

ISSN 1561-2430 (Print)

ISSN 2524-2415 (Online)

УДК 004.932:528.8

<https://doi.org/10.29235/1561-2430-2026-62-2-164-176>

Поступила в редакцию 04.06.2025

Received 04.06.2025

**А. А. Интякова<sup>1</sup>, С. В. Абламейко<sup>1,2</sup>**<sup>1</sup>*Белорусский государственный университет, Минск, Республика Беларусь*<sup>2</sup>*Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Республика Беларусь***АНАЛИЗ МНОГОКАНАЛЬНЫХ СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ  
ГОРОДСКИХ ТЕРРИТОРИЙ МЕТОДАМИ ГЛУБОКОГО ОБУЧЕНИЯ**

**Аннотация.** Рассмотрена проблема классификации растительного покрова в плотно застроенной урбанизированной среде по изображениям дистанционного зондирования Земли сверхвысокого разрешения. Целью исследования является разработка и валидация нейросетевого алгоритма для высокоточной семантической сегментации городских зеленых зон с использованием мультимодальных данных. Предложена усовершенствованная архитектура сверточной нейронной сети U-Net, модифицированная для работы с 7-канальным входным тензором (RGB, NIR, RedEdge, DSM, NDVI). В основе подхода лежит стратегия раннего слияния (Early Fusion), объединяющая спектральные измерения с перспективными структурными характеристиками (цифровой моделью поверхности, DSM). Для преодоления дисбаланса классов применена функция Focal Loss. Предложенная модель позволила надежно разделить спектрально идентичные ярусы (трава и деревья) и исключить ложные срабатывания на зеленых антропогенных объектах. Итоговый показатель Mean IoU составил 0,725. Полнота (Recall) обнаружения лесных массивов достигла 0,93, а сложного миноритарного класса «Кустарники» – 0,77. Эксперимент на независимой тестовой территории подтвердил высокую способность модели к генерализации (F1-score 0,93). Комплексирование 7 каналов информации и применение модифицированной U-Net полностью оправдано для задач точного расчета площадей насаждений и экологического мониторинга в концепции «Умный город».

**Ключевые слова:** семантическая сегментация, городские зеленые зоны, дистанционное зондирование Земли, глубокое обучение, сверточные нейронные сети, U-Net, цифровая модель поверхности, мультиспектральные изображения, пространственная кросс-валидация

**Для цитирования.** Интякова, А. А. Анализ многоканальных спутниковых изображений городских территорий методами глубокого обучения / А. А. Интякова, С. В. Абламейко // Весці Нацыянальнай акадэміі навук Беларусі. Сэрыя фізіка-матэматычных навук. – 2026. – Т. 62, № 2. – С. 164–176. <https://doi.org/10.29235/1561-2430-2026-62-2-164-176>

**Anastasia A. Intyakova<sup>1</sup>, Sergey V. Ablameyko<sup>1,2</sup>**<sup>1</sup>*Belarusian State University, Minsk, Republic of Belarus*<sup>2</sup>*United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus***ANALYSIS OF MULTICHANNEL SATELLITE IMAGES OF URBAN AREAS  
USING DEEP LEARNING METHODS**

**Abstract.** This paper examines the problem of land cover classification in densely populated urban environments using ultra-high-resolution Earth observation images. The aim of the study is to develop and validate a neural network algorithm for a high-precision semantic segmentation of urban green spaces using multimodal data. An improved U-Net convolutional neural network architecture, modified for the use with a 7-channel input tensor (RGB, NIR, RedEdge, DSM, and NDVI), is proposed. The approach is based on the Early Fusion strategy, which combines spectral measurements with promising structural characteristics (digital surface model, DSM). Focal Loss is used to overcome the class imbalance. The proposed model reliably separated spectrally identical layers (grass and trees) and eliminated false positives on green anthropogenic objects. The final Mean IoU was 0.725. The recall for detecting forested areas reached 0.93, and for the complex minority class “Shrubs” it reached 0.77. The experiment on an independent test site confirmed the model’s high generalizability (F1-score 0.93). The integration of seven data channels and the use of a modified U-Net are fully justified for the tasks of accurate calculating forest areas and environmental monitoring in the Smart City concept.

**Keywords:** semantic segmentation, urban green spaces, Earth remote sensing, deep learning, convolutional neural networks, U-Net, digital surface model, multispectral imagery, spatial cross-validation

**For citation.** Intyakova A. A., Ablameyko S. V. Analysis of multichannel satellite images of urban areas using deep learning methods. *Vestsi Natsyyanal'nai akademii navuk Belarusi. Seryya fizika-matematychnykh navuk = Proceedings of the National Academy of Sciences of Belarus. Physics and Mathematics series*, 2026, vol. 62, no. 2, pp. 164–176 (in Russian). <https://doi.org/10.29235/1561-2430-2026-62-2-164-176>

**Введение.** Современные методы мониторинга городской среды базируются на комплексировании разнородных данных дистанционного зондирования Земли (ДЗЗ). Для задач семантической сегментации растительного покрова критически важным является использование не только оптического диапазона, но и специализированных спектральных каналов, а также информации о пространственной структуре объектов [1].

Задача выделения растительного покрова на спутниковых снимках относится к классу задач семантической сегментации (Semantic Segmentation) [2]. В отличие от классификации изображений, где всему снимку присваивается одна метка, сегментация требует присвоения метки класса каждому пикселю изображения, формируя маску, топологически соответствующую входным данным. Учет пространственного контекста имеет решающее значение для классификации землепользования, что делает применение глубокого обучения наиболее перспективным подходом [3].

Современные модели ориентированы на баланс между высоким пространственным разрешением и вычислительной эффективностью. В литературе выделяются три ключевые группы архитектур:

- модели на базе сети U-Net, которые являются наиболее распространенными для данной задачи [4; 5], например, архитектура SA-SC-U-Net (Self-Attention with Separable Convolutions), являющаяся одной из ведущих по состоянию на 2026 г. (точность – до 91,78 %, IoU – 81,82 %), использует механизм самовнимания для захвата глобального контекста [6];

- модели на основе трансформеров: архитектуры типа SegFormer и CTSA-DeepLabV3+ все чаще применяются благодаря способности моделировать глобальные контекстные связи, что критически важно для идентификации разрозненных зеленых зон [7];

- гибридные модели: объединяют сверточные нейронные сети (CNN) с трансформерами для задач сверхвысокого разрешения (например, SR-UGSnet), эффективно преобразуя данные низкого разрешения в детализированные карты [8].

Традиционные архитектуры CNN, такие как AlexNet, VGG, имеют структуру кодировщика, который последовательно уменьшает размерность изображения, выделяя абстрактные признаки. Однако завершающие полносвязные слои утрачивают информацию о пространственном расположении объектов, что неприемлемо для задач сегментации [2]. Прорывом стало появление полносверточных сетей (Fully Convolutional Networks – FCN), где полносвязные слои заменены на сверточные, а для восстановления разрешения используются операции транспонированной свертки [2]. Однако FCN страдали от низкой точности на границах объектов из-за потери деталей в процессе пулинга. Дальнейшим развитием этой идеи стала архитектура U-Net, предложенная в 2015 г. [9]. Изначально разработанная для биомедицины, она показала выдающиеся результаты в ДЗЗ благодаря способности обучаться на малых выборках и высокой точности локализации границ.

Очевидно, что выбор архитектуры напрямую влияет на точность конечных результатов, особенно в условиях ограниченной разметки. В данной работе предлагается усовершенствованная модификация U-Net, адаптированная для работы с 7-канальными мультимодальными данными (RGB + NIR + RedEdge + DSM + NDVI) с целью высокоточной семантической сегментации городских зеленых зон.

**Выбор данных для задачи сегментации городских зеленых зон.** Специфика задачи сегментации городских зеленых зон (ГЗЗ) на изображениях дистанционного зондирования требует комплексного анализа спектральных и высотных данных. В городском планировании ГЗЗ определяются как открытые пространства, включающие парки, уличные деревья, флору, водные объекты и ухоженные газоны.

Существует множество способов классификации показателей ГЗЗ на изображениях ДЗЗ [10]. В рамках данного исследования мы исходим из гипотезы, что для высокоточной сегментации недостаточно использовать только плоские спектральные изображения, полученные в результате съемки сверху (со спутников или БПЛА). Отличительной чертой нашего подхода является комплексирование плановых измерений (вид сверху) с пространственно-структурными характеристиками, имитирующими перспективный (наземный) обзор.

Традиционный вид сверху представляет собой меру спектральной яркости, непосредственно полученную из оптического (RGB) и ближнего инфракрасного (NIR) диапазонов. Это позволяет надежно идентифицировать наличие фотосинтезирующей биомассы [11]. Однако такой подход не дает представления о вертикальной структуре объектов, которую мы обычно воспринимаем, находясь на земле (перспективный вид). Для воссоздания этого структурного измерения в анализ интегрируются трехмерные цифровые модели поверхности (DSM). Слой DSM представляет собой растровое изображение, кодирующее абсолютные высоты объектов над уровнем моря, что позволяет алгоритму буквально «видеть» объем и высоту растительности.

Важным аспектом является пространственное разрешение снимков. Изображение с высоким пространственным разрешением (в нашем случае – 8 см/пиксель) дает возможность более детально рассматривать зеленые объекты, что критически важно для различения мелких элементов городской среды – отдельно стоящих кустарников, узких газонов. В условиях плотной застройки пространственное разрешение зачастую играет более важную роль, чем частотное (спектральное), так как базовые зеленые объекты уверенно идентифицируются в комбинации видимого (RGB) и инфракрасного (NIR) диапазонов [11].

Набор данных, используемый в исследовании, состоит из двух сопряженных растровых слоев: цифрового ортофотоплана и цифровой модели поверхности (DSM). Мультиспектральный ортофотоплан представлен в формате GeoTIFF (16 бит) и содержит 5 спектральных каналов. В свою очередь, слой цифровой модели поверхности представляет собой одноканальный растр в формате 32-bit Float, содержащий абсолютные значения высот над уровнем моря.

Для максимального раскрытия потенциала нейросети исходное признаковое пространство было расширено. Итоговый входной тензор для задачи сегментации городских зеленых зон включает 7 каналов:

- 1) 3 канала RGB (Red, Green, Blue): обеспечивают базовый визуальный контекст и информацию о цвете объектов в видимом диапазоне;
- 2) NIR (ближний инфракрасный диапазон): ключевой канал для идентификации здоровой растительности за счет высокой отражательной способности клеточной структуры листьев;
- 3) RedEdge (крайний красный диапазон): переходная зона между поглощением хлорофилла (красный) и рассеянием листа (NIR), канал высокочувствителен к состоянию растений и помогает отличать разные типы зелени;
- 4) DSM (цифровая модель поверхности – высота): обеспечивает структурную информацию (перспективный вид), позволяя разделить спектрально идентичные классы по ярусам (например, высокие деревья и низкий газон);
- 5) NDVI (нормализованный вегетационный индекс): расчетный физический признак, подаваемый в сеть в явном виде.

Такое расширение и использование семи каналов позволяет нейросети одновременно анализировать данные о высоте объектов для геометрического уточнения их границ и усиленные спектральные признаки (RedEdge и NDVI) для максимально надежного детектирования хлорофилла на фоне искусственных городских поверхностей.

**Модифицированная модель U-Net для семантической сегментации многоканальных изображений. Предлагаемая модель U-Net.** Классическая архитектура U-Net изначально проектировалась для работы с одноканальными (медицинскими) или трехканальными (RGB) изображениями [9]. Для текущей задачи сегментации растительности входной слой U-Net был модифицирован для приема семи каналов вместо стандартных трех.

Стандартный сверточный слой в библиотеках глубокого обучения (PyTorch) ожидает на входе 4-мерный тензор размерности  $(B, C, H, W)$ , где  $B$  – размер батча;  $C$  – количество каналов;  $H$  и  $W$  – высота и ширина изображения.

В данной работе размерность входного тензора по оси каналов увеличена с  $C = 3$  до  $C = 7$ . Это потребовало модификации весов первого сверточного блока кодировщика (Encoder).

Входной слой сети DoubleConv был переконфигурирован следующим образом:

$$\text{Input : (Batch, 7, 256, 256)} \xrightarrow{\text{Conv2d}} (\text{Batch, 64, 256, 256}). \quad (1)$$

Это означает, что ядра свертки (фильтры) первого слоя имеют глубину 7. Такое архитектурное решение реализует стратегию раннего слияния признаков (Early Fusion). В отличие от раздельной обработки спектра и высоты двумя независимыми сетями, ранняя интеграция позволяет нейросети уже на первом этапе обучения выявлять сложные нелинейные зависимости между разнородными данными. Например, сеть получает возможность обучить специфические фильтры, которые активируются только при одновременном выполнении условий:

- высокое значение в каналах NIR и NDVI (спектральный признак живой биомассы);
- низкое значение в канале DSM (геометрический признак объекта на уровне земли).

Такая комбинация признаков однозначно идентифицирует класс «Газон» или «Трава», позволяя отличать его от спектрально похожих «Зеленых крыш» (где NDVI высокий, но и DSM высокий) или от искусственных покрытий стадионов (где DSM низкий, но и NDVI низкий).

**Обучение модели.** Эффективность нейросетевой модели определяется не только ее архитектурой, но и выбранной стратегией обучения. Специфика спутниковых данных – высокая вариативность сцен, наличие шумов и выраженная гетерогенность объектов – требует применения специализированных методов настройки гиперпараметров и адаптации функции потерь.

В задачах семантической сегментации урбанизированных территорий наблюдается фундаментальная проблема *дисбаланса классов* (Class Imbalance Problem) [10]. Статистический анализ масок обучающей выборки показал крайне неравномерное распределение пикселей объектов:

- объекты искусственного происхождения (класс «Не растительность») и фоновые значения занимают доминирующее положение (мажоритарные классы);
- целевые классы растительности, особенно «Умеренная растительность» (кустарники), представлены разреженно и часто составляют менее 5–10 % от общей площади снимка (миноритарные классы).

При использовании стандартной функции потерь (*Cross Entropy Loss*) нейронная сеть стремится минимизировать глобальную ошибку по пути наименьшего сопротивления – предсказывая мажоритарный класс для всех пикселей. Это приводит к смещению модели (*Bias*) и неспособности детектировать редкие объекты.

На начальных этапах исследования применялась взвешенная перекрестная энтропия, однако эксперименты показали, что простого взвешивания недостаточно для уверенной классификации сложных примеров (Hard Mining). Для решения этой проблемы была внедрена функция потерь Focal Loss [12]. В отличие от стандартной энтропии, Focal Loss добавляет модулирующий множитель  $(1 - p_i)^\gamma$ , который уменьшает вклад простых, хорошо классифицируемых примеров (таких как большие участки асфальта или фона) и фокусирует обучение на трудных примерах (границы объектов, разреженные кустарники).

Формула функции потерь имеет вид

$$FL(p_i) = -\alpha_i(1 - p_i)^\gamma \log(p_i), \quad (2)$$

где  $p_i$  – вероятность правильного класса;  $\gamma = 0,2$  – параметр фокусировки (focusing parameter), регулирующий степень подавления простых примеров;  $\alpha_i$  – весовой коэффициент класса.

Для балансировки вклада классов были применены веса  $\alpha_i$ , рассчитанные методом обратной частоты с ручной корректировкой:

- класс «Фон»: вес снижен до 0,1, чтобы минимизировать влияние технических границ снимка (padding);

– класс «Кустарники»: присвоен максимальный вес 1,69, что вынуждает модель активно корректировать веса даже при единичных ошибках на этом классе.

В качестве алгоритма оптимизации выбран Adam (Adaptive Moment Estimation) [13]. В отличие от стохастического градиентного спуска (SGD), Adam адаптирует скорость обучения индивидуально для каждого параметра, используя оценки первого и второго моментов градиентов, что обеспечивает быструю сходимость на сложной поверхности функции потерь.

Для предотвращения переобучения (*Overfitting*) внедрен комплекс методов регуляризации:

– L2-регуляризация (Weight Decay) – в оптимизатор добавлен штраф  $\lambda = 1 \cdot 10^{-4}$ , пропорциональный квадрату весов, это ограничивает рост параметров, делая границы принятия решений более плавными и снижая чувствительность к шуму;

– Dropout (слой исключения) – перед выходным классификатором добавлен слой Dropout ( $p = 0,5$ ), который случайно обнуляет 50 % нейронов на каждой итерации, это эквивалентно обучению ансамбля прореженных сетей и предотвращает коадаптацию признаков;

– ранняя остановка (Early Stopping): обучение прекращается, если ошибка на валидации не уменьшается в течение 15 эпох; сохраняются веса модели, показавшей наилучший результат (Best Model Checkpoint), а не веса последней эпохи.

Реализована динамическая стратегия *Learning Rate Scheduling* с помощью планировщика ReduceLROnPlateau. Если метрика ошибки (*Val Loss*) не улучшается в течение 5 эпох (*patience*), скорость обучения автоматически снижается в 2 раза. Это позволяет модели на финальных этапах совершать малые шаги для тонкой настройки весов в окрестности глобального минимума.

**Экспериментальные исследования. Предварительная обработка набора данных.** Используемая в работе гетерогенная структура данных (разная разрядность, физический смысл и диапазоны значений) требует применения специализированной стратегии нормализации, так как использование входных данных с большим разбросом значений приводит к нестабильности градиентов при обратном распространении ошибки. Для приведения всех каналов к единому диапазону [0, 1] была реализована многоступенчатая процедура предварительной обработки.

В первую очередь проводится нормализация спектральных каналов. Поскольку исходные данные имеют 16-битную глубину ( $N = 16$ ), простое конвертирование в 8 бит привело бы к потере радиометрической точности в канале RedEdge. Поэтому применяется линейное масштабирование значений цифровых отсчетов (DN) по формуле

$$X_{\text{norm}} = \frac{X_{\text{raw}}}{2^N - 1} = \frac{X_{\text{raw}}}{65535,0}, \quad (3)$$

где  $X_{\text{raw}}$  – исходное значение пикселя;  $N = 16$  – битовая глубина. Данная операция выполняется для всех пяти спектральных каналов, сохраняя относительные спектральные характеристики объектов.

Параллельно с этим обрабатывается канал высоты (DSM), который имеет иную природу распределения. Для него применяется стратегия Min-Max нормализации с предварительной фильтрацией выбросов (значений  $H < -100$ , возникающих из-за ошибок сенсора). Масштабирование производится на основе глобальных статистик сцены ( $H_{\text{min}} = -10,0$  м,  $H_{\text{max}} = 10,0$  м) по формуле

$$H_{\text{norm}} = \frac{H_{\text{curr}} - H_{\text{min}}}{H_{\text{max}} - H_{\text{min}}}. \quad (4)$$

В случае, если перепад высот на снимке отсутствует ( $H_{\text{max}} = H_{\text{min}}$ ), канал заполняется нулями для предотвращения деления на ноль.

Дополнительно, для улучшения сепарабельности растительности, во входной тензор был добавлен расчетный канал вегетационного индекса NDVI (Normalized Difference Vegetation Index). Он вычисляется на основе нормализованных каналов NIR и Red как

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}. \quad (5)$$

Полученные значения, лежащие в диапазоне  $[-1, 1]$ , линейно трансформируются в диапазон  $[0, 1]$ . Итоговый входной тензор формируется путем конкатенации спектральных каналов, нормализованной высоты и индекса NDVI, образуя массив размерности  $(7, H, W)$ .

Поскольку исходные снимки имеют размеры, превышающие объем видеопамати графического ускорителя, применяется метод скользящего окна (Sliding Window). Изображение разбивается на фрагменты фиксированного размера  $256 \times 256$  пикселей. Реализована стратегия нарезки с перекрытием (Overlap Strategy), где шаг окна (*stride*) составляет 128 пикселей. Это обеспечивает 50%-ное перекрытие соседних патчей, что позволяет увеличить объем обучающей выборки в 4 раза и устраняет проблему краевых эффектов на границах фрагментов.

Для предотвращения переобучения модели применяется аугментация данных «на лету» (On-the-fly augmentation). Учитывая инвариантность аэрофотоснимков к ориентации, используются геометрические трансформации: случайные повороты на  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  и зеркальные отражения по горизонтали и вертикали. Каждая трансформация применяется с вероятностью  $p = 0,5$ , что обеспечивает модели высокую вариативность примеров.

**Постобработка и визуализация.** Поскольку исходные ортофотопланы имеют разрешение, значительно превышающее объем видеопамати (VRAM) графического ускорителя, прямая подача всего снимка в сеть невозможна. Для решения этой задачи был разработан программный модуль `predict_smooth.py`, реализующий метод скользящего окна со сглаживанием (Gaussian Blending Sliding Window).

Алгоритм обработки:

- *декомпозиция* – исходные 7-канальные данные (RGB + NIR + RedEdge + DSM + NDVI) считываются с диска, изображение разбивается на фрагменты (патчи) размером  $256 \times 256$  пикселей с шагом 128 (перекрытие 50 %);

- *взвешивание* – для устранения эффекта «шахматной сетки» на границах патчей применяется гауссовское сглаживание, пикселям в центре патча присваивается больший вес, чем пикселям на краях, что обеспечивает бесшовную склейку;

- *предикция и сборка* – нейронная сеть генерирует карты вероятностей для каждого класса, которые затем объединяются в единый растровый массив размером с исходный снимок.

Результат работы алгоритма экспортируется в формат GeoTIFF с сохранением системы координат (CRS), что позволяет использовать полученную карту непосредственно в ГИС-системах (QGIS, ArcGIS). Результат семантической сегментации представлен на рис. 1.

Визуальный анализ демонстрирует высокую эффективность предложенного подхода:

- эффективное разделение ярусов: алгоритм уверенно различает высокие деревья (темно-зеленые зоны) и низкорослую травянистую растительность, что достигается благодаря каналу DSM – объекты с идентичным спектральным откликом, но с разной высотой отнесены к разным классам;

- геометрическая точность: в отличие от ранних экспериментов, финальная модель корректно оконтуривает здания и дороги (класс «Асфальт»), не допуская ложных срабатываний на зеленых крышах, что подтверждается высоким показателем IoU для антропогенных объектов (0,86).

**Сравнительный анализ результатов. Метрики оценки качества.** Поскольку в спутниковых снимках часто наблюдается дисбаланс классов (например, площадь асфальта значительно превышает площадь кустарников), стандартная метрика *Accuracy* (доля правильных ответов) не является показательной. Для объективной оценки качества семантической сегментации используются специализированные метрики.

Intersection over Union (IoU, индекс Жаккара) является основным стандартом в задачах сегментации. Она рассчитывается как отношение площади пересечения предсказанной области и истинной разметки (*Ground Truth*) к площади их объединения:

$$\text{IoU} = \frac{A \cap B}{A \cup B} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \quad (6)$$



Рис. 1. Карта классификации растительного покрова, полученная с помощью разработанной нейросети

Fig. 1. Vegetation classification map generated using the developed neural network

где TP (True Positive) – верно классифицированные пиксели целевого класса (например, дерево); FP (False Positive) – фон, ошибочно классифицированный как дерево; FN (False Negative) – дерево, ошибочно классифицированное как фон.

$F1$ -score (коэффициент Дайса) представляет собой гармоническое среднее между точностью (Precision) и полнотой (Recall):

$$F1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN}. \quad (7)$$

Использование комплекса этих метрик позволяет объективно оценить, насколько точно нейронная сеть выделяет контуры растительности и минимизирует ложные срабатывания на антропогенных объектах. В итоговом анализе используется усреднение метрик по всем классам (Macro-Average), что исключает доминирование мажоритарных классов (фона и асфальта) в общей оценке.

Для оценки качественного прироста точности было проведено визуальное сравнение карт, полученных в среде QGIS, а также количественный анализ на основе вычисленных метрик. Сравнивались результаты традиционного метода пороговой классификации индекса NDVI (Normalized Difference Vegetation Index), пиксельной классификации и разработанной нейросетевой модели U-Net.

**Сравнение с индексным методом (NDVI).** Анализ фрагментов карты позволил выявить ключевые недостатки индексного метода, которые были успешно устранены в текущей работе.

1. *Проблема ложных срабатываний на антропогенных объектах:*

– NDVI – традиционный метод, основанный исключительно на спектральной яркости, часто ошибочно классифицировал крыши зданий (особенно окрашенные в зеленый цвет или имеющие шероховатую текстуру) как газоны, это связано с тем, что старые материалы кровли могут иметь спектральный отклик, близкий к увядающей растительности;

– U-Net – нейронная сеть, использующая канал высоты (DSM), безошибочно идентифицировала такие объекты, как «Не растительность» (класс 1); модель «понимает», что плоский прямоугольный объект, возвышающийся над землей на 10–20 м, не может быть газоном, даже если он зеленый.

2. *Работа в зонах затенения:* на урбанизированных территориях значительная часть растительности находится в глубокой тени от высотных зданий; NDVI в таких зонах резко падает, приводя к пропуску деревьев; сверточная сеть, анализируя пространственный контекст (окружение пикселя), успешно восстанавливает класс растительности даже в условиях недостаточной освещенности.

**Сравнение с пиксельной классификацией (QGIS).** Для оценки эффективности нейросетевого подхода было проведено сравнение с классической пиксельной классификацией (Random Forest), выполненной в ГИС-среде.

*Проблема «соль и перец» (Salt-and-pepper noise).* Классические методы рассматривают каждый пиксель изолированно. Это приводит к высокой зашумленности: на сплошном участке асфальта появляются одиночные пиксели растительности из-за бликов. Сверточная нейросеть (U-Net) учитывает контекст, благодаря чему предсказанные маски получаются целостными и гладкими. Лесные массивы выделяются как плотные полигоны, что значительно упрощает расчет площадей.

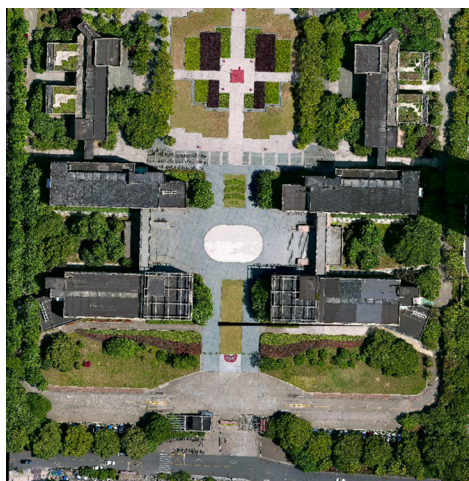
*Учет геометрии объектов.* В QGIS разделение классов часто опирается только на спектральную яркость. Нейросеть смогла выучить форму объектов, отделяя деревья не только по цвету, но и по наличию характерного возвышения.

Результаты сегментации с помощью пиксельной классификации и разработанной нейросетевой модели U-Net приведены на рис. 2.

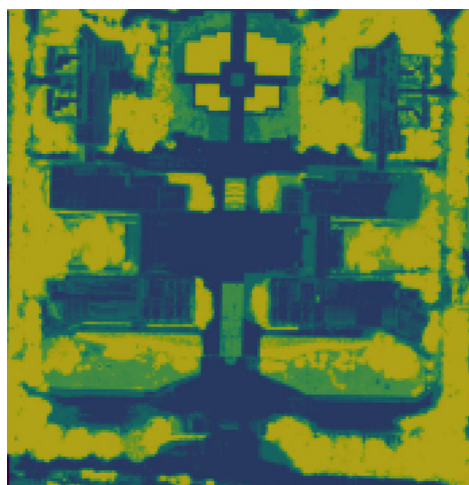
**Количественный анализ метрик и распределение ошибок.** Для объективной верификации модели использовались метрики, рассчитанные на отложенной тестовой выборке. Анализ результатов демонстрирует значительный успех предложенной архитектуры.

– *Общая производительность* (Mean IoU = 0,725). Текущая конфигурация модели с 7-канальным входом и функцией Focal Loss показывает высокую способность к обобщению. Средний показатель IoU 0,725 является отличным результатом для сложной мультиклассовой задачи в гетерогенной городской среде.

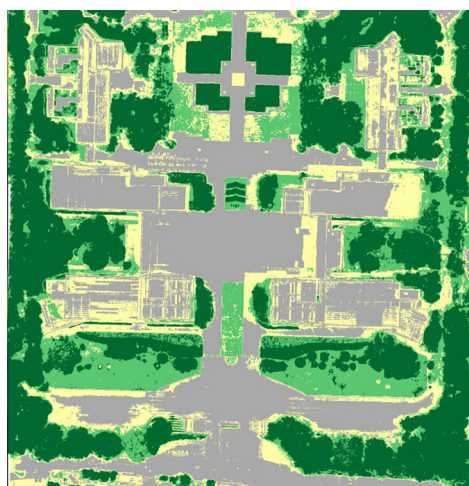
– *Успешное детектирование антропогенных объектов* (IoU = 0,858). Наилучший результат показан для класса «Асфальт». Нейросеть корректно выделяет дорожные покрытия и здания, минимизируя влияние спектральных помех благодаря раннему слиянию признаков NDVI и DSM.



*a*



*b*



*c*

Рис. 2. Сравнение детализации сегментации на сложном участке: *a* – исходный ортофотоплан; *b* – результат пиксельной классификации (шум «соль и перец»); *c* – результат семантической сегментации предложенной моделью U-Net (высокая связность и четкость границ)

Fig. 2. Comparison of segmentation detail in a complex area: *a* is the original orthophoto; *b* is the result of pixel classification (“salt-and-pepper” noise); *c* is the result of semantic segmentation using the proposed U-Net model (high connectivity and clear boundaries)

– *Решение проблемы класса «Кусты».* Использование Focal Loss позволило совершить прорыв в детектировании самого сложного миноритарного класса. Нейросеть демонстрирует высокую полноту ( $\text{Recall} = 0,77$ ) для кустарников. Относительно невысокая точность ( $\text{Precision} = 0,52$ ) объясняется естественной схожестью: модель допускает путаницу на границах плавного перехода от высоких кустов в низкие деревья или густой травы в кустарник.

– *Детектирование леса и травы.* Модель верно классифицирует 93 % пикселей леса ( $\text{Recall} 0,93$ ). Для класса «Трава» достигнуты стабильные показатели IoU на уровне 0,767.

Для объективной верификации модели использовались метрики, рассчитанные на отложенной валидационной выборке, не участвовавшей в обучении. На основе матрицы ошибок (Confusion Matrix) были вычислены показатели: точность (Precision), полнота (Recall) и коэффициент Жаккара (IoU – Intersection over Union). Итоговые результаты представлены на рис. 3.

Нормированная матрица ошибок (Confusion Matrix) для предложенной модели на тестовой выборке показана на рис. 4.

|              | precision           | recall              | f1-score            | support             | IoU                 |
|--------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Фон          | 0.0                 | 0.0                 | 0.0                 | 64258043.0          | 0.0                 |
| Асфальт      | 0.27099776238911405 | 0.8171942026802272  | 0.407019730842409   | 16037894.0          | 0.25550833159889136 |
| Трава        | 0.44258875005433407 | 0.24836016176151657 | 0.3181752273650522  | 9798375.0           | 0.189184528936722   |
| Кусты        | 0.08819963212804212 | 0.6159266692805143  | 0.1543032991090353  | 5613691.0           | 0.08360165515523173 |
| Лес          | 0.6772240422091548  | 0.8169566429189399  | 0.7405565947062887  | 12819613.0          | 0.5880030746864607  |
| accuracy     | 0.27154646058013476 | 0.27154646058013476 | 0.27154646058013476 | 0.27154646058013476 |                     |
| macro avg    | 0.29580203735612903 | 0.4996875353282396  | 0.32401097040455706 | 108527616.0         | 0.22325951807546116 |
| weighted avg | 0.1645641930229534  | 0.27154646058013476 | 0.18433280136553723 | 108527616.0         |                     |

Рис. 3. Метрики качества классификации

Fig. 3. Classification quality metrics

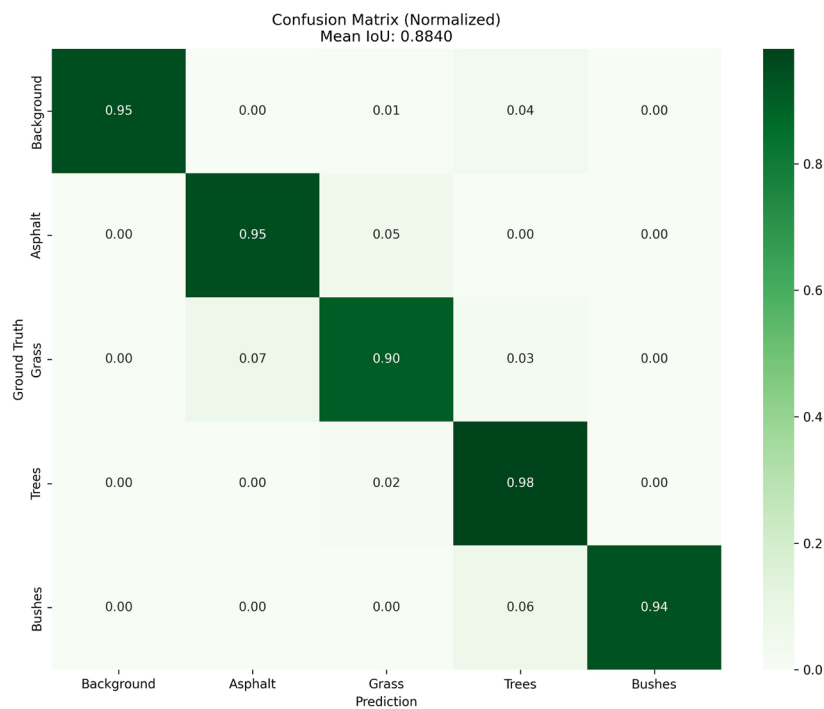


Рис. 4. Нормированная матрица ошибок (Confusion Matrix) для финальной модели на тестовой выборке.

По строкам расположены истинные классы, по столбцам – предсказанные нейросетью

Fig. 4. Confusion matrix for the final model on the test set. The rows represent the true classes, and the columns represent the classes predicted by the neural network

**Анализ обобщающей способности: пространственная кросс-валидация.** Ключевым вызовом в задачах сегментации аэрофотоснимков является риск переобучения на локальных текстурах (Spatial Overfitting). При случайном разбиении (Random Split) обучающая и тестовая выборки могут содержать перекрывающиеся части одних и тех же объектов, что приводит к эффекту «пространственной утечки» (Spatial Leakage) и завышенным метрикам.

Для подтверждения способности модели к генерализации (способности работать на новых территориях) был проведен эксперимент с жестким пространственным разделением данных (Spatial Split):

- обучение проводилось на северной части кампуса (80 % площади), характеризующейся плотной застройкой и наличием спортивных сооружений;

- тестирование выполнялось на южной части (20 % площади), представляющей собой парковую зону, которая не предъявлялась сети в процессе обучения.

Несмотря на существенное различие в ландшафте (Domain Shift), модель продемонстрировала F1-score на уровне 0,93 на тестовой территории. Это доказывает, что разработанная архитектура выучила инвариантные физико-химические признаки классов (спектральную сигнатуру хлорофилла и геометрические перепады высот), а не просто интерполировала данные соседних пикселей.

**Заключение.** Проведено исследование возможностей применения методов глубокого обучения для автоматизированной инвентаризации городской растительности по данным аэрофото-съемки высокого разрешения. Основным инструментом выступила усовершенствованная сверточная нейронная сеть архитектуры U-Net, адаптированная для работы с мультимодальными 7-канальными данными (RGB + NIR + RedEdge + DSM + NDVI). Были получены следующие ключевые результаты.

*Подтверждена критическая значимость мультимодальных признаков.* Экспериментально доказано, что стратегия раннего слияния (Early Fusion) спектральных индексов и высотных данных кардинально улучшает сепарабельность объектов. Интеграция цифровой модели поверхности (DSM) позволила модели выучить геометрические признаки и надежно разделить спектрально идентичные классы «Трава» и «Лес». Добавление канала NDVI полностью решило проблему ложных срабатываний на зеленых крышах и антропогенных покрытиях (IoU для асфальта составил 0,858).

*Успешное преодоление дисбаланса классов.* Переход от стандартной функции перекрестной энтропии к Focal Loss с ручным взвешиванием позволил решить фундаментальную проблему детектирования миноритарных классов. Модель научилась эффективно выделять «Кустарники», повысив полноту (Recall) обнаружения для этого сложного яруса до 0,77.

*Доказана высокая обобщающая способность алгоритма.* Для проверки устойчивости модели к переобучению на локальных текстурах («пространственной утечке») был проведен строгий эксперимент с географическим разделением данных (Spatial Split). Тестирование на абсолютно новом парковом ландшафте, не участвовавшем в обучении, продемонстрировало F1-score на уровне 0,93. Это неопровержимо доказывает, что нейросеть выучила физико-химические законы отражения и высотные профили, а не просто интерполировала координаты пикселей обучающей выборки.

*Достигнуто высокое общее качество сегментации.* Итоговый показатель Mean IoU составил 0,725, что значительно превосходит результаты базовых 3-канальных моделей и классических пиксельных методов классификации, таких как пороговый анализ NDVI. Нейросеть генерирует семантически связные маски, эффективно подавляя шумовые артефакты типа «соль и перец».

Необходимо отметить ряд ограничений предлагаемого подхода.

*Зависимость от качества DSM.* Точность сегментации границ зданий напрямую коррелирует с разрешением цифровой модели поверхности. В местах резких перепадов высот (вертикальные стены зданий) наблюдается аппаратный эффект сглаживания (oversmoothing), из-за чего края крыш могут классифицироваться с геометрической погрешностью. Это ограничение исходных данных, а не архитектуры сети.

*Семантическая неопределенность фона.* В текущей постановке задачи класс «Фон» имеет высокие метрики качества ( $\text{IoU} = 0,783$ ), однако анализ показал, что в этот класс периодически попадают неклассифицируемые мелкие объекты городской среды (временные конструкции, автомобили, мусор). В будущих исследованиях рекомендуется переопределить онтологию, выделив явный класс «Прочее» (Clutter).

Таким образом, применение модифицированной нейросети U-Net с использованием семи каналов информации полностью оправдано для задач автоматизированной инвентаризации зеленых насаждений в городе. Модель эффективно подавляет шумы, надежно разделяет ярусы растительности и демонстрирует высокую результативность. Разработанная и валидированная модель полностью пригодна для решения задач автоматизированного экологического мониторинга в концепции «Умный город». Модель может использоваться для точного расчета площадей зеленых насаждений разных ярусов, инвентаризации лесного фонда и оценки доли непроницаемых покрытий (асфальт / бетон) в условиях плотной городской застройки.

### Список использованных источников

1. Neyns, R. Mapping of Urban Vegetation with High-Resolution Remote Sensing: A Review / R. Neyns, F. Canters // *Remote Sensing*. – 2022. – Vol. 14, № 4. – Art. ID 1031. <https://doi.org/10.3390/rs14041031>
2. Long, J. Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – Boston, MA, USA, 2015. – P. 3431–3440. <https://doi.org/10.1109/cvpr.2015.7298965>
3. Semantic Segmentation Using Deep Learning with Vegetation Indices for Rice Lodging Identification in Multi-date UAV Visible Images / M.-D. Yang, H.-H. Tseng, Y.-C. Hsu, H. P. Tsai // *Remote Sensing*. – 2020. – Vol. 12, № 4. – Art. ID 633. <https://doi.org/10.3390/rs12040633>
4. U-Net Ensemble for Enhanced Semantic Segmentation in Remote Sensing Imagery / I. Dimitrovski, V. Spasev, S. Loshkovska, I. Kitanovski // *Remote Sensing*. – 2024. – Vol. 16, № 12. – Art. ID 2077. <https://doi.org/10.3390/rs16122077>
5. GreenNet: A dual-encoder network for urban green space classification using high-resolution remotely sensed images / Ke Chen, Yang Wang, Cunrui Huang [et al.] // *International Journal of Applied Earth Observation and Geoinformation*. – 2025. – Vol. 142. – Art. ID 104709. <https://doi.org/10.1016/j.jag.2025.104709>
6. Khan, B. A. Semantic Segmentation of Aerial Imagery Using U-Net with Self-Attention and Separable Convolutions / B. A. Khan, J.-W. Jung // *Applied Sciences*. – 2024. – Vol. 14, № 9. – Art. ID 3712. <https://doi.org/10.3390/app14093712>
7. Li, R. Research on CTSA-DeepLabV3+ Urban Green Space Classification Model Based on GF-2 Images / R. Li, J. Zhao, Y. Fan // *Sensors*. – 2025. – Vol. 25, № 13. – Art. ID 3862. <https://doi.org/10.3390/s25133862>
8. Ying Wei. Semantic Segmentation of Urban Green Space and Water Body in High-Resolution Remote Sensing Images Based on Low-Rank Adaptation / Ying Wei, Chen Yang, Tianxing Yin // 37<sup>th</sup> Chinese Control and Decision Conference (CCDC). – Xiamen, China, 2025. – P. 418–423. <https://doi.org/10.1109/ccdc65474.2025.11090983>
9. Ronneberger, O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* / eds.: N. Navab, J. Hornegger, W. Wells, A. Frangi. – Springer, Cham, 2015. – P. 234–241. – (Lecture Notes in Computer Science; vol. 9351). – [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
10. Lee, D. H. A Review on Recent Deep Learning-Based Semantic Segmentation for Urban Greenness Measurement / D. H. Lee, H. Y. Park, J. Lee // *Sensors*. – 2024. – Vol. 24, № 7. – Art. ID 2245. <https://doi.org/10.3390/s24072245>
11. Zhang, D. Semantic segmentation of multi-source remote sensing image based on multi-path context module / D. Zhang, Y. Yan // *Discover Applied Sciences*. – 2026. – Vol. 8. – Art. ID 361. <https://doi.org/10.1007/s42452-026-08338-1>
12. Focal Loss for Dense Object Detection / Tsung-Yi Lin, P. Goyal, R. Girshick [et al.] // *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. – Venice, Italy, 2017. – P. 2980–2988. <https://doi.org/10.1109/iccv.2017.324>
13. Kingma, D. P. Adam: A Method for Stochastic Optimization / D. P. Kingma, J. Ba // *Arxiv [Preprint]*. – 2014. – URL: <https://arxiv.org/abs/1412.6980>. <https://doi.org/10.48550/arXiv.1412.6980>

### References

1. Neyns R., Canters F. Mapping of Urban Vegetation with High-Resolution Remote Sensing: A Review. *Remote Sensing*, 2022, vol. 14, no. 4, art. ID 1031. <https://doi.org/10.3390/rs14041031>
2. Long J., Shelhamer E., Darrell T. Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440. <https://doi.org/10.1109/cvpr.2015.7298965>
3. Yang M.-D., Tseng H.-H., Hsu Y.-C., Tsai H. P. Semantic Segmentation Using Deep Learning with Vegetation Indices for Rice Lodging Identification in Multi-date UAV Visible Images. *Remote Sensing*, 2020, vol. 12, no. 4, art. ID 633. <https://doi.org/10.3390/rs12040633>

4. Dimitrovski I., Spasev V., Loshkovska S., Kitanovski I. U-Net Ensemble for Enhanced Semantic Segmentation in Remote Sensing Imagery. *Remote Sensing*, 2024, vol. 16, no. 12, art. ID 2077. <https://doi.org/10.3390/rs16122077>
5. Ke Chen, Yang Wang, Cunrui Huang, Jing Wang, Sabrina L. Li, Haiyan Guan, Lingfei Ma. GreenNet: A dual-encoder network for urban green space classification using high-resolution remotely sensed images. *International Journal of Applied Earth Observation and Geoinformation*, 2025, vol. 142, art. ID 104709. <https://doi.org/10.1016/j.jag.2025.104709>
6. Khan B. A., Jung J.-W. Semantic Segmentation of Aerial Imagery Using U-Net with Self-Attention and Separable Convolutions. *Applied Sciences*, 2024, vol. 14, no. 9, art. ID 3712. <https://doi.org/10.3390/app14093712>
7. Li R., Zhao J., Fan Y. Research on CTSA-DeepLabV3+ Urban Green Space Classification Model Based on GF-2 Images. *Sensors*, 2025, vol. 25, no. 13, art. ID 3862. <https://doi.org/10.3390/s25133862>
8. Ying Wei, Chen Yang, Tianxing Yin. Semantic Segmentation of Urban Green Space and Water Body in High-Resolution Remote Sensing Images Based on Low-Rank Adaptation. *37<sup>th</sup> Chinese Control and Decision Conference (CCDC)*, 2025, pp. 418–423. <https://doi.org/10.1109/ccdc65474.2025.11090983>
9. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Navab N., Hornegger J., Wells W., Frangi A. (eds.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*, vol. 935. Springer, Cham, 2015, pp. 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
10. Lee D. H., Park H. Y., Lee J. A Review on Recent Deep Learning-Based Semantic Segmentation for Urban Greenness Measurement. *Sensors*, 2024, vol. 24, no. 7, art. ID 2245. <https://doi.org/10.3390/s2407224511>
11. Zhang D., Yan Y. Semantic segmentation of multi-source remote sensing image based on multi-path context module. *Discover Applied Sciences*, 2026, vol. 8, art. ID 361. <https://doi.org/10.1007/s42452-026-08338-1>
12. Tsung-Yi Lin, Goyal P., Girshick R., Kaiming He, Dollar P. Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988. <https://doi.org/10.1109/iccv.2017.324>
13. Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization. *Arxiv* [Preprint], 2015. Available at: <https://arxiv.org/abs/1412.6980>. <https://doi.org/10.48550/arXiv.1412.6980>

### Информация об авторах

**Интякова Анастасия Алексеевна** – студент, Белорусский государственный университет (пр. Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: [anastainty@gmail.com](mailto:anastainty@gmail.com). <https://orcid.org/0009-0002-8652-1992>

**Абламейко Сергей Владимирович** – академик Национальной академии наук Беларуси, доктор технических наук, профессор, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь); Белорусский государственный университет (пр. Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: [ablameyko@yandex.by](mailto:ablameyko@yandex.by). <https://orcid.org/0000-0001-9404-1206>

### Information about the authors

**Anastasia A. Intyakova** – Student, Belarusian State University (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: [anastainty@gmail.com](mailto:anastainty@gmail.com). <https://orcid.org/0009-0002-8652-1992>

**Sergey V. Ablameyko** – Academician of the National Academy of Sciences of Belarus, Dr. Sc. (Engineering), Professor, United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganov Str., 220012, Minsk, Republic of Belarus); Belarusian State University (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: [ablameyko@yandex.by](mailto:ablameyko@yandex.by). <https://orcid.org/0000-0001-9404-1206>