затрагивают территории приграничных стран.

Для своевременного обнаружения на конкретном участке территории фактов возникновения и развития неблагоприятных явлений, ситуаций и стихийных бедствий необходимо, с целью скорейшей ликвидации этих последствий, использовать информацию от разных источников: космических снимков, данные аэросъемки, включая полученные с БПЛА, а также специальные данные от наземных средств мониторинга, включая мобильные.

Широкое внедрение информационных систем, основанных на обработке данных ДЗЗ, получаемых с помощью космических спутников и БПЛА, ставит задачу разработки эффективных методов и алгоритмов обработки изображений, в которых нуждаются потребители для распознавания природно-техногенных объектов в процессе мониторинга окружающей среды.

Постановка задачи

Задачей решаемой представленными нейросетевыми моделями является выделение параметров объектов на изображениях для их идентификации и классификации. Идентификация и классификация объектов осуществляется для участков исходных цветных изображений размером 128 x 128 пикселей.

Пример кадра видеосъемки БПЛА пожара на нефтебазе приведен на рис. 1.



Рис. 1. Пример кадра видеосъемки БПЛА

Также использовались файлы съемки спутником Landsat 8 в диапазоне, полученном комбинацией красного, синего, зеленого и панхроматического каналов, и тепловом инфракрасном диапазоне. Данные получены из Sentinel Hub [1]. Файлы съемки

спутником Landsat 8 в диапазоне, полученном комбинацией красного, синего, зеленого и панхроматического каналов, и тепловом инфракрасном диапазоне (см. рис. 2 – 3).

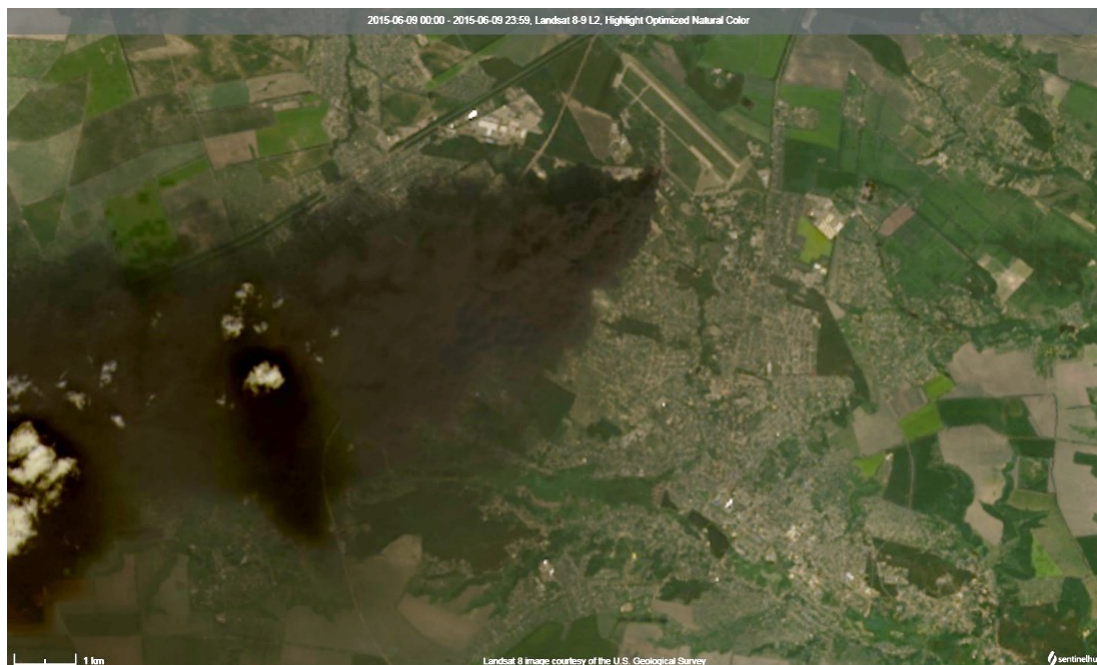


Рис. 2. Съемка спутником Landsat 8 в диапазоне, полученном комбинацией красного, синего, зеленого и панхроматического каналов

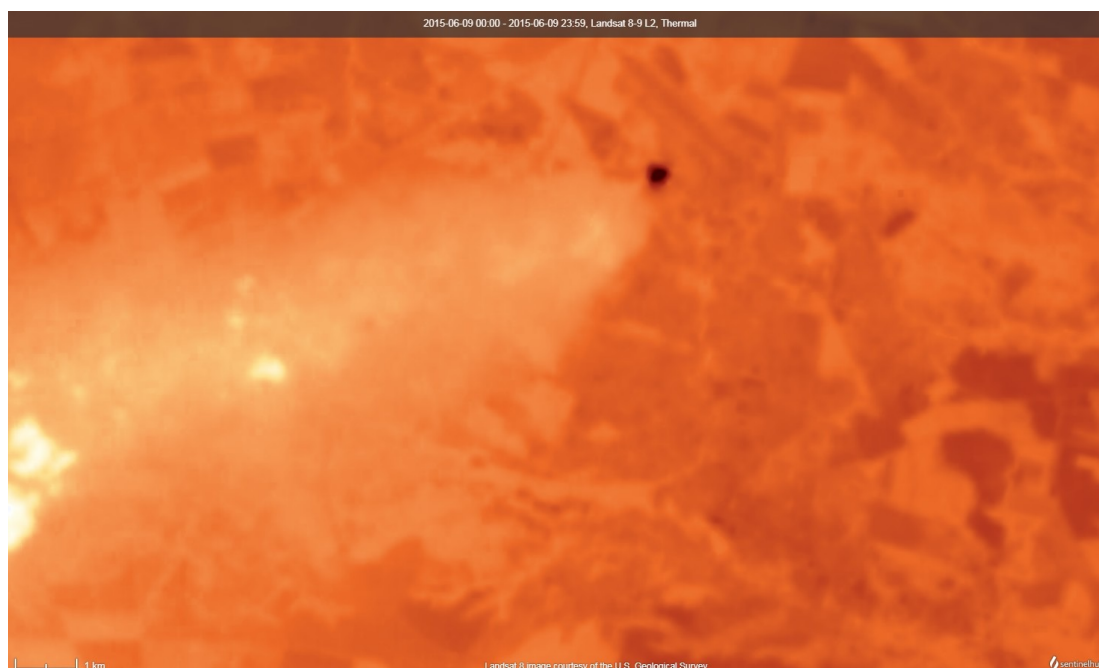


Рис. 3. Съемка спутником Landsat 8 в тепловом инфракрасном диапазоне

Разрешение цветных каналов составляет 30 метров на пиксель, Разрешение панхроматического канала – 15 метров на пиксель. Разрешение теплового инфракрасного канала – 100 метров на пиксель.

Классификация осуществляется на четыре

класса: «Пожар», «Задымление», «Растительность» и «Строения». Пример изображений для каждого класса приведен на рис. 4.



Рис. 4. Примеры участков изображений для идентификации и классификации

Для расширения набора данных, используемых для обучения моделей были использованы следующие аугментации (в случайных комбинациях):

- поворот изображения на углы 0, 90, 180 и 270 градусов;
- отражение по вертикали и горизонтали;
- масштабирование на случайную величину из диапазона [0.9; 2].

Описание подхода

Идентификация и классификация участка изображения разделяется на два этапа:

- 1) выделение информативных признаков, позволяющих идентифицировать содержимое изображения;
- 2) выбор класса, которому согласно выделенным признакам, максимально соответствует анализируемый участок изображения.

Основой используемого подхода к выделению признаков является использование нейросетевой модели автоэнкодера, который формирует дескриптор изображения в виде массива из 512 элементов (числа с плавающей запятой). Получаемый массив нормируется по L2-норме.

Автоэнкодеры состоят из двух частей: энкодера g и декодера f . Энкодер переводит входной сигнал в его представление (дескриптор) $h = g(x)$, а декодер восстанавливает сигнал по его коду $x = f(h)$. Автоэнкодер, изменяя g и f , стремится выучить тождественную функцию $x = f(g(x))$, минимизируя функционал ошибки $L(x, f(g(x)))$. При этом семейства функций энкодера g и декодера f ограничены так, чтобы автоэнкодер был вынужден отбирать наиболее важные свойства сигнала.

Кодировщик представляет собой набор из трех сверточных блоков с понижением размерности и одного полносвязного слоя. Каждый сверточный блок состоит из двух сверточных слоев (например, Conv2D_1.1 и Conv2D_1.2 для первого блока) и слоя понижения пространственной размерности

(например, MaxPooling2D_1 для первого блока). Сверточные слои выполняют функцию выделения признаков заданного размера и формирования карт признаков, по одной на каждый фильтр. Полносвязный слой выполняет функцию преобразования полученного набора карт признаков в одномерный вектор чисел с плавающей запятой.

Кодировщик представляет собой набор из трех сверточных блоков с понижением размерности и одного полносвязного слоя. Каждый сверточный блок состоит из двух сверточных слоев (например, Conv2D_1.1 и Conv2D_1.2 для первого блока) и слоя понижения пространственной размерности (например, MaxPooling2D_1 для первого блока). Сверточные слои выполняют функцию выделения признаков заданного размера и формирования карт признаков, по одной на каждый фильтр. Полносвязный слой выполняет функцию преобразования полученного набора карт признаков в одномерный вектор чисел с плавающей запятой.

Декодировщик также состоит из одного полносвязного слоя и трех сверточных блоков с повышением размерности. Каждый сверточный блок состоит из слоя повышения пространственной размерности (например, UpSampling2D_2 для первого блока) и двух сверточных слоев (например, Conv2D_4.1 и Conv2D_4.2 для первого блока).

Многозональное изображение может быть представлено в виде трехмерного куба, две оси которого имеют пространственную, а третья – спектральную размерность (непосредственно связанную с излучательной и отражательной характеристиками объекта) [2].

Для получения дескриптора участка исходного цветного изображения размером 128 x 128 пикселей достаточно подать это изображение на вход автоэнкодера и получить выходные значения указанного скрытого слоя. Т.к. вектор размерностью 512 отобразить не представляется возможным, то для вывода данных был использован алгоритм понижения размерности – метод главных компонент.

С помощью указанного метода размерность дескрипторов была уменьшена с 512 до двух, что уже может быть отображено на плоскости.

На рис. 5 приведен пример отображения

16 дескрипторов (для четырех различных групп изображений, включающих различные типы поверхностей).

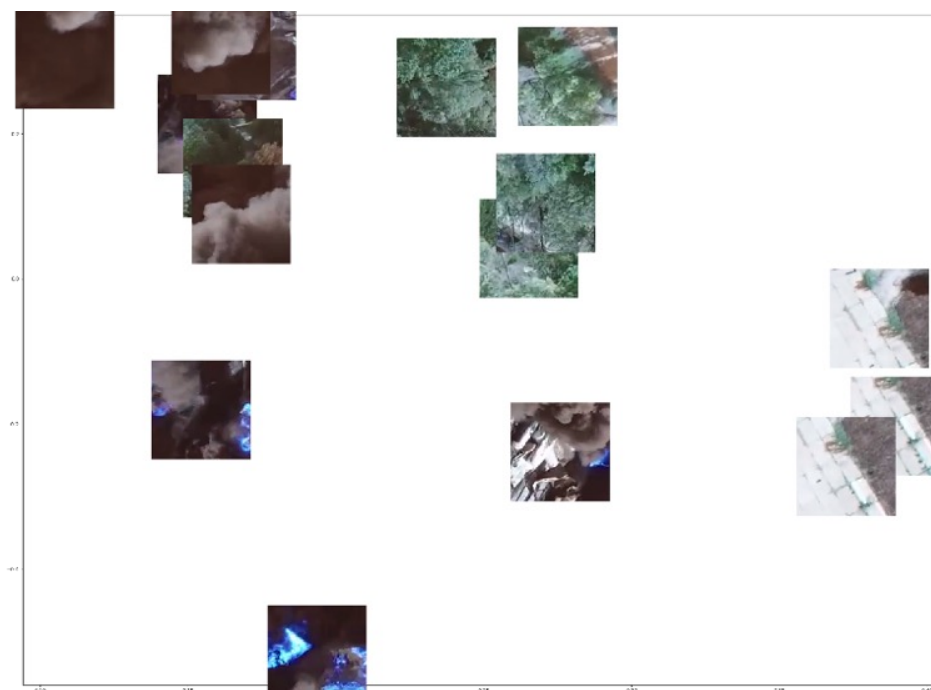


Рис.5. Расположение двумерных проекций дескрипторов на плоскости в виде миниатюр

Текстурные характеристики

Текстурные характеристики выражают разность значений интенсивности соседних пикселей изображения, при этом значения оценок

этой разности отличаются для различных типов подстилающей поверхности.

Примеры вычисленных текстурных характеристик приведены в Таблице 1.

Таблица 1. Примеры вычисленных текстурных характеристик

Объект	Характеристика	
	ASM	Contrast
Дым	5.262e-03	4.249e+01
Лес	2.215e-04	2.681e+02
Строения	3.915e-04	1.560e+02
Очаг возгорания	2.594e-03	1.656e+02

Вычисление текстурных характеристик снимка выполняется в так называемом «скользящем окне». Размер «скользящего окна» – 4×4 пикселя. Т. е. выбранные текстурные характеристики вычисляются для небольших участков изображения. Значения текстурных характеристик собираются в матрицы, которые преобразовываются к целочисленным значениям в диапазоне от 0 до 255. Полученная матрица представляет собой полутоновое изображение, которое позволяет сравнить значения оценок текстурных характеристик для различных участков изображений.

Текстурные характеристики могут вычисляться на основе полутоновых пространственно-зависимых матриц по каждому из четырех направлений – по горизонтали, по вертикали или по двум диагоналям, с учетом выбранного расстояния d , после чего эти четыре оценки усредняются для получения единой оценки.

Расстояние d используется при вычислении матрицы относительных частот $P(i, j, d, a)$ (вычисляется для пары соседних пикселей, разделенных расстоянием d по выбранному направлению a).

Для оценки текстуры используются

текстурные характеристики Харалика.

Мультимодальные данные

В литературе рассматривается ряд способов использования мультимодальных данных для глубоких нейронных сетей. В качестве данных различной модальности рассматриваются: цветные изображения, карта глубины, радарные снимки, цифровая модель рельефа, цифровая модель поверхности, данные ультразвуковых датчиков, текстовое описание, аудио- и видеопоследовательности, а также данные с иных спектральных датчиков и их производные признаки (вегетационные индексы) [3 – 5].

Работа с мультимодальными данными направлена на получение лучшего общего представления по сравнению с одномодальными. Поскольку прямая конкатенация извлеченных признаков из данных с различной модальностью часто приводит к неточным результатам, общее представление (дескриптор) может быть улучшено с использованием описанного выше глубокого автоэнкодера. Для этого необходимо, чтобы используемые данные в каждой модальности могли быть представлены в виде карты. Также необходимо согласование координат для данных карт, то есть они должны быть представлены в виде многозонального изображения.

Далее размер входного слоя преобразуется к размеру $128 \times 128 \times N$, где N – количество карт различной модальности. А для выходного сверточного слоя Conv2D_6.2 количество фильтров $F_c = N$. Обучение модели и получение дескриптора не отличается от подхода, описанного выше.

Вариативные информативные признаки

Построение дескрипторов объектов основано на использовании комбинированных информативных признаков многоспектральных изображений. Важной характеристикой признаков является вариативность.

Вариативность – свойство и мера изменчивости объектов, их числовых и нечисловых характеристик, состояний и связей. Локальная вариативность информативных признаков может оцениваться только их текстурными характеристиками.

Алгоритм построения дескрипторов объектов с использованием обученного автоэнкодера включает следующие шаги.

Шаг 1. Получение изображения изображений ДЗЗ, которые включают ближний инфракрасный спектр.

Шаг 2. Вычисление карт информативных признаков многоспектральных изображений (вегетационные индексы NDVI и цветовые признаки).

Шаг 3. Вычисление карт характеристик вариативности (текстурные характеристики Харалика и фрактальные характеристики).

Шаг 4. Комбинирование полученных карт в одну многомерную карту информативных признаков.

Шаг 5. Разбиение многомерной карты на перекрывающиеся участки размером 128×128 .

Шаг 6. Вычисление дескрипторов с

использованием автоэнкодера.

При выделении информативных признаков дополнительно добавляется цветовая информация, инвариантная к масштабу и поворотам изображения. Для этого для каждого обрабатываемого участка изображения строится его поканальная гистограмма. При этом для сокращения объема данных применяется квантование значений яркости не на 256 уровней, а на 16. Далее полученные гистограммы объединяются в один массив размером 48 элементов: [16 значений синего канала, 16 значений зеленого канала, 16 значений красного канала]. Полученный массив также нормируется по L2-норме.

Таким образом, результатом идентификации является массив из 560 элементов, который и используется для классификации.

Для классификации разработана нейросетевая модель, состоящая из двух многослойных перцептронов, объединенных в виде нейросетевого ансамбля. Каждый перцептрон в отдельности анализирует входной массив данных, после чего выходные слои формируют взвешенное общее решение по классу входных данных. Сама модель реализована на языке программирования Python с использованием библиотеки Keras [6].

Таким образом, в целом разработанный подход можно представить в виде следующих шагов.

Шаг 1. Берем участок цветного изображения размером 128×128 пикселей.

Шаг 2. Для этого изображения строим дескриптор размером 512 элементов, используя разработанный автоэнкодер.

Шаг 3. Для этого же изображения строим набор дополнительных признаков, описывающих цветовые характеристики изображения, представляющий собой массив из 48 элементов, полученный из поканальной гистограммы, как описано ниже.

Шаг 4. Объединяем полученные на шагах 2 и 3 массивы в один размерностью 560.

Шаг 5. Полученный объединенный массив отправляем на вход модели классификации, в результате исполнения которой получаем массив из четырех чисел, которые указывают на степень близости изображенного тому или иному классу.

Шаг 6. Индекс, соответствующий максимальному значению массива, полученного на шаге 5, и является искомым индексом класса.

Обучение моделей

Обучение модели классификации:

Функция потерь (loss) – categorical cross entropy [7].

- размер обучающей выборки: 4000 изображений (не включая аугментацию);

- размер валидационной выборки: 800 изображений (не включая аугментацию);

- результирующая точность на валидационной выборке: 99,19 %;

- обучающий алгоритм: Adam [8] с параметрами: $lr = 0.0001$, $beta_1 = 0.9$, $beta_2 = 0.999$, $epsilon = 1 \cdot 10^{-8}$, $decay = 0.0$.

Тестирование

Тестирование осуществлялось на валидационной выборке. При этом оценивалась точность как для каждого класса в отдельности, так и для всех классов в целом. Полученные результаты тестирования приведены в Таблице 2.

Из-за несбалансированности классов

в исходных данных требуется дополнительная оценка. Данные, полученные в результате, сведены в матрицы неточностей (confusion matrix), приведенные в Таблице 3. Значения в матрицах приведены в относительных величинах – количество объектов принадлежащих классу к общему количеству объектов всех классов в выборке.

Таблица 2. Результаты тестирования сегментаторов

Классы	Точность, %
Задымление	86.9
Растительность	99.7
Строения	100
Пожар	98.9
Общая точность	99.19

Таблица 3. Матрицы неточностей

Предсказанные классы	Реальные классы			
	Задымление	Растительность	Строения	Пожар
Задымление	0.285	0.003	0.039	0.001
Растительность	0.000	0.222	0.001	0.000
Строения	0.000	0.000	0.224	0.000
Пожар	0.000	0.002	0.001	0.224

Для оценки качества классификации были вычислены соответствующие значения точности (precision), полноты (recall) и F1-меры [8] (см. Таблицу 4).

Таблица 4. Значения точности (precision), полноты (recall) и F1-меры

Классы	Precision	Recall	F1
Задымление	0.869	1.000	0.930
Растительность	0.997	0.981	0.989
Строения	1.000	0.848	0.918
Пожар	0.989	0.994	0.991

Наибольшее количество ошибок возникло на участках, соответствующих границе задымления и пожара, а также в случае горения строений.

Выводы

Для идентификации и классификации объектов на изображениях, полученных с помощью съемочных средств БПЛА и орбитального базирования, предложена нейросетевая модель классификации, основанная на использовании оригинального автоэнкодера и построенная по архитектуре ансамбля многослойных перцептронов.

Получаемые дескрипторы представляют собой массив из 512 чисел с плавающей запятой. В качестве основы для построения дескрипторов использован нейросетевой автоэнкодер, архитектура которого зависит от модальности и цветности исследуемых данных. Отличительными чертами разработанных моделей является использование как многомодальных и разноцветных данных, так

и использование дополнительных информативных признаков вариативности данных. Для оценки вариативности данных использованы текстурные признаки Харалика, построенные для локальных участков изображения. При выделении информативных признаков дополнительно добавляется цветовая информация, инвариантная к масштабу и поворотам изображения и основанная на построении поканальных гистограмм.

Модель реализована с использованием библиотеки Keras. Использование предложенной модели для классификации на четыре класса: «Пожар», «Задымление», «Растительность» и «Строения», позволяет достичь точности классификации выше 99 %.

Работа выполнена при поддержке БРФФИ (договор № Ф21УКРГ-007 от 30.04.2021).

ЛИТЕРАТУРА

1. **Sentinel Hub EO Browser** [Electronic resource]. – Mode of access: https://apps.sentinel-hub.com/eo-browser/?zoom=13&lat=50.20921&lng=30.23931&themeId=DEFAULT-THEME&visualizationUrl=https://services.sentinel-hub.com/ogc/wms/e35192fe-33a1-41f3-b798-b755e771c5a5&datasetId=AWS_LOTL1&fromTime=2015-06-09T00:00:00.000Z&toTime=2015-06-09T23:59:59.999Z&layerId=1_TRUE_COLOR/ – Date of access: 28.05.2022.
2. **Орешкина Л. В.** Обнаружение и распознавание класса объектов на многозональных изображениях дистанционного зондирования / Л. В. Орешкина // Информатика. – 2019. – №. 2 (6). – С. 79-85.
3. **R. Kuga [et al.]** Multi-task learning using multi-modal encoder-decoder networks with Kuga R. Multi-task learning using multi-modal encoder-decoder networks with shared skip connections / R. Kuga [et al.] // Proceedings of the IEEE International Conference on Computer Vision Workshops. – 2017. – С. 403-411.
4. **Chen K.** Deep Semantic Segmentation of Aerial Imagery Based on Multi-Modal Data / K. Chen [et al.] // IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. – 2018. – С. 6219-6222.
5. **Feng D. [et al.]** Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges // IEEE Transactions on Intelligent Transportation Systems. – 2020.
6. **Kingma, D. P.** Adam: A Method for Stochastic Optimization / D. P. Kingma, J. Ba // 3rd Intern. Conf. for Learning Representations, 7–9 May, 2015, San Diego. – San Diego, 2015. – Vol. 1412. – P. 6980.
7. **Sokolova, M.** Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation / M. Sokolova, N. Japkowicz, S. Szpakowicz // Advances in Artificial Intelligence : 19th Australian Joint Conf. on Artificial Intelligence, Hobart, 4–8 Dec., 2006. – Hobart, 2006. – P. 1015–1021.

REFERENCES

1. **Sentinel Hub EO Browser** [Electronic resource]. – Mode of access: https://apps.sentinel-hub.com/eo-browser/?zoom=13&lat=50.20921&lng=30.23931&themeId=DEFAULT-THEME&visualizationUrl=https://services.sentinel-hub.com/ogc/wms/e35192fe-33a1-41f3-b798-b755e771c5a5&datasetId=AWS_LOTL1&fromTime=2015-06-09T00:00:00.000Z&toTime=2015-06-09T23:59:59.999Z&layerId=1_TRUE_COLOR/ – Date of access: 28.05.2022.
2. **Oreshkina L.V.** Object class detection and recognition in multi-zone remote sensing images / L.V. Oreshkina // Informatics. – 2019. – №. 2 (6). – С. 79-85. [In russian]
3. **Kuga R.** Multi-task learning using multi-modal encoder-decoder networks with shared skip connections / R. Kuga [et al.] // Proceedings of the IEEE International Conference on Computer Vision Workshops. – 2017. – С. 403-411.
4. **Chen K.** Deep Semantic Segmentation of Aerial Imagery Based on Multi-Modal Data / K. Chen [et al.] // IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. – 2018. – С. 6219-6222.
5. **Feng D. [et al.]** Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges // IEEE Transactions on Intelligent Transportation Systems. – 2020.
6. **Kingma, D. P.** Adam: A Method for Stochastic Optimization / D. P. Kingma, J. Ba // 3rd Intern. Conf. for Learning Representations, 7–9 May, 2015, San Diego. – San Diego, 2015. – Vol. 1412. – P. 6980.
7. **Sokolova, M.** Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation / M. Sokolova, N. Japkowicz, S. Szpakowicz // Advances in Artificial Intelligence : 19th Australian Joint Conf. on Artificial Intelligence, Hobart, 4–8 Dec., 2006. – Hobart, 2006. – P. 1015–1021.

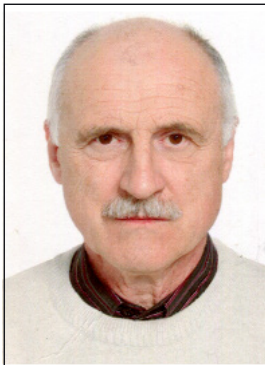
DOUDKIN A. A., **GANCHENKO V. V.**, **INYUTIN A. V.**, **MARUSHKO E. E.**

IDENTIFICATION AND CLASSIFICATION OF OBJECTS IN IMAGES OBTAINED BY UAV AND ORBITAL BASE IMAGING EQUIPMENT

The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

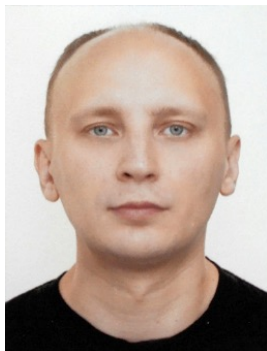
To identify and classify objects on images obtained using UAV imaging and orbital-based imaging, a neural network classification model based on the use of an autoencoder and built on the architecture of an ensemble of multilayer perceptrons is proposed. Additionally, at the stage of highlighting informative features, is added a color information, which is based on the per-channel histograms and is invariant to the scale and rotations of the image. The model is implemented using the Keras library. The use of the proposed model for classification into four classes: “Fire”, “Smoke”, “Vegetation” and “Buildings”, allows to achieve a classification accuracy above 99%.

Keywords: metal fracture; texture features; macrogeometric descriptors.



Дудкин Александр Арсентьевич, (13.10.1950 – 16.10.2022) доктор технических наук, профессор – ученый в области технической кибернетики и информатики, области научных интересов: цифровая обработка сигналов и изображений; распознавание образов; архитектуры и модели систем компьютерного зрения и высокопроизводительной обработки информации. Автор более 300 научных работ, включая 3 монографии и 90 статей.
Doudkin A., prof., research interests – digital signal and image processing, pattern recognition; architectures and models of computer vision systems and high-performance information processing.

E-mail: doudkin@newman.bas-net.by



Ганченко Валентин Вячеславович, кандидат технических наук, область научных интересов – цифровая обработка сигналов, дистанционное зондирование.

Ganchenko V., PhD, research interests – digital signal processing, remote sensing.

E-mail: ganchenko@lsi.bas-net.by



Инютин Александр Владимирович, область научных интересов – обработка изображений и распознавание образов в системах компьютерного зрения, обработка изображений дистанционного зондирования.

Inyutin A., research interests – image processing and pattern recognition in computer vision systems, remote sensing image processing.

E-mail: avin@lsi.bas-net.by



Марушко Евгений Евгеньевич, область научных интересов – обработка изображений и распознавание образов в системах компьютерного зрения, обработка изображений дистанционного зондирования, обработка телеметрических данных в системах управления космическими аппаратами.

Marushko Y., research interests – image processing and pattern recognition in computer vision systems, remote sensing image processing, telemetry processing in spacecraft control systems.

E-mail: marushko@lsi.bas-net.by