ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION

УДК 004.93
https://doi.org/10.37661/1816-0301-2026-23-3-7-23

Поступила в редакцию | Received 17.04.2026
Подписана в печать | Accepted 11.05.2026
Опубликована | Published 30.09.2026

Модифицированная архитектура DeepLabV3+ с интеграцией глобального контекста и контрастивного обучения с учителем для семантической сегментации изображений сверхвысокого разрешения

А. А. Козлов✉, С. В. Абламейко
✉E-mail: anatoly.kozlov.ak@yandex.ru

Белорусский государственный университет,
пр. Независимости, 4, Минск, 220030, Беларусь

Аннотация

Цели. Целью исследования является разработка современного метода семантической сегментации видеокадров сверхвысокого разрешения (4K) в области задач дистанционного зондирования Земли (ДЗЗ) с помощью модификации сверточной нейронной сети DeepLabV3+.

Методы. Метод направлен на решение двух критических проблем: ограниченности рецептивного поля модели при работе с локальными фрагментами изображения в силу больших затрат видеопамяти графических процессоров и некачественного разделения классов со схожими цветовыми характеристиками и текстурой. Базовая архитектура была дополнена двумя модулями механизма внимания для улучшения локализации объектов, а стандартная функция потерь cross entropy – алгоритмом контрастивного обучения с учителем (Supervised Contrastive Learning, SupCon) для лучшего разделения схожих классов. В Atrous Spatial Pyramid Pooling (ASPP)-модуль была встроена линейная модуляция (Feature-wise Linear Modulation, FiLM) для внедрения глобального контекста в виде признаков, полученных из исходного изображения.

Результаты. Результаты проведенных экспериментов показали, что предложенный подход успешно справляется с минимизацией артефактов на границах классов, обладает высокой способностью различать визуально схожие классы, а также повышает точность сегментации в условиях специфического набора данных. Предложенная архитектура превосходит базовую по метрике mean Intersection over Union на 0,45 %.

Заключение. Разработанная модификация DeepLabV3+ позволяет эффективно решать задачи семантической сегментации изображений сверхвысокого разрешения в области ДЗЗ, обеспечивая высокую точность при незначительном увеличении вычислительной сложности.

Ключевые слова: дистанционное зондирование Земли, сверточные нейронные сети, семантическая сегментация, контрастивное обучение, мониторинг терриконов, механизм внимания

© Козлов А. А., Абламейко С. В., 2026

Check for updates

CC BY

Благодарности. Авторы выражают благодарность сотрудникам научно-исследовательской лаборатории прикладной механики механико-математического факультета Белорусского государственного университета за предоставленное оборудование и данные для проведения эксперимента.

Для цитирования. Козлов, А. А. Модифицированная архитектура DeepLabV3+ с интеграцией глобального контекста и контрастивного обучения с учителем для семантической сегментации изображений сверхвысокого разрешения / А. А. Козлов, С. В. Абламейко // Информатика. – 2026. − Т. 23, № 3. – С. 7–23. – https://doi.org/10.37661/1816-0301-2026-23-3-7-23.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Modified DeepLabV3+ architecture with global context integration and supervised contrastive learning for semantic segmentation of ultra-high-resolution image

Anatoly A. Kozlov✉, Sergey V. Ablameyko
✉E-mail: anatoly.kozlov.ak@yandex.ru

Belarusian State University,
av. Nezavisimosti, 4, Minsk, 220030, Belarus

Abstract

Objectives. We propose a modern method for semantic segmentation of ultra-high-resolution (4K) video frames in the field of Earth remote sensing using a modification of the DeepLabV3+ convolutional neural network.

Methods. The method is aimed at solving two critical problems: the limited receptive field of the model when working with local image fragments due to GPU memory constraints, and the poor separation of classes with similar color characteristics and texture. The baseline architecture was supplemented with two attention mechanism modules to improve object localization, and the standard cross-entropy loss function was supplemented with a Supervised Contrastive Learning (SupCon) algorithm for better separation of similar classes. Feature-wise Linear Modulation (FiLM) was embedded into the Atrous Spatial Pyramid Pooling (ASPP) module to introduce global context in the form of features extracted from the original image.

Results. The results of the experiments showed that the proposed approach successfully minimizes artifacts at class boundaries, has a high ability to distinguish visually similar classes, and improves segmentation accuracy under the conditions of a specific dataset. The proposed architecture outperforms the baseline by 0,45 % in terms of mean Intersection over Union.

Conclusion. The developed modification of DeepLabV3+ effectively addresses the challenges of semantic segmentation of ultra-high-resolution images in the field of Earth remote sensing, achieving high accuracy with a negligible increase in computational complexity.

Keywords: Earth remote sensing, convolutional neural networks, semantic segmentation, contrastive learning, tailings pond monitoring, attention mechanism

Acknowledgments. The authors express their gratitude to the staff of the Research Laboratory of Applied Mechanics of the Faculty of Mechanics and Mathematics of Belarusian State University for providing the equipment and data used in the experiment.

For citation. Kozlov A. A., Ablameyko S. V. *Modified DeepLabV3+ architecture with global context integration and supervised contrastive learning for semantic segmentation of ultra-high-resolution image*. Informatika [*Informatics*], 2026, vol. 23, no. 3, pp. 7–23 (In Russ.). https://doi.org/10.37661/1816-0301-2026-23-3-7-23.

Conflict of interests. The authors declare of no conflict of interest.

Введение

Внедрение новых технологий в область мониторинга технических объектов, в частности шламохранилищ, является приоритетной задачей в деле предотвращения экологических катастроф и чрезвычайных ситуаций [1, 2]. Катастрофические последствия обвалов объектов такого рода в Бразилии и Канаде показывают, что и сейчас существуют высокие риски аварий на шламохранилищах [3, 4]. Традиционные методы контроля могут оказаться недостаточными для предотвращения подобных катастроф, поэтому в настоящее время активно ведутся исследования по внедрению глубокого обучения в мониторинг и задачи ДЗЗ [5].

Республика Беларусь является одним из крупнейших производителей калийных удобрений, и проблема мониторинга прорыва терриконов и подтопления окружающих сельскохозяйственных угодий стоит очень остро [6]. Именно поэтому развитие методов глубокого обучения для сегментации кадров сверхвысокого разрешения является важным направлением в области искусственного интеллекта и ДЗЗ.

Исследования в области анализа данных ДЗЗ в последние годы почти полностью проводятся с помощью методов глубокого обучения. Первые успехи в области семантической сегментации были связаны с появлением полносверточных нейронных сетей [7] и архитектуры U-Net [8]. С тех пор исследователи в области компьютерного зрения добились больших успехов в разработке архитектур для решения этой задачи. Среди них выделяются модели, которые решают проблему недостатка контекста и ограниченности рецептивного поля. К таким относятся PSPNet [9], которая использует Pyramid Pooling Module, и семейство DeepLab (V1, V2, V3, V3+) [10–13], где внимание уделено расширению рецептивного поля за счет расширенных сверток и ASPP-модуля. Визуальные трансформеры также показали свою эффективность и получают широкое распространение в современных исследованиях [14, 15], но их применение требует огромного объема видеопамяти, что является серьезным аргументом против их использования для сегментации изображений сверхвысокого разрешения.

Отсутствие механизмов внимания у классических архитектур сверточных нейронных сетей приводит к тому, что модели теряют информацию на том этапе, когда карты признаков максимально сжимаются в один вектор. Это явилось серьезной мотивацией для разработки механизмов внимания. Так же, как зрение человека фокусируется на главном объекте, интересующем его на данный момент, механизмы внимания позволили нейронным сетям концентрироваться на информативных признаках, все больше игнорируя фон и шум. В задачах ДЗЗ это критически необходимо для успешной сегментации [16].

Самым простым существующим на данный момент подходом является блок Squeeze-and-Excitation (SE) [17]. Недостаток этого блока заключается в том, что он фокусируется исключительно на межканальных зависимостях, игнорируя пространственные. Более продвинутые модули (Convolutional Block Attention Module, CBAM) [18] объединили в себе канальное и пространственное внимание, но у них были большие траты вычислительной мощности.

Одним из современных механизмов внимания является модуль механизма внимания (Coordinate Attention, CA) [19]. Этот легковесный модуль кодирует и встраивает в карту признаков координаты горизонтальной и вертикальной осей. Он позволяет эффективно уточнять сегментацию на границах классов со сложной геометрией, не сильно добавляя количество параметров модели. Данный компромиссный вариант является наиболее эффективным.

С точки зрения многомерного анализа можно рассматривать выходные признаки сверточной нейронной сети как элементы векторного пространства. Вполне очевидным фактом является то, что визуально схожие классы будут находиться в этом пространстве очень близко друг к другу. Таким образом, обучение модели с применением традиционной функции потерь (cross-entropy) может приводить к ошибочной классификации пикселей, обладающих схожими цветовыми характеристиками, но принадлежащих к разным семантическим классам. Например, в рассматриваемом случае это вода и небо, а также шламохранилище и регион слежения, который является частью шламохранилища. Решением данной проблемы стало внедрение методов контрастивного обучения [20, 21]. Алгоритм SupCon позволяет обучать модель таким образом, чтобы векторы признаков пикселей одного класса стягивались в пространстве признаков, а расстояние между векторами разных классов увеличивалось [22]. В случае специфического набора данных использование алгоритма SupCon повысит устойчивость модели к шуму и улучшит точность сегментации.

После тщательного анализа существующих подходов для решения описанных выше проблем авторы предлагают дополнить современную архитектуру СНС DeepLabV3+ следующими методами:

– интеграцией глобального контекста с помощью модуля FiLM в виде вектора признаков, полученного из исходного изображения, для лучшего понимания моделью расположения локального фрагмента относительно всей сцены;

– добавлением блоков CA для уточнения сегментации границ классов;

– использованием обучения с помощью SupCon для эффективного разделения визуально схожих классов.

1. Методология и архитектура

В качестве модели для извлечения признаков была выбрана архитектура ResNet-50 [23]. Выбор ResNet-50 обоснован тем, что данная архитектура – это оптимальное решение с точки зрения выбора между вычислительными затратами и точностью. Данный факт подтверждается большим количеством работ, содержащих в себе сравнительный анализ качества работы моделей на одинаковых данных. Согласно результатам сравнительного анализа в области классификации медицинских изображений [24] ResNet-101 дает всего 0,23 % прироста к точности при том, что имеет почти в два раза больше параметров. Более того, некоторые работы демонстрируют сопоставимую точность моделей ResNet-50 и ResNet-101 [25, 26]. Все это подтверждает целесообразность выбора менее ресурсоемкой архитектуры.

Избранным механизмом внимания в данной работе является модуль CA. Его превосходство над другими существующими механизмами внимания подтверждено экспериментально [27]. В отличие от существующих распространенных методов SE [28] и CBAM, CA сочетает в себе учет пространственной информации и захват длинных за-

висимостей в картах признаков. Также его значимым преимуществом является легковесность, а эффективность в задачах ДЗЗ подтверждена надежными источниками [29]. Эти факты позволяют сделать вывод о том, что выбор CA в качестве механизма внимания является наиболее эффективным решением для семантической сегментации и ДЗЗ.

В настоящей работе стандартная функция потерь cross entropy была дополнена контрастивным обучением с учителем SupCon. Другие методы контрастивного обучения самоконтролируемые, т. е. не используют разметку, что не подходит для рассматриваемой ситуации, когда в изображении встречаются визуально схожие классы. Эффективность метода контрастивного обучения с учителем подтверждена экспериментально [30], а также успешно применима в ДЗЗ.

1.1. Стратегия фрагментации и многомасштабной обработки. В настоящее время наиболее распространенные аппаратные ускорители (графические процессоры) имеют объем видеопамяти в 24 Гб. Библиотека torchinfo позволила оценить количество видеопамяти, которое необходимо для запуска процесса обучения или инференса на одном элементе набора данных. Для обработки моделью, являющейся предложенной модификацией DeepLabV3+, видеокадров разрешения 4K необходимо около 40 Гб видеопамяти. Существуют серверные решения, которые позволяют производить такие вычисления, но авторы ограничены возможностями графического процессора NVDIA GeForce RTX 3080 Ti, который обладает 12 Гб видеопамяти. В связи с этим к видеокадрам была применена стратегия декомпозиции, или фрагментации (деления на фрагменты).

Для обработки данных было принято решение применить стратегию разбиения изображения на локальные фрагменты размерностью 512×512. Выбор данной размерности обусловлен необходимостью достижения баланса между сохранением локальной контекстной информации фрагмента и вычислительными затратами его обработки. Непосредственно разбиение осуществляется по регулярной сетке с шагом в 400 пикселей. Это обеспечивает перекрытие в 112 пикселей – 21,8 % от размера фрагмента. Перекрытие позволяет минимизировать наличие разрывов областей классов на границах фрагментов в результате сегментации, так как при формировании прогноза целых изображений в пересечениях фрагментов выходные значения модели суммируются и усредняются. Координаты левого верхнего угла фрагментов по ширине задаются последовательностью $x \in \{0, 400, 800, \ldots, 2800, 3200, 3328\}$. Последний фрагмент смещается с целью выравнивания по правой границе кадра. Аналогичным образом разбиение реализовано и по высоте кадра и задается последовательностью $y \in \{0, 400, 800, \ldots, 1200, 1600, 1650\}$. Следовательно, фрагменты целиком покрывают все изображение.

В связи с ограничениями вычислительных мощностей вынужденное применение стратегии фрагментации создает проблему отсутствия глобального контекста. Поле зрения модели ограничено локальным фрагментом, что может привести к неточностям сегментации. Способность модели к дифференциации классов, имеющих схожие цветовые характеристики и текстуру, ухудшается. Особенно это вероятно для ситуации, когда фрагмент оказался внутри области класса.

1.2. Интеграция глобального контекста. Схема интеграции глобального контекста в качестве модификации архитектуры модели показана на рис. 1. Исходный кадр подвергается обработке с помощью легковесной модели MobileNetV3 [31], благодаря чему получается

вектор признаков размерностью 960, в котором содержится глобальная информация. Затем этот вектор проходит через двуслойную нейронную сеть, которая уже является частью основной модели предложенной модификации. В итоге получаются два вектора размерностями 256 каждый.

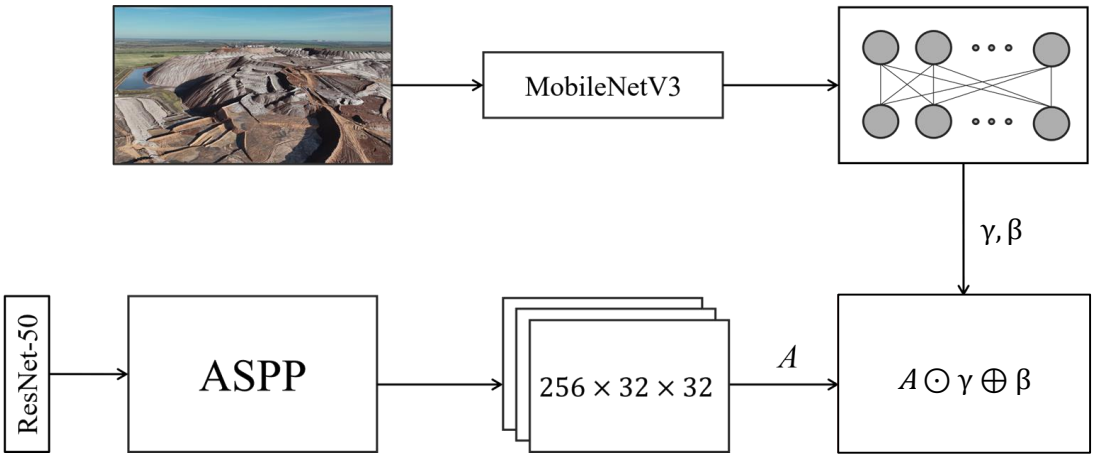

Рис. 1. Схема интеграции глобального контекста

Fig. 1. Global context integration diagram

Полученные векторы комбинируются с выходной картой признаков модуля ASPP с помощью блока FiLM [32], что можно записать в виде следующего выражения:

$$\hat{F} = \gamma \odot F_{ASPP} \oplus \beta, \quad (1)$$

где $\odot$ – поэлементное умножение; $\oplus$ – поэлементное сложение; γ и β – векторы с обучаемыми параметрами, которые получаются из глобального контекста, а F_{ASPP} – выходная карта признаков ASPP-модуля. Данная модификация позволяет изменять активации нейронной сети на основе глобальной специфики конкретного видеокадра.

1.3. Координатно-ориентированные механизмы внимания. С целью повышения точности локализации геометрически сложных объектов в архитектуру внедрены блоки механизма внимания. На рис. 2 представлена схема работы этого модуля.

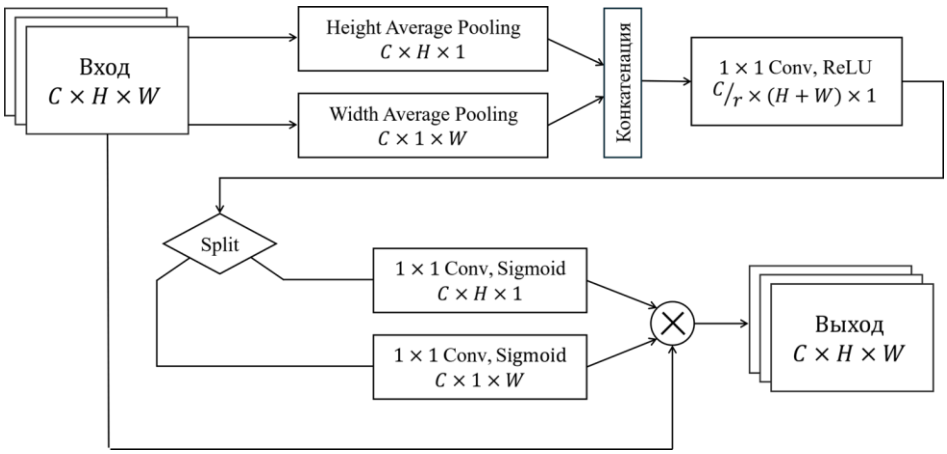

Рис. 2. Схема механизма внимания

Fig. 2. Coordinate Attention mechanism diagram

Механизм внимания реализует эффективную пространственную селекцию признаков путем декомпозиции глобального пулинга на две одномерные операции вдоль осей. Полученные дескрипторы координат проходят через общую нелинейную трансформацию, что позволяет модели вычислять весовые коэффициенты с учетом кросс-координатных зависимостей.

Применение двух последовательных масок внимания обеспечивает точную локализацию объектов и подавление фонового шума, сохраняя при этом высокую вычислительную эффективность.

1.4. Внедрение контрастивного обучения. Суть контрастивного обучения состоит в том, что выходные признаки модели проецируются в конкретное линейное пространство признаков. Вычисляя схожесть векторов в рамках этого пространства, модель обучается так, чтобы векторы одинаковых классов сближались, а разных – отдалялись. Для реализации модели на ее выход встраивается проекционный модуль, который преобразует выходную карту признаков модели в векторы-признаки. Иначе говоря, именно он проецирует признаки в конкретное векторное пространство. В итоге для каждого пикселя исходного фрагмента размерностью 512×512 получаются векторы размерностью 128 (рис. 3).

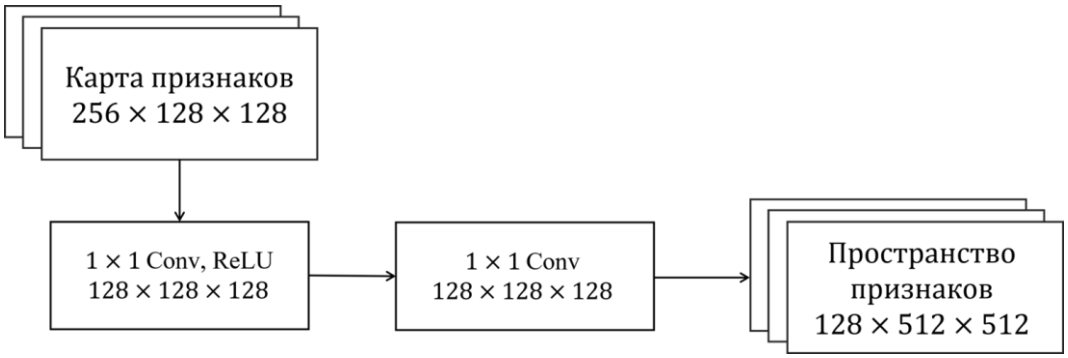

Рис. 3. Схема проекционного модуля
Fig. 3. Projection module diagram

Сам алгоритм заключается в том, что все выбранные векторы проецируются на многомерную единичную сферу и с помощью скалярного произведения вычисляется косинусовое сходство между всеми векторами. Таким образом, чем меньше расстояние между векторами, которые принадлежат одному классу, тем меньше функция потерь SupCon. Все вышесказанное можно выразить с помощью формулы

$$L_i = \frac{-1}{|P(i)|}\sum_{p\in P(i)} \log \frac{\exp\left(z_i \cdot {}^{z_p}/_{\tau}\right)}{\sum_{a\in A(i)} \exp\left(z_i \cdot {}^{z_a}/_{\tau}\right)}, \quad (2)$$

где z_i – нормализованные векторы-признаки; $P(i)$ – подмножество выбранных векторов, принадлежащих тому же классу; $A(i)$ – множество выбранных векторов, исключая вектор, для которого в данный момент проводятся вычисления; τ – параметр, который регулирует интенсивность влияния расстояния на обучение.

Естественно, применять алгоритм ко всем векторам будет невозможно с точки зрения нагрузки на графический процессор, ведь тогда нужно будет считать функцию потерь попарно для всех 262 144 векторов. Поэтому для оптимизации вычислительных затрат каждый раз будут выбираться 128 случайных векторов и SupCon будет применяться только к ним.

В итоге получим комбинацию двух функций потерь, сконструированную следующим образом:

$$Loss = Loss_{cross\ entropy} + \omega \cdot Loss_{SupCon}. \quad (3)$$

1.5. Предлагаемая архитектура модификации. На рис. 4 изображена схема предложенной модификации. Проекционный модуль в конце декодера включается только во время обучения модели. Два модуля механизма внимания имеют разное количество каналов входных карт признаков. Первый модуль, расположенный перед конкатенацией низкоуровневых и высокоуровневых признаков, предполагает наличие 48 каналов у входной карты признаков, а второй модуль – 256 каналов.

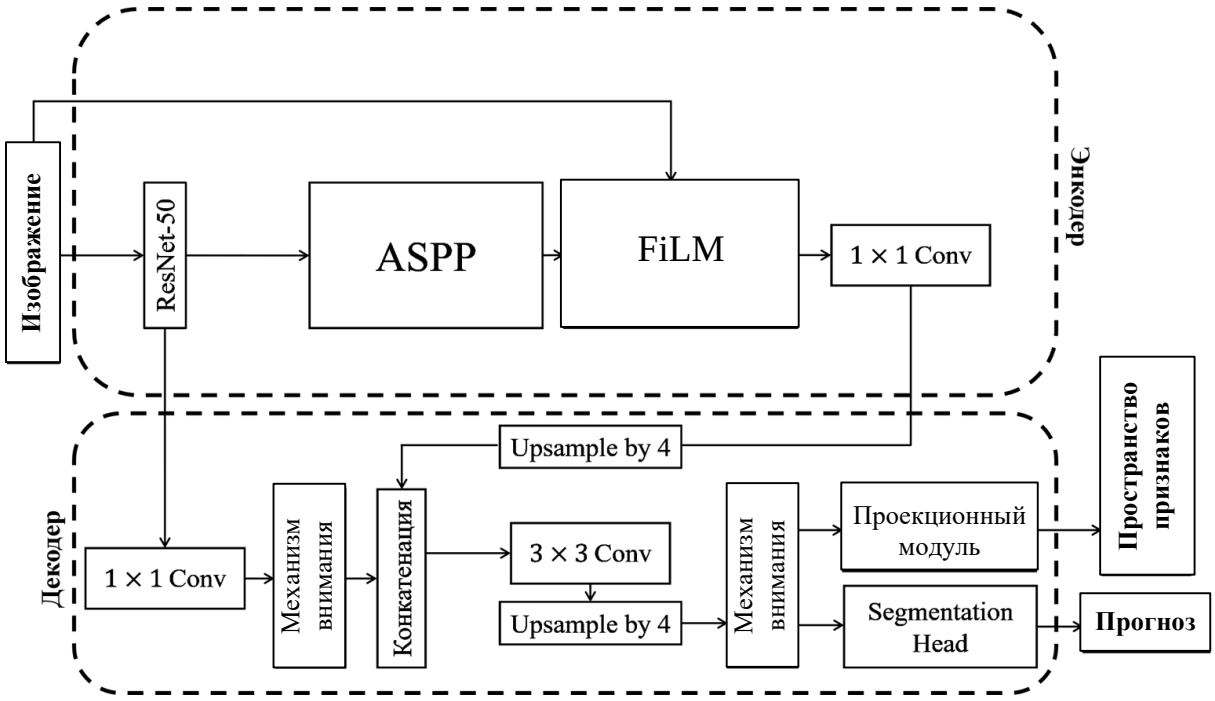

Рис. 4. Предложенная модификация архитектуры DeepLabV3+ с интеграцией механизмов внимания, адаптивной калибровки признаков и проекционного модуля контрастивного обучения

Fig. 4. The proposed modification of the DeepLabV3+ architecture, with integration of attention mechanisms, adaptive feature calibration, and a projection module for contrastive learning

В начале модели расположен энкодер ResNet-50. У данной модели две выходные карты признаков: низкого уровня (размерностью 256×128×128) и высокого (размерностью 2048×32×32). Карта признаков высокого уровня поступает на вход модуля ASPP, который в свою очередь имеет выходную карту признаков размерностью 256×32×32.

В модуле FiLM объединяются выходная карта признаков модуля ASPP и сжатый при помощи MobileNetV3 и двуслойной нейронной сети полный 4К-кадр. Интеграция глобального контекста подробно описана в подразд. 1.2. Выходная карта признаков модуля FiLM имеет ту же размерность, что и выходная карта признаков модуля ASPP.

Карта уровней низкого уровня проходит через свертку 1×1, уменьшая число каналов, и попадает на вход в первый блок CA размерностью 48×128×128. Модуль CA не меняет размерность карты признаков, а лишь применяет к карте признаков веса внимания. Далее конкатенируются высоко- и низкоуровневые признаки. В результате конкатенации получается карта признаков размерностью 304×128×128.

Входная и выходная карты признаков второго модуля CA имеют размерность 256×128×128. После модуля CA признаки поступают на вход в голову сегментации при инференсе, а при обучении – еще и в проекционный модуль.

2. Эксперимент и результаты

2.1. Описание набора данных. Набор данных для проведения эксперимента был сформирован из 110 размеченных видеокадров сверхвысокого разрешения. Точное разрешение – 3840×2160 пикселей. Кадры получены профессиональной съемкой с помощью беспилотного летательного аппарата. Каждый кадр разбивались на 60 фрагментов размером 512×512 пикселей описанным выше способом. Таким образом, набор данных состоит из 6600 элементов.

2.2. Конфигурация обучения и метрики оценки качества. Обучение было реализовано с помощью программного интерфейса PyTorch и библиотеки CUDA для запуска вычислений на аппаратных ускорителях. В качестве такого ускорителя использовался графический процессор NVDIA GeForce RTX 3080 Ti с 12 Гб видеопамяти.

В табл. 1 представлены гиперпараметры, использованные при обучении исследуемых моделей. Для обеспечения сопоставимости результатов скорость обучения (Learning rate) была зафиксирована на одном уровне для всех архитектур. Размер батча выбран в соответствии с максимальным количеством доступной видеопамяти. Обучение проводилось в течение 100 эпох. Стоит отметить, что для предложенной модификации введены дополнительные параметры: температурный коэффициент τ и весовой коэффициент ω, которые используются в функции контрастивных потерь.

Таблица 1

Список гиперпараметров для обучения сравниваемых моделей

Table 1

Hyperparameters for training the compared models

	Learning rate	Batch size	Epochs	τ	ω
Модификация	0,0004	12	100	0,7	0,1
DeepLabV3+	0,0004	14	100	–	–
U-Net++	0,0004	4	100	–	–

Для количественной оценки качества сегментации выбраны следующие метрики: средний коэффициент пересечения над объединением (mIoU), коэффициент сходства Дайса (Dice Coefficient) и общая точность.

Коэффициент mIoU отражает среднее отношение площади пересечения предсказанной маски с эталонной к площади их объединения по всем классам. Коэффициент сходства Дайса оценивает меру близости двух выборок и более чувствителен к корректному выделению мелких структур в условиях дисбаланса классов.

Общая точность (Overall Accuracy) рассчитывается как доля пикселей изображения, классифицированных верно, что позволяет оценить общую эффективность модели на всем наборе данных.

Результаты сравнительной оценки качества сегментации и вычислительной сложности моделей приведены в табл. 2. Предложенная модификация демонстрирует наилучшие показатели по всем метрикам, достигая значения mIoU 98,85 %, что на 0,45 % выше базовой архитектуры DeepLabV3+ и на 3,08 % выше модели U-Net++. Столь высокие показатели свидетельствуют об эффективности внедрения механизмов внимания и интеграции глобальных признаков.

По количеству параметров представленная модель лишь на 3 % превосходит стандартную. При этом она практически в два раза легче архитектуры U-Net++, что делает ее более пригодной для развертывания в системах реального времени при значительно более высокой точности.

Таблица 2

Результаты оценки качества сравниваемых моделей

Table 2

Performance evaluation results of the compared models

	mIoU, %	Dice, %	Overall Accuracy, %	Total parameters
Модификация	**98,85**	**99,42**	**99,44**	**27 490 000**
DeepLabV3+	98,40	99,19	99,23	26 680 000
U-Net++	95,77	97,77	98,78	48 990 000

2.3. Визуальный анализ результатов. Качественный анализ результатов сегментации видеокадра (рис. 5) показывает, что предложенные модификации значительно улучшают топологическую связность сегментированных объектов. В то время как базовый DeepLabV3+ демонстрирует разрывы в узких структурах (например, линия класса фона на увеличенных фрагментах), представленная модель сохраняет непрерывную и четкую границу. Это улучшение в первую очередь связано с механизмом CA, который обеспечивает точную пространственную локализацию, и модулем FiLM, интегрирующим глобальный контекст для предотвращения локальных ошибок классификации. Кроме того,

использование SupCon обеспечивает лучшую различимость классов, что приводит к более плавным переходам и снижению шума на стыках нескольких классов. Такие визуальные улучшения согласуются с наблюдаемым увеличением метрики mIoU на 0,45 %, подтверждая, что модель лучше улавливает геометрию сцены.

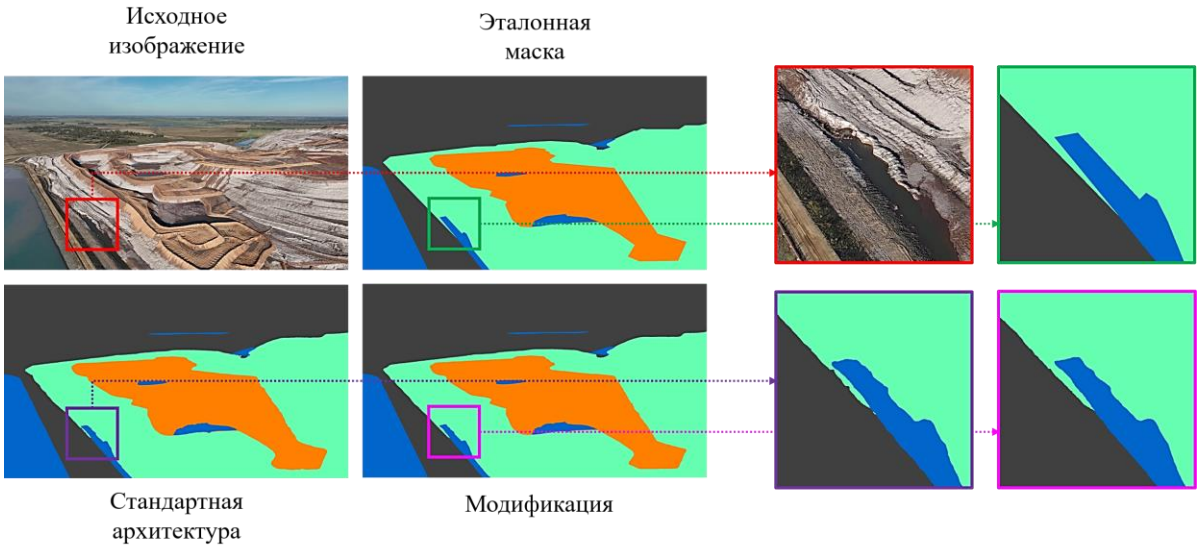

Рис. 5. Результаты сегментации видеокадра

Fig. 5. Video frame segmentation results

Заключение

В работе представлен модифицированный метод семантической сегментации видеокадров сверхвысокого разрешения (4K) для задач мониторинга шламохранилищ. Предложенная архитектура на базе DeepLabV3+ эффективно решает проблему ограниченного рецептивного поля при обработке локальных фрагментов изображения за счет интеграции глобального контекста.

Разработан комбинированный подход к обучению, объединяющий стандартную кросс-энтропию и функцию потерь SupCon. Это дало возможность существенно повысить сепарабельность визуально схожих классов, минимизировав ошибки классификации.

Адаптация модели под конкретные условия съемки позволила достичь прироста метрики mIoU на 0,45 % по сравнению с базовой архитектурой DeepLabV3+.

Практическая значимость работы заключается в возможности автоматизации экологического контроля промышленных объектов, что снижает риск возникновения техногенных аварий.

Вклад авторов. *А. А. Козлов* разработал программное обеспечение для обучения моделей, предложил архитектурные модификации, выполнил разметку набора данных, провел вычислительные эксперименты, осуществил анализ и интерпретацию результатов. *С. В. Абламейко* сформулировал задачу, установил научные контакты для получения исходных данных и утвердил окончательную версию статьи для публикации.

Список использованных источников

1. Цзян, Ч. Выявление структуры землепользования в зоне влияния Солигорского калийного комбината по данным дистанционного зондирования / Ч. Цзян, А. Н. Червань // Природопользование. – 2025. – № 1. – С. 51–63.

2. Roche, C. Mine tailings storage: Safety is no accident / C. Roche, K. Thygesen, E. Baker // Rapid Response Assessment / United Nations Environment Programme, GRID-Arendal. – Arendal, 2017. – P. 6–10.

3. The 2019 Brumadinho tailings dam collapse: Possible cause and impacts of the worst human and environmental disaster in Brazil / L. H. S. Rotta, E. Alcântara, E. Park [et al.] // International Journal of Applied Earth Observation and Geoinformation. – 2020. – Vol. 90. – Art. 102119. – URL: https://www.sciencedirect.com/science/article/pii/S0303243420300192?via%3Dihub (date of access: 16.03.2026). – https://doi.org/10.1016/j.jag.2020.102119.

4. Santamarina, J. C. Why coal ash and tailings dam disasters occur / J. C. Santamarina, L. A. Torres-Cruz, R. C. Bachus // Science. – 2019. – Vol. 364, iss. 6440. – P. 526–528. – https://doi.org/10.1126/science.aax1927.

5. Mining and tailings dam detection in satellite imagery using deep learning / R. Balaniuk, O. Isupova, S. Reece // Sensors. – 2020. – Vol. 20, no. 23. – Art. 6936. – URL: https://www.mdpi.com/1424-8220/20/23/6936 (date of access: 16.03.2026). – https://doi.org/10.3390/s20236936.

6. Стельмах, В. И. Экологические проблемы добычи калийных солей и некоторые пути их преодоления / В. И. Стельмах // Обеспечение экономической безопасности Республики Беларусь: теория и практика : материалы Респ. науч.-практ. конф., Минск, 19 дек. 2019 г. / Акад. М-ва внутр. дел Респ. Беларусь. – Мн., 2019. – С. 103–105.

7. Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, Massachusetts, USA, 7–12 June 2015. – Boston, 2015. – P. 3431–3440. – https://doi.org/10.1109/CVPR.2015.7298965.

8. 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 : 18th Intern. Conf., Munich, 5–9 Oct. 2015 / ed.: N. Navab [et al.]. – Cham, 2015. – Pt. III. – P. 234–241. – (Lecture Notes in Computer Science ; vol. 9351). – https://doi.org/10.1007/978-3-319-24574-4_28.

9. Pyramid scene parsing network / H. Zhao, J. Shi, X. Qi [et al.] // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 21–26 July 2017. – Honolulu, 2017. – P. 2881–2890. – https://doi.org/10.1109/CVPR.2017.660.

10. Encoder-decoder with atrous separable convolution for semantic image segmentation / L. C. Chen, Y. Zhu, G. Papandreou [et al.] // Proc. of the European Conf. on Computer Vision (ECCV), Munich, Germany, 8–14 Sept. 2018. – Munich, 2018. – P. 801–818. – https://doi.org/10.1007/978-3-030-01234-2_49.

11. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs / L. C. Chen, G. Papandreou, I. Kokkinos [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2018. – Vol. 40, iss. 4. – P. 834–848. – https://doi.org/10.1109/TPAMI.2017.2699184.

12. Rethinking atrous convolution for semantic image segmentation / L.-C. Chen, G. Papandreou, F. Schroff, H. Adam. – 2017. – URL: https://arxiv.org/pdf/1706.05587 (date of access: 16.03.2026). – https://doi.org/10.48550/arXiv.1706.05587.

13. Encoder-decoder with atrous separable convolution for semantic image segmentation / L. C. Chen, Y. Zhu, G. Papandreou [et al.] // Proc. of the European Conf. on Computer Vision (ECCV), Munich, Germany, 8–14 Sept. 2018. – Munich, 2018. – P. 801–818. – https://doi.org/10.1007/978-3-030-01234-2_49.

14. An image is worth 16x16 words: Transformers for image recognition at scale / A. Dosovitskiy, L. Beyer, A. Kolesnikov [et al.] // Intern. Conf. on Learning Representations (ICLR), Vienna, Austria, 4 May 2021. – Vienna, 2021. – P. 1–21. – https://doi.org/10.48550/arXiv.2010.11929.

15. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers / S. Zheng, J. Lu, H. Zhao [et al.] // Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. – Nashville, 2021. – P. 6881–6890. – https://doi.org/10.1109/CVPR46437.2021.00681.

16. Attention mechanisms in semantic segmentation of remote sensing images / V. V. Godase, S. Takale, R. Ghodake, A. Mulani // Journal of Advancement in Electronics Signal Processing. – 2025. – Vol. 2. – P. 45–58.

17. Hu, J. Squeeze-and-excitation networks / J. Hu, L. Shen, G. Sun // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018. – Salt Lake City, 2018. – P. 7132–7141. – https://doi.org/10.1109/CVPR.2018.00745.

18. CBAM: Convolutional block attention module / S. Woo, J. Park, J. Y. Lee, I. S. Kweon // Proc. of the European Conf. on Computer Vision (ECCV), Munich, Germany, 8–14 Sept. 2018. – Munich, 2018. – P. 3–19. – https://doi.org/10.1007/978-3-030-01234-2_1.

19. Hou, Q. Coordinate attention for efficient mobile network design / Q. Hou, D. Zhou, J. Feng // Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. – Nashville, 2021. – P. 13713–13722. – https://doi.org/10.1109/CVPR46437.2021.01350.

20. Big self-supervised models are strong semi-supervised learners / T. Chen, S. Kornblith, K. Swersky [et al.] // Advances in Neural Information Processing Systems (NeurIPS). – 2020. – Vol. 33. – P. 20735–20747. – https://doi.org/10.48550/arXiv.2006.10029.

21. A comprehensive survey on contrastive learning / H. Hu, X. Wang, Y. Zhang [et al.] // Neurocomputing. – 2024. – Vol. 610. – Art. 128645. – URL: https://www.sciencedirect.com/science/article/abs/pii/S0925231224014164?via%3Dihub (date of access: 16.03.2026). – https://doi.org/10.1016/j.neucom.2024.128645.

22. Supervised contrastive learning / P. Khosla, P. Teterwak, C. Wang [et al.] // Advances in Neural Information Processing Systems (NeurIPS). – 2020. – Vol. 33. – P. 18661–18673. – https://doi.org/10.48550/arXiv.2004.11362.

23. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. – Las Vegas, 2016. – P. 770–778. – https://doi.org/10.1109/CVPR.2016.90.

24. Comparative analysis of deep learning architectures for medical image classification / P.-N. Bui, D.-T. Le, J. Bum [et al.] // Bioengineering. – 2023. – Vol. 10, iss. 10. – Art. 1249. – URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10669434/ (date of access: 16.03.2026). – https://doi.org/10.3390/bioengineering10101249.

25. Kansal, K. ResNet-based deep learning approaches for histopathological breast cancer classification / K. Kansal, A. Singh, D. Srivastava, K. Kansal // 2025 Intern. Conf. on Artificial Intelligence and Emerging Technologies (ICAIET), Bhubaneswar, Odisha, India, 28–30 Aug. 2025. – Bhubaneswar, 2025. – P. 1–4. – https://doi.org/10.1109/ICAIET.2025.11210953.

26. Kordemir, M. A mask R-CNN approach for detection and classification of brain tumours from MR images / M. Kordemir, K. K. Cevik, A. Bozkurt // Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. – 2024. – Vol. 11, iss. 7. – URL: https://www.tandfonline.com/doi/full/10.1080/21681163.2023.2301391 (date of access: 16.03.2026). – https://doi.org/10.1080/21681163.2023.2301391.

27. A comparative analysis of attention mechanisms in 3D CNN-based hyperspectral image super-resolution / M. S. Zitouni, M. Q. Alkhatib, N. Aburaed, H. Ahmad // 2024 14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Helsinki, Finland, 09–11 Dec. 2024. – Helsinki, 2024. – P. 1–5. – https://doi.org/10.1109/WHISPERS.2024.10876507.

28. A new semantic segmentation method for remote sensing images integrating coordinate attention and SPD-Conv / Z. Yang, Q. Wu, F. Zhang [et al.] // Symmetry. – 2023. – Vol. 15, iss. 5. – Art. 1037. – URL: https://www.mdpi.com/2073-8994/15/5/1037 (date of access: 16.03.2026). – https://doi.org/10.3390/sym15051037.

29. Li, Z. A comparative study of Contrastive Self-Supervised Learning (CSSL): Methods, technologies, and applications / Z. Li // Transactions on Computer Science and Intelligent Systems Research. – 2025. – Vol. 10. – P. 92–97. – https://doi.org/10.62051/2hgr2j26.

30. Wang, Y. Multi-label guided supervised contrastive learning for earth observation pretraining / Y. Wang, C. M. Albrecht, X. X. Zhu // 2024 IEEE Intern. Geoscience and Remote Sensing Symp. (IGARSS), Athens, Greece, 7–12 July 2024. – Athens, 2024. – P. 7568–7571. – https://doi.org/10.1109/IGARSS53475.2024.10642289.

31. Searching for MobileNetV3 / A. Howard, M. Sandler, B. Chen [et al.] // Proc. of the IEEE/CVF Intern. Conf. on Computer Vision (ICCV), Seoul, Korea (South), 27 Oct. – 02 Nov. 2019. – Seoul, 2019. – P. 1314–1324. – https://doi.org/10.1109/ICCV.2019.00140.

32. FiLM: Visual reasoning with a general conditioning layer / E. Perez, F. Strub, H. de Vries [et al.] // Proc. of the AAAI Conf. on Artificial Intelligence, New Orleans, Louisiana, USA, 2–7 Febr. 2018. – New Orleans, 2018. – Vol. 32, iss. 1. – P. 3942–3951. – https://doi.org/10.1609/aaai.v32i1.11823.

References

1. Jiang C., Chervan A. N. *Identification of land use structure in the zone of influence of the Soligorsk potash plant based on remote sensing data*. Prirodopolzovanie [*Nature Management*], 2025, no. 1, pp. 51–63 (In Russ.).

2. Roche C., Thygesen K., Baker E. Mine tailings storage: Safety is no accident. *Rapid Response Assessment.* United Nations Environment Programme, GRID-Arendal. Arendal, 2017, pp. 6–10.

3. Rotta L. H. S., Alcântara E., Park E., Negri R. G., Lin Y. N. The 2019 Brumadinho tailings dam collapse: Possible cause and impacts of the worst human and environmental disaster in Brazil. *International Journal of Applied Earth Observation and Geoinformation*, 2020, vol. 90, art. 102119. Available at: https://www.sciencedirect.com/science/article/pii/S0303243420300192?via%3Dihub (accessed: 16.03.2026). https://doi.org/10.1016/j.jag.2020.102119.

4. Santamarina J. C., Torres-Cruz L. A., Bachus R. C. Why coal ash and tailings dam disasters occur. *Science*, 2019, vol. 364, iss. 6440, pp. 526–528. https://doi.org/10.1126/science.aax1927.

5. Balaniuk R., Isupova O., Reece S. Mining and tailings dam detection in satellite imagery using deep learning. *Sensors*, 2020, vol. 20, no. 23, art. 6936. Available at: https://www.mdpi.com/1424-8220/20/23/6936 (accessed 16.03.2026). https://doi.org/10.3390/s20236936.

6. Stelmakh V. I. *Environmental problems of potash salt mining and some solutions*. Obespechenie jekonomicheskoj bezopasnosti Respubliki Belarus': teorija i praktika: materialy Respublikanskoj nauchno-prakticheskoj konferencii, Minsk, 19 dekabrja 2019 g. [*Ensuring the Economic Security of the Republic of Belarus: Theory and Practice: Proceedings of the Republican Scientific-practical Conference, Minsk, 19 December 2019*]. Minsk, Akademija Ministerstva vnutrennih del Respubliki Belarus', 2019, pp. 103–105 (In Russ.).

7. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, Massachusetts, USA, 7–12 June 2015*. Boston, 2015, pp. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.

8. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 : 18th International Conference, Munich, 5–9 October 2015*. In N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (eds.). Cham, 2015, pt. III, pp. 234–241 (Lecture Notes in Computer Science; vol. 9351). https://doi.org/10.1007/978-3-319-24574-4_28.

9. H. Zhao, Shi J., Qi X., Wang X., Jia J. Pyramid scene parsing network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 21–26 July 2017*. Honolulu, 2017, pp. 2881–2890. https://doi.org/10.1109/CVPR.2017.660.

10. Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015*. San Diego, 2015, pp. 1–14. https://doi.org/10.48550/arXiv.1412.7062.

11. Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, vol. 40, iss. 4, pp. 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.

12. Chen L.-C., Papandreou G., Schroff F., Adam H. *Rethinking atrous convolution for semantic image segmentation*, 2017. Available at: https://arxiv.org/pdf/1706.05587 (accessed: 16.03.2026). https://doi.org/10.48550/arXiv.1706.05587.

13. Chen L.-C., Zhu Y., Papandreou G., Schroff F., Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018*. Munich, 2018, pp. 801–818. https://doi.org/10.1007/978-3-030-01234-2_49.

14. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., …, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR), Vienna, Austria, 4 May 2021*. Vienna, 2021, pp. 1–21. https://doi.org/10.48550/arXiv.2010.11929.

15. Zheng S., Lu J., Zhao H., Zhu X., Luo Z., Wang Y. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021*. Nashville, 2021, pp. 6881–6890. https://doi.org/10.1109/CVPR46437.2021.00681.

16. Godase V. V., Takale S., Ghodake R., Mulani A. Attention mechanisms in semantic segmentation of remote sensing images. *Journal of Advancement in Electronics Signal Processing*, 2025, vol. 2, pp. 45–58.

17. Hu J., Shen L., Sun G. Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018*. Salt Lake City, 2018, pp. 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.

18. Woo S., Park J., Lee J. Y., Kweon I. S. CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018*. Munich, 2018, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.

19. Hou Q., Zhou D., Feng J. Coordinate attention for efficient mobile network design. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021*. Nashville, 2021, pp. 13713–13722. https://doi.org/10.1109/CVPR46437.2021.01350.

20. Chen T., Kornblith S., Swersky K., Norouzi M., Hinton G. Big self-supervised models are strong semi-supervised learners. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020, vol. 33, pp. 20735–20747. https://doi.org/10.48550/arXiv.2006.10029.

21. Hu H., Wang X., Zhang Y., Chen Q., Guan Q. A comprehensive survey on contrastive learning. *Neurocomputing*, 2024, vol. 610, art. 128645. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0925231224014164?via%3Dihub (accessed 16.03.2026). https://doi.org/10.1016/j.neucom.2024.128645.

22. Khosla P., Teterwak P., Wang C., Sarna A., Tian Y., …, Krishnan D. Supervised contrastive learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020, vol. 33, pp. 18661–18673. https://doi.org/10.48550/arXiv.2004.11362.

23. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*. Las Vegas, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

24. Bui P.-N., Le D.-T., Bum J., Kim S., Song S. J., Choo H. Comparative analysis of deep learning architectures for medical image classification. *Bioengineering*, 2023, vol. 10, iss. 10, art. 1249. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10669434/ (accessed 16.03.2026). https://doi.org/10.3390/bioengineering10101249.

25. Kansal K., Singh A., Srivastava D., Kansal K. ResNet-based deep learning approaches for histopathological breast cancer classification. *2025 International Conference on Artificial Intelligence and Emerging Technologies (ICAIET), Bhubaneswar, Odisha, India, 28–30 August 2025*. Bhubaneswar, 2025, pp. 1–4. https://doi.org/10.1109/ICAIET.2025.11210953.

26. Kordemir M., Cevik K. K., Bozkurt A. A mask R-CNN approach for detection and classification of brain tumours from MR images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2024, vol. 11, iss. 7. Available at: https://www.tandfonline.com/doi/full/10.1080/21681163.2023.2301391 (accessed 16.03.2026). https://doi.org/10.1080/21681163.2023.2301391.

27. Zitouni M. S., Alkhatib M. Q., Aburaed N., Ahmad H. A comparative analysis of attention mechanisms in 3D CNN-based hyperspectral image super-resolution. *2024 14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Helsinki, Finland, 09–11 December 2024*. Helsinki, 2024, pp. 1–5. https://doi.org/10.1109/WHISPERS.2024.10876507.

28. Yang Z., Wu Q., Zhang F., Zhang X., Chen X., Gao Y. A new semantic segmentation method for remote sensing images integrating coordinate attention and SPD-Conv. *Symmetry*, 2023, vol. 15, iss. 5, art. 1037. Available at: https://www.mdpi.com/2073-8994/15/5/1037 (accessed 16.03.2026). https://doi.org/10.3390/sym15051037.

29. Li Z. A comparative study of Contrastive Self-Supervised Learning (CSSL): Methods, technologies, and applications. *Transactions on Computer Science and Intelligent Systems Research*, 2025, vol. 10, pp. 92–97. https://doi.org/10.62051/2hgr2j26.

30. Wang Y., Albrecht C. M., Zhu X. X. Multi-label guided supervised contrastive learning for earth observation pretraining. *2024 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Athens, Greece, 7–12 July 2024*. Athens, 2024, pp. 7568–7571. https://doi.org/10.1109/IGARSS53475.2024.10642289.

31. Howard A., Sandler M., Chen B., Wang W., Chen L.-C., Tan M. Searching for MobileNetV3. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 27 October – 02 November 2019*. Seoul, 2019, pp. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.

32. Perez E., Strub F., Vries H. de, Dumoulin V., Courville A. FiLM: Visual reasoning with a general conditioning layer. *Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, 2–7 February 2018*. New Orleans, 2018, vol. 32, iss. 1, pp. 3942–3951. https://doi.org/10.1609/aaai.v32i1.11823.

Информация об авторах

Козлов Анатолий Анатольевич, аспирант кафедры веб-технологий и компьютерного моделирования, Белорусский государственный университет.
E-mail: anatoly.kozlov.ak@yandex.ru

Абламейко Сергей Владимирович, доктор технических наук, академик Национальной академии наук Республики Беларусь, профессор кафедры веб-технологий и компьютерного моделирования, Белорусский государственный университет.
E-mail: ablameyko@bsu.by

Information about the authors

Anatoly A. Kozlov, Postgraduate Student at the Department of Web Technologies and Computer Modeling, Belarusian State University.
E-mail: anatoly.kozlov.ak@yandex.ru

Sergey V. Ablameyko, Dr. Sci. (Eng.), Acad. of the National Academy of Sciences of Belarus, Prof. at the Department of Web Technologies and Computer Modeling, Belarusian State University.
E-mail: ablameyko@bsu.by