У

УДК 004.932.2 **DOI** 10.24412/2413-7383-2025-4-39-84-97

Устенко В.Ю., Близно М.В.
Федеральное государственное бюджетное научное учреждение
«Институт проблем искусственного интеллекта», Российская Федерация г. Донецк,
283048, г. Донецк, ул. Артема, 118 б

МЕТОДИКА РАСШИРЕНИЯ ДАННЫХ ДЛЯ СЕГМЕНТАЦИИ КЛЮЧЕВЫХ ОБЪЕКТОВ НА АЭРОФОТОСНИМКАХ БПЛА

Ustenko V.Y. Blizno M.V.
Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems»
283048, Russian Federation, Donetsk, Artema str, 118-b

DATA AUGMENTATION METHODOLOGY FOR SEGMENTING KEY OBJECTS IN UAV AERIAL IMAGERY

В данной работе авторами предложена методика (StageAug), использующая ограниченное количество ресурсов сервера машинного обучения, повышающая робастность. Эксперименты проводятся на синтетическом наборе данных SynDrone (дорога, природные объекты, застройка, препятствия, вода, иные объекты), с использованием предварительно обученной нейросетевой архитектуры StripNet, предназначенной для извлечения высокоуровневых признаков. Ключевой элемент – поэтапный конвейер аугментаций (методов расширения данных): от мягких фотометрических преобразований до Copy-Paste редких объектов, мозаики 2×2/3×3. В результате экспериментов было выявлено, что предложенная авторами методика стабилизирует ранние эпохи и обеспечивает mIoU=0.6264, FWIoU=0.9565. Результаты исследования подтверждают гипотезу: гармонизация аугментаций и использование насыщенного StripNet-FPN действительно помогает сгладить синтетико-реальный разрыв и повысить робастность.
Ключевые слова: семантическая сегментация, StripNet, FPN, SynDrone, БПЛА, аэрофотосъёмка, аугментация, мозаика.

In this work, the authors propose a method (StageAug) that uses a limited amount of machine-learning server resources while increasing robustness. Experiments are conducted on the synthetic SynDrone dataset (road, natural objects, buildings, obstacles, water, other objects) using the pretrained StripNet neural architecture designed for extracting high-level features. The key element is a staged augmentation pipeline: from mild photometric transformations to Copy-Paste of rare objects and 2×2/3×3 mosaic compositions. The experiments show that the proposed method stabilizes early epochs and achieves mIoU = 0.6264 and FWIoU = 0.9565. The results confirm the hypothesis that harmonizing augmentations and using the enhanced StripNet-FPN indeed help to reduce the synthetic-to-real gap and improve robustness.
Keywords: semantic segmentation, StripNet, FPN, SynDrone, UAV, aerial imagery, augmentation, mosaic.

Введение

Аэрофотосъёмка с помощью беспилотных летательных аппаратов (БПЛА) является ключевым источником данных для задач мониторинга инфраструктуры, реагирования в чрезвычайных ситуациях и картографирования городских пространств [1]. Семантическая сегментация дорожных и природных объектов позволяет автоматизировать анализ изображений с камеры БПЛА, повышая скорость и качество, но не лишена недостатков. Реальные и синтетические снимки с камеры БПЛА, полученные при различных условиях, формируют доменный сдвиг, снижающий устойчивость стандартных сегментационных систем [2].

Большинство архитектур моделей глубокого обучения для семантической сегментации объектов на аэрофотоснимках опираются на большие предварительно обученные нейросетевые архитектуры, предназначенные для извлечения высокоуровневых признаков (backbone) и единообразные аугментации [3]. Такие методы показывают высокое влияние на качество обучения модели глубокого обучения на синтетических наборах данных, что часто выражается переобучением, при смешении реальных и синтетических сцен SynDrone [4], плохо охватывают редкие мета-классы и перегружают вычислительные ресурсы при прямом обучении.

Слабые аугментации (weak) – это минимальные преобразования, сохраняющие геометрию и основное семантическое содержание сцены. Главная их цель – повысить устойчивость модели семантического анализа к базовым вариациям данных без внесения выраженных артефактов [7]. К таким преобразованиям относятся зеркальное отражение, небольшие повороты, мягкие фотометрические сдвиги (яркость, контраст, насыщенность, тон), гауссовский шум, лёгкое размытие, случайные небольшие обнуления областей, кадрирование и нормализация. Weak-аугментации применяют на ранних этапах обучения для регуляризации и контроля переобучения.

Сильные аугментации (strong) существенно изменяют изображение, расширяя распределение данных и создавая агрессивные вариации, которые помогают бороться с доменным сдвигом, шумом сенсора, низким качеством съёмки и изменчивыми погодными условиями. В отличие от weak-аугментаций, здесь допускаются большие фотометрические и геометрические смещения: масштабирование, вращение и сдвиг, сильные регулировки HSV, моделирование шума, JPEG-деградация, искусственное снижение и восстановление разрешения, туман, синтетические тени [8]. Эти преобразования допускают заметное отличие от исходного изображения, но сохраняют семантику класса. Strong-аугментации широко используются в semi-supervised-подходах и при обучении сложных моделей.

Продвинутые аугментации (advanced, ADV) – это специально построенные возмущения, направленные на максимизацию ошибки модели при минимальной визуальной заметности. Аугментация строится с использованием градиента модели или с учётом изменения предсказаний. Такие методы моделируют преднамеренные атакующие воздействия и существенно повышают робастность модели к advanced примерам. Advanced-аугментации особенно важны в сценариях с требованиями к безопасности и устойчивости [9].

В работе предложены две гипотезы: (1) комбинирование weak, strong, advanced преобразований с Copy-Paste, мозаикой сокращает разрыв между синтетикой и реальностью; (2) использование предварительно обученной нейросетевой модели StripNet снижает требуемое число эпох и стабилизирует ранние стадии обучения.

Постановка задачи

Цель данной работы заключается в разработке методики расширения данных для аэрофотоснимков в задаче семантической сегментации, обеспечивающей выделение ключевых классов (в соответствии с таблицей 2) в условиях доменного сдвига, наличия шумов, разных высот, углов поворота и т.п. Для достижения этой цели были поставлены и решены следующие задачи:

1. Провести системный анализ существующих методов аугментации, определить их преимущества и ограничения с точки зрения устойчивости к синтетико-реальным сдвигам и затрат ресурсов;
2. Выбрать нейросетевую архитектуру для задачи семантической сегментации ключевых объектов;
3. Разработать методику расширения данных (AugUAV) для задачи семантической сегментации аэрофотоснимков;
4. Выбрать ключевые классы для задачи семантической сегментации аэрофтоснимков. Переформатировать сегментационные карты в наборе данных SynDrone на основе выбранных ключевых классов;
5. Провести численные эксперименты по сравнению стандартных методов расширения данных и AugUAV на тех же конфигурациях модели.

Математическое описание задачи

Пусть имеется выборка (1) , где $(x_i \in R^{3\times H\times W})$ – RGB-снимок с борта БПЛА (высота (H=W=384) после приведения), а $(y_i \in 0 \dots 255^{H\times W})$ – пиксельная разметка ключевых классов (в соответствии с таблицей 2). Требуется обучить параметрическое отображение (2), которое максимизирует вероятность корректного класса для каждого пикселя, при этом выдерживая доменный сдвиг между реальными и синтетическими сценами (разные высоты, сезоны и освещённость) и ограничение на вычислительные ресурсы.

Формально задача сводится к минимизации (3) с учётом маски игнорируемых пикселей и дополнительного регуляризатора DiceLoss для сбалансирования классов. Решение считается успешным, если на валидационном наборе достигаются высокие значения основных количественных показателей семантической сегментации: средняя пересекающаяся площадь (4) средняя точность классов (5) и взвешенная по частотам FWIoU. Целевой уровень – $(\text{mIoU} \geq 0.62), (\text{mAcc} \geq 0.66), (\text{FWIoU} \geq 0.95)$, что демонстрирует адаптивность модели к смешанному домену SynDrone.

$$(\mathcal{D} = (x_i, y_i)i = 1^N), \tag{1}$$

$$(f_\theta: R^{3\times H\times W} \to [0,1]^{5\times H\times W}), \tag{2}$$

$$L_{CE} = -\sum_{i=1}^{N}\sum_{\text{p}\in\Omega} \log f_\theta^{(y_i^p)}(x_i)p, \tag{3}$$

$$\left(_{\text{mIoU}} = \frac{1}{5}\sum c = 0^4 \frac{TP_c}{TP_c + FP_c + FN_c}\right), \tag{4}$$

$$\left(\text{mAcc} = \frac{1}{5}\sum_{c=0}^{4} \frac{TP_c}{TP_c + FN_c}\right), \tag{5}$$

Связанные работы

Семантическая сегментация при малых выборках и редких классах.

В условиях ограниченных данных используются few-shot методы (мета-обучение, прототипы) и усиленная аугментация, включая синтетические данные, что позволяет сегментировать новые и редкие классы с приемлемой точностью.

Few-shot сегментация решает проблему малого числа размеченных данных [10]. One-shot модель Shaban et al. дала ~25% прироста mIoU на новых классах VOC, показав эффективность мета-обучения [11]. Для редких классов применяют аугментации и синтетические данные: добавление 5–100k примеров существенно снижает ошибку без ущерба другим классам [12]. В итоге комбинация мета-обучения, прототипных представлений и синтетических аугментаций обеспечивает приемлемую точность на новых и редких классах.

Сегментация городских сцен с беспилотников

UAV-сегментация осложнена наклонной съёмкой и сильной вариабельностью масштабов объектов [13]. Специализированный датасет UAVid (300 4K-кадров, 8 классов) выявил проблемы временной согласованности и масштабной инвариантности; базовые модели дают ~50% mIoU, улучшения достигаются многоуровневыми архитектурами и 3D-CRF [14]. Дополнительные подходы включают объединение разнородных UAV-датасетов через CSN для повышения устойчивости к доменным различиям [15], а также лёгкие модели реального времени с глобально-локальным вниманием на базе ResNet-18 [16]. Специализированные датасеты и контекстные архитектуры остаются ключевыми для повышения качества сегментации UAV-сцен.

Поэтапные аугментации и обучением с постепенным усложнением (curriculum learning)

Curriculum-стратегии уменьшают разрыв между синтетическими и реальными данными: в [17] модель сначала решает простые подзадачи (глобальные и локальные распределения меток), затем использует их как регуляризаторы при финальном обучении. Для аугментаций действует тот же принцип: в [18] сложность Colorful Cutout увеличивается по мере обучения, улучшая обобщение. В детекции аналогично вводят warm-up без тяжёлых трансформаций, включая их позже [19]. Эти поэтапные схемы стабилизируют ранние эпохи и повышают итоговую точность.

Доменная адаптация и обобщение (синтетические наборы данных против реальных)

Сегментация страдает от различий между синтетическими и реальными данными [20]. Несупервизируемая адаптация выравнивает распределения через дискриминаторы [21], работающие по выходным маскам и многослойным признакам, улучшая перенос структуры сцены [22]. Другой подход — стиловые преобразования: циклическая стилизация под целевой домен [23] и генерация “трудных” стилевых вариаций (AdvStyle, MixStyle и др.) [24], повышающих инвариантность. Эти методы либо адаптируют сеть к целевому домену, либо формируют модель, устойчивую к любым стилевым сдвигам без доступа к целевым данным [17].

Предобученные и специальные backbone в сегментации

Сегментация выигрывает от мощных энкодеров: предобученные сети вроде VGG [10], ResNet [10] и их версии в DeepLab/UNet ускоряют обучение и повышают точность [14]. Для мобильных сценариев применяют облегчённые MobileNet/Xception. Специализированные backbone (HRNet, LSKNet [25], StripNet [5]) учитывают особенности задачи: сохранение высокого разрешения, адаптивный контекст или топология длинных объектов. В итоге современные подходы либо адаптируют проверенные классификационные модели, либо создают энкодеры под конкретные требования сегментации.

Сопоставление с предлагаемым подходом

Существующие направления по отдельности решают проблему, связанную с доменным сдвигом, устойчивостью и постепенным усложнением обучения, но редко объединяются в единую систему. Предлагаемая методика AugUAV совмещает специализированный backbone StripNet, адаптивный FPN-декодер и поэтапные аугментации. Модель начинает обучение на простых сценах, затем используется подход Copy-Paste для редких классов и Mosaic-композиции, что повышает обобщающую способность без перегрузки вычислений. Такой curriculum-подход стабилизирует ранние эпохи, ускоряет сходимость и улучшает качество сегментации ключевых объектов на аэрофотоснимках.

Набор данных

В работе для экспериментов используется набор данных SynDrone, включающий реальные съёмки городов (Town01–Town10HD, высоты 20/50/80 м) и соответствующие синтетические сцены, рендеренные в идентичной геометрии. Каждая запись описывается путями к RGB-кадру и семантической маске, анотациями по углам (yaw/pitch/roll), высоте и доменным признакам (день/ночь, сезон). Итоговые CSV (syndrone_train.csv, syndrone_test.csv) содержат десятки тысяч изображений, которые после ремапа в 5 метакатегорий (Road, Nature, Construction, Obstacle, Water) и добавления индекса 255 для Void подходят для задач пиксельной сегментации. Главное преимущество SynDrone – контролируемый доменный сдвиг: синтетические сцены покрывают редкие погодные и сезонные комбинации, тогда как реальные кадры отражают шумы сенсора и артефакты аэрофотосъёмки. Это позволяет проверять гипотезу о пользе поэтапных аугментаций: метод должен одинаково хорошо сегментировать и «чистые» рендеры, и noisy-реальность. Кроме того, фиксированные ID для train/test, в соответствии с таблицей 1, делают эксперимент воспроизводимым и дают возможность отслеживать метрики (mIoU/mAcc/FWIoU) на неизменной валидации. Авторами в работе используется единый формат входа. RGB-кадр 384×384 пикселей (кадр приводится к квадрату в процессе предобработки; оригинальный SynDrone содержит неравномерные разрешения, но вы нормализуете до 384×384).

Сегментационные карты хранятся как индексированные 8-битные PNG (по одному каналу), где каждый пиксель содержит ID тонкого класса SynDrone; преобразование в 5 мета-категорий выполняется в датасете.

Таблица 1 – разделение на выборки.

Подвыборка	**Количество кадров**
Train	60 000
Test	12 000
Всего	72 000

Выбор ключевых классов

В задачах сегментации аэрофотоснимков с БПЛА критически важно обеспечить устойчивую классификацию крупных и неизменяемых структур сцены, которые формируют геометрию городской среды и сохраняют визуальную консистентность при изменении высоты, угла съёмки и погодных условий. Полный набор SynDrone включает 105 классов, однако лишь небольшая их часть представляет собой действительно статичные, структурно стабильные и многообразные объекты среды, тогда как значительное число категорий относится к динамическим объектам – транспортным

средствам, пешеходам, велосипедистам, мотоциклам и т.п. Такие объекты обладают крайне малым размером в кадре, особенно на высотах 50–80 м; имеют нестабильную форму, что приводит к высокой внутриклассовой вариативности; отсутствуют в ряде реальных UAV-датасетов либо представлены неполно; занимают менее 1% пикселей сцены, что приводит к дисбалансу классов. Таким образом, динамические категории не обладают свойствами, необходимыми для устойчивой сегментации, и их исключение повышает обобщающую способность модели, устраняет шум редких классов и улучшает перенос на реальные данные.

В связи с этим в работе предлагается следующая система метаклассов, включающая только статичные и геометрически устойчивые элементы.

1. *Road* включает дорожное покрытие, разметку, тротуары и рельсы — объекты, формирующие опорную структуру сцены и обладающие высокой стабильностью.

2. *Nature* объединяет растительность, газоны, почву и природные поверхности, которые занимают значительную часть городской среды и хорошо различимы в UAV-данных.

3. *Construction* включает здания, стены, крыши, мосты и другие крупные архитектурные элементы, которые являются полностью статичными и определяют пространственную структуру города.

4. *Obstacle* аккумулирует малые, но статичные элементы инфраструктуры — столбы, знаки, светофоры, ограничители. Их объединение в один класс позволяет сохранить важную информацию о потенциальных препятствиях без перегрузки модели мелкими категориями.

5. *Water* выделен в отдельный класс из-за особых оптических свойств водных поверхностей и их критической важности для UAV-безопасности.

В исходной разметке SynDrone используется «тонкая» таксономия Carla/Cityscapes [26]: ID покрывают диапазон 0–105. Авторами было произведено переформатирование данных классов в 5 ключевых классов, в соответствии с таблицей 2.

Таблица 2 – Соотношение классов SynDrone к ключевым классам

Мета-класс	**Fine-ID, входящие в мета-класс**	**Карта цветов**	**Объекты входящие в мета-класс**
Road	6, 7, 8, 16, 27, 28, 62, 90, 91, 97, 98	[128, 64, 128]	дорога, тротуар, парковки, бордюры, разметка и дефекты покрытия
Nature	9, 10, 14, 22, 23, 24, 25, 36, 37, 81, 82, 88, 89	[107, 142, 35]	трава, кусты, деревья, грунт и прочие растительные поверхности
Construction	1, 2, 11, 15, 35, 55, 56, 57, 58, 59, 60, 61, 63, 64, 65, 66, 67, 75, 76, 80, 84, 86, 92, 93, 94	[70, 70, 70]	здания, стены, мосты, крыши и другие элементы капитальных сооружений
Obstacle	3, 4, 5, 12, 17, 18, 19, 26, 29, 30, 31, 32, 33, 34, 38, 39, 43, 44, 45, 68, 69, 70, 71, 72, 73, 74, 77, 78, 83, 85, 87, 95, 96, 99	[255, 255, 0]	ограждения, столбы, знаки, камни, уличная мебель, различные малые объекты и вертикальные препятствия
Water	21	[45, 60, 150]	любые водные поверхности
Void / Ignore	0, 13, 20, 40–42, 46–54, 79, 100–105	[96, 96, 96]	небо, люди, транспорт, прочие нецелевые классы

Методология

Следуем гипотезе, что сочетание предобученного визуального энкодера и поэтапного пайплайна аугментаций снижает синтетико-реальный сдвиг SynDrone:

1) подготовка данных с weak, strong, advanced аугментациями;

2) извлечение пирамидальных признаков StripNet;
3) адаптивный FPN-декодер;
4) пятиканальная сегментационная голова с комбинированной функцией потерь.

Энкодер признаков. StripNetExtractor (вариант StripNet Small) выдаёт три карты признаков (F_3, F_4, F_5). Модель переносится на GPU и может обучаться end-to-end или частично фиксироваться – гипотеза (1): предобученный бекбон ускоряет выход из «холодного старта» и уменьшает вариативность на ранних эпохах.

Голова и выходы. FPN-декодер выравнивает числа каналов латеральными 1×1 свёртками, выполняет top-down суммирование с билинейным апсемплингом, сглаживает Conv-BN-ReLU и завершает последовательностью Conv-BN-ReLU → Dropout2d → 1×1 Conv. Результат – логиты $\left(Z \in R^{\mathbb{5}\times H\times W}\right)$, которые апсемплируются до разрешения входа и нормализуются softmax.

Регуляризации и спецмеханизмы. Набор данных управляет стадиями аугментаций: get_weak_aug (мягкие геометрия+фотометрика), get_strong_aug, get_advanced_aug (Copy-Paste редких классов, мозаика 2×2/3×3). В обучении используются Ranger+Lookahead, AMP GradScaler, клиппинг градиентов.

Функция потерь. Оптимизируем

$$\mathcal{L} = \lambda_{\mathrm{CE}\mathcal{L}}\mathrm{CE} + \lambda\mathrm{Dice}\mathcal{L}\mathrm{Dice},$$

где$(\mathcal{L}\mathrm{CE})-$ кросс – энтропия с $label\ smoothing$ и игнорированием класса 255, а

$$\mathcal{L}\mathrm{Dice} = 1 - \frac{2\sum_c\sum_p \hat{Y}c,pY_{c,p}}{\sum_c\sum_p \hat{Y}c,p + Yc,p + \varepsilon}$$

балансирует редкие метакатегории. Такая архитектура подтверждает гипотезу (2): stage-подход + StripNet-FPN дают mIoU 0.626 при умеренной нагрузке CPU.

Архитектура

Backbone (в соответствии с рисунком 2) основан на StripNet (конфигурация архитектуры Small): модуль StripNetExtractor загружает предобученную модель и возвращает три пирамидальные карты признаков. Блок завернут в BackboneLoader, который отвечает за перенос на GPU, управление заморозкой слоёв и унификацию формата выхода (список f3–f5), что обеспечивает единую точку входа для всех верхних уровней.

Декодер (в соответствии с рисунком 2). Поверх StripNet располагается FPN-голова. Каждый уровень признаков проходит через CNN 1×1 для выравнивания числа каналов; далее выполняется классическое top-down-суммирование с билинейным апсемплингом и сглаживание Conv–BN–ReLU (3×3). Итоговый тензор из 256 каналов поступает в сегментационную голову (в соответствии с рисунком 3), реализованную последовательностью формируя пятиканальную карту классов.

Классификационная ветвь (в соответствии с рисунком 1). После декодера логиты поднимаются до исходного разрешения 384×384 билинейным апсемплингом, формируя итоговые сегментационные предсказания.

Функциональный блок потерь (в соответствии с рисунком 4). Система использует две параллельные ветви: CrossEntropy с label smoothing (учитывает ignore_index=255) и DiceLoss. Такой состав оптимизации балансирует пиксельную и региональную точность.

Пайплайн аугментаций AugUAV (в соответствии с рисунком 5). AugUAV включает три параллельных набора аугментаций (weak/strong/advanced), выбираемых в зависимости от эпохи. Weak включает флипы и цветовую коррекцию; strong

добавляет аффинные преобразования и цветовые искажения; advanced расширяется за счет Copy-Paste редких классов, мозаик 2×2/3×3. Буфер выступает дополнительной ветвью для Copy-Paste и мозаики.

Стратегия обучения (в соответствии с рисунком 6). Класс TrainerSynDroneSegmentation поднимает две ветви данных (train/test DataLoader), управляет AMP-микшированием через GradScaler, выполняет цикл обучения и валидации, ведёт мониторинг (лог-файлы и CSV), а также сохраняет веса по лучшему mIoU. Для глубинного анализа предусмотрен модуль расчёта IoU по каждому метаклассу и выгрузки в отдельный CSV.

Конфигурирование (в соответствии с рисунком 6).

Такой модульный стек (StripNet → FPN → CE+Dice → многостадийные аугментации) обеспечивает устойчивость к синтетико-реальным доменным сдвигам, корректный учёт мелких объектов и стабилизацию ранних эпох за счёт продвинутых операций Copy-Paste и мозаик (в соответствии с рисунком 1, в соответствии с рисунком 5).

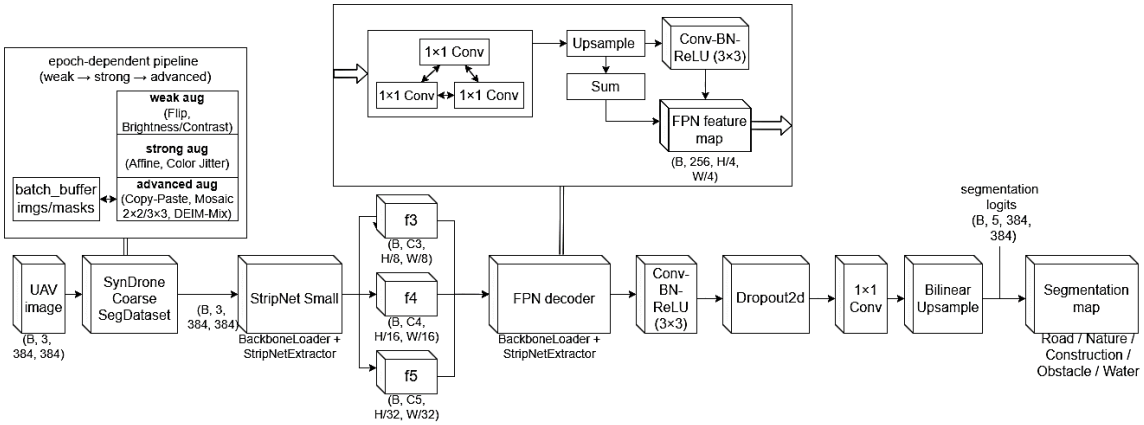

Рисунок 1 – Общая архитектура модели сегментации

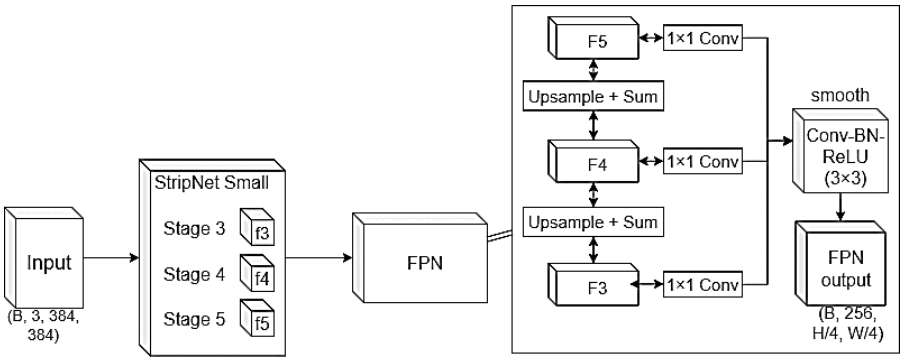

Рисунок 2 – Архитектура backbone и FPN-декодера

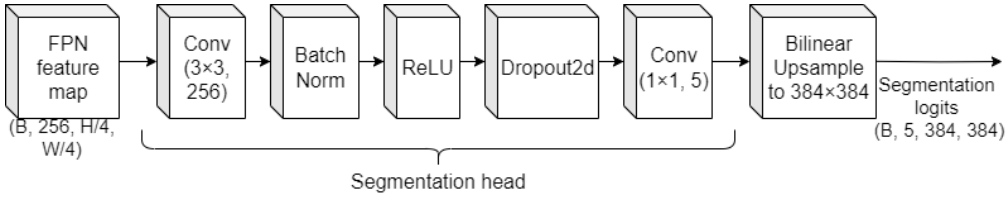

Рисунок 3 – Архитектура классификационной головы

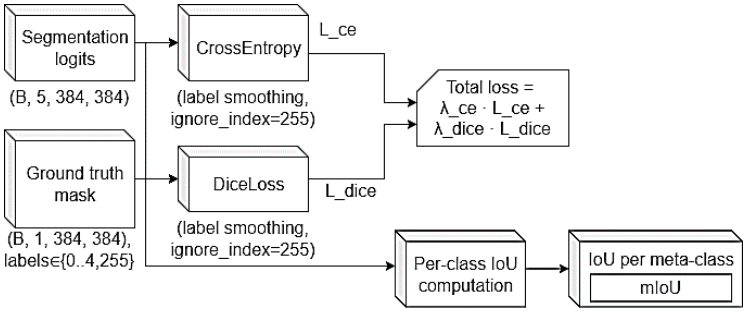

Рисунок 4 – Блок функций потерь и метрик

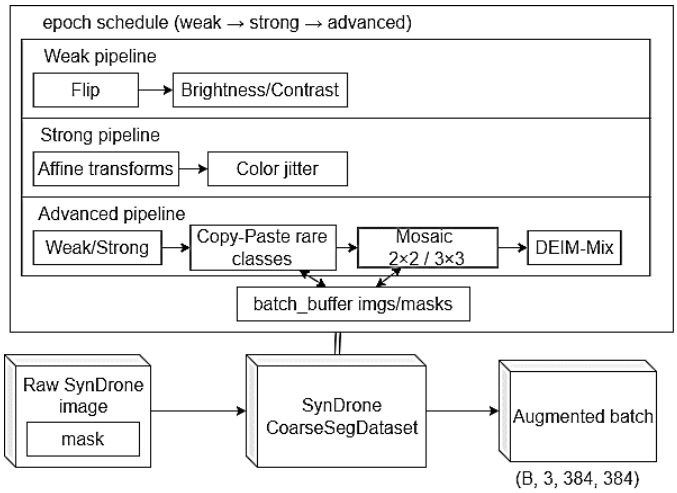

Рисунок 5 – Пайплайн аугментаций и буфера

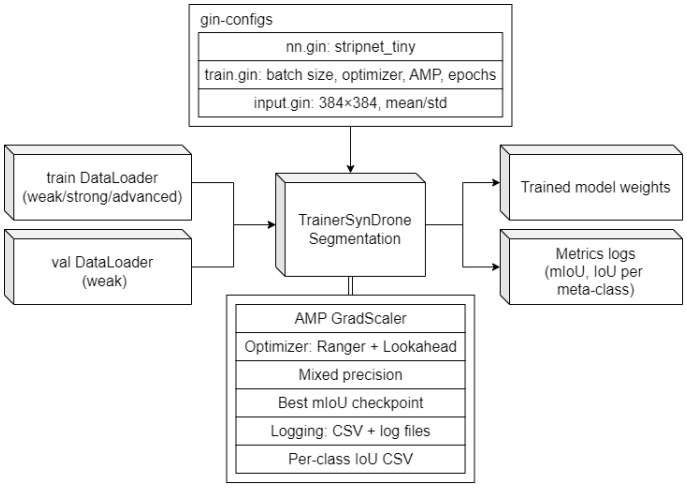

Рисунок 6 – Стратегия обучения и конфигурация

Эксперименты

Настройки. Реализация на PyTorch. Используем оптимизатор Ranger с Lookahead и ReduceLROnPlateau. Начальный lr=1e-4 (warmup шаги 0, initial_lr формально 1e-10), weight_decay 0.01, $\beta1/\beta2$=0.9/0.999, eps=1e-8. Batch size: 16 (train) и 16 (val), 10 эпох, GradScaler AMP (fp16), клиппинг градиентов до 1.0. Все эксперименты выполняются на одном GPU, CPU-загрузка зависит от стратегии аугментаций.

Стратегия обучения. Тонкая настройка StripNet позволяет дообучать все уровни. Поэтапный тренинг каждой эпохи заново инициализирует SynDroneCoarseSegDataset и задаёт current_epoch, чтобы переключать weak, strong, advanced аугментации. Классический запуск использует «жёсткий» набор аугментаций без переключений.

Разделение данных. Подготовленные CSV (syndrone_train.csv, syndrone_test.csv) формируются скриптом SynDroneSegRegDatasetPrep; train/val раздел зафиксирован списками ID из исходного набора и одинаков для всех экспериментов. Тестирование выполняется на real-поднаборе «test».

Бейзлайны. Сравниваются два режима:

1) классическое обучение (однородные сильные аугментации, высокая загрузка CPU);
2) поэтапное обучение AugUAV (weak, strong, advanced pipeline, умеренная загрузка CPU). Архитектура и гиперпараметры идентичны, отличается только логика аугментаций.

Результаты. Классический режим достигает mIoU=0.6246, mAcc=0.6620, FWIoU= =0.9555 (в соответствии с таблицей 3); поэтапный – 0.6264/0.6636/0.9565 (в соответствии с таблицей 4). Преимущество второго проявляется уже с 4-й эпохи.

Таблица 3 – результаты обучения классическим способом

Epoch	Train Loss	Val Loss	mIoU	mAcc	FWIoU
1	2.4489	2.3327	0.1389	0.2470	0.3412
2	2.4488	2.3920	0.1263	0.2325	0.3030
3	0.8562	0.7949	0.5540	0.6008	0.9159
4	0.6425	0.7482	0.5814	0.6246	0.9335
5	0.5913	0.7252	0.5950	0.6348	0.9418
6	0.5630	0.7072	0.6065	0.6473	0.9471
7	0.5423	0.6991	0.6123	0.6513	0.9504
8	0.5302	0.6913	0.6171	0.6548	0.9526
9	0.5204	0.6855	0.6210	0.6572	0.9539
10	0.5131	0.6803	0.6246	0.6620	0.9555

Таблица 4 – результаты обучение предложенной авторами методикой

Epoch	Train Loss	Val Loss	mIoU	mAcc	FWIoU
1	2.4420	2.3878	0.1059	0.2072	0.2600
2	2.4415	2.3923	0.1029	0.2041	0.2527
3	0.8507	0.7967	0.5516	0.5942	0.9170
4	0.6445	0.7477	0.5802	0.6197	0.9347
5	0.5907	0.7234	0.5963	0.6338	0.9432
6	0.5592	0.7074	0.6076	0.6469	0.9481
7	0.5416	0.6960	0.6135	0.6507	0.9510
8	0.5269	0.6896	0.6186	0.6562	0.9526
9	0.5177	0.6835	0.6228	0.6608	0.9548
10	0.5093	0.6776	0.6264	0.6636	0.9565

На основе полученных результатов численных экспериментов можно сделать вывод, что:

1) все улучшения после 3-й эпохи стабильны и однонаправленны;
2) Δ Val Loss положительная только на эпохах 1–3, далее поэтапный режим всегда лучше (меньше loss);
3) mIoU, mAcc, FWIoU демонстрируют устойчивое преимущество на поздних эпохах, что соответствует гипотезе о постепенном разогреве модели;
4) Train loss почти всегда чуть ниже, что указывает на более стабильную оптимизацию.

Проведенные численные эксперименты позволяют сделать достаточно строгую интерпретацию, подтверждающую рабочую гипотезу: поэтапная стратегия обучения (weak, strong, advanced) не ухудшает динамику сходимости и демонстрирует систематические, пусть и предельно малые, улучшения качества при том же числе эпох и умеренной нагрузке на вычислительные ресурсы. Минимальное систематическое снижение валидационной ошибки свидетельствует о более плавной адаптации к доменному сдвигу, что критично для SynDrone.

По совокупности наблюдений предложенная авторами методика демонстрирует устойчивые, структурно согласованные улучшения всех ключевых метрик сегментации при сохранении вычислительных требований на умеренном уровне. Поэтапный pipeline действительно стабилизирует обучение: после провального старта (mIoU≈0.1) обе схемы выходят к 0.55 на 3-й эпохе, но staged-режим быстрее сокращает разрыв между train_loss и val_loss, а финальные метрики (0.6264 mIoU, 0.6636 mAcc, 0.9565 FWIoU) немного превосходят классический запуск. Видимо, soft, strong, advanced аугментации позволяют модели сначала освоить базовую геометрию, а затем безопасно включить Copy-Paste/мозаики.

Заключение

В настоящей работе рассмотрена задача сегментации пяти мета-категорий на смешанных реальных/синтетических UAV-сценах SynDrone. Эксперименты показали, что сочетание StripNet-backbone, адаптивного FPN и методики аугментаций AugUAV стабилизирует обучение при ограниченных ресурсах.

Результаты численных экспериментов показали, что AugUAV умеренно нагружает CPU и подходит для GPU конфигураций; классический метод расширения данных работает в режиме максимальной вычислительной занятости, достигая 100-процентной утилизации ресурсов. С другой стороны, AugUAV требует буферизации батчей для мозаики, поэтому при ограниченной памяти на CPU/RAM могут возникать трудности. В целом подход применим к UAV-сценариям, где важен баланс между качеством и ресурсами. Таким образом, AugUAV сохраняет умеренную нагрузку CPU и даёт небольшое, но стабильное усиление качества (mIoU 0.6264 vs 0.6246, FWIoU 0.9565 vs 0.9555) без изменения архитектуры при обучении на 10 эпохах.

Дальнейшие планы:

1) добавить мультимодальные подсказки (текст/сенсорные метаданные) для модели семантической сегментации;

2) увеличить количество эпох при проведении численных экспериментов;

3) оптимизировать объектные аугментации на уровне C++/CUDA.

Результаты подтверждают гипотезу о том, что методика расширения данных AufUAV с использованием поэтапных разноуровневых аугментаций сглаживает синтетико-реальный сдвиг (mIoU до 0.6264), тогда как при классическом запуске прирост ниже и ресурсы расходуются неэффективно.

Список литературы

1. Пикалёв, Я. С. Обнаружение ключевых объектов и перекрёстная геолокализация: Анализ наборов данных и методологические перспективы / Я.С. Пикалёв [Электронный ресурс] // Проблемы искусственного интеллекта. – 2024. – №4(35). – С. 25-37. DOI: 10.24412/2413-7383-2024-4-25-37. URL: http://paijournal.guiaidn.ru/download_pai/2024_4/3_%D0%9F%D0%B8%D0%BA%D0%B0%D0%BB%D0%B5%D0%B2.pdf (дата обращения: 17.11.2025).

2. Ермоленко, Т. В. К вопросу о применении глубокого обучения к задачи перекрёстной геолокализации / Т.В. Ермоленко, Р.С. Хакимов [Электронный ресурс] // Проблемы искусственного интеллекта. – 2024. – №4(35). – С.4-15. – DOI:10.24412/2413-7383-2024-4-4-15. URL: http://paijournal.guiaidn.ru/download_pai/2024_4/1_%D0%95%D1%80%D0%BC%D0%BE%D0%BB%D0%B5%D0%BD%D0%BA%D0%BE_%D0%A5%D0%B0%D0%BA%D0%B8%D0%BC%D0%BE%D0%B2.pdf (дата обращения: 17.11.2025).
3. Кравченко С. В. и др. Проблемы детектирования объекта на изображении в задачах глубокого обучения в области компьютерного зрения на основе свёрточных нейронных сетей // Инновации и инвестиции. 2020. №6. . 2020. С. 194–197.
4. Rizzoli G. и др. SynDrone -- Multi-modal UAV Dataset for Urban Scenarios [Электронный ресурс]. 2023. URL: https://arxiv.org/pdf/2308.10491 (дата обращения: 14.11.2025).
5. Qu G. и др. StripNet [Электронный ресурс] // Proceedings of the 26th ACM international conference on Multimedia. 2018. С. 283–291. URL: https://dl.acm.org/doi/10.1145/3240508.3240553 (дата обращения: 16.11.2025).
6. Lin T.-Y. и др. Feature Pyramid Networks for Object Detection [Электронный ресурс]. 2017. URL: https://arxiv.org/pdf/1612.03144v2 (дата обращения: 10.11.2025).
7. Berthelot D. и др. MixMatch: A Holistic Approach to Semi-Supervised Learning [Электронный ресурс]. 2019. URL: https://arxiv.org/pdf/1905.02249 (дата обращения: 18.11.2025).
8. Foi A. и др. Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data [Электронный ресурс] // IEEE Transactions on Image Processing. 2008. С. 1737–1754. URL: https://www.researchgate.net/publication/23249686_Practical_Poissonian-Gaussian_Noise_Modeling_and_Fitting_for_Single-Image_Raw-Data (дата обращения: 18.11.2025).
9. Miyato T. и др. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning [Электронный ресурс] // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019. Т. 41, № 8. С. 1979–1993. URL: https://ieeexplore.ieee.org/document/8417973 (дата обращения: 18.11.2025).
10. Catalano N., Matteucci M. Few Shot Semantic Segmentation: a review of methodologies, benchmarks, and open challenges [Электронный ресурс]. 2024. URL: https://arxiv.org/pdf/2304.05832v2 (дата обращения: 16.11.2025).
11. Shaban A. и др. One-Shot Learning for Semantic Segmentation [Электронный ресурс]. 2017. URL: https://arxiv.org/pdf/1709.03410 (дата обращения: 16.11.2025).
12. Beery S. и др. Synthetic Examples Improve Generalization for Rare Classes [Электронный ресурс]. 2019. URL: https://arxiv.org/pdf/1904.05916 (дата обращения: 16.11.2025).
13. Зуев, В. М. Сравнение обнаружения объектов средствами искусственного интеллекта в сравнении с классическими методами / В.М. Зуев [Электронный ресурс] // Проблемы искусственного интеллекта. – 2024. – Т. 34 (3). – С. 4-10. URL: http://paijournal.guiaidn.ru/download_pai/2024_3/3_%D0%97%D1%83%D0%B5%D0%B2.pdf (дата обращения: 17.11.2025).
14. Lyu Y. и др. UAVid: A Semantic Segmentation Dataset for UAV Imagery [Электронный ресурс]. 2020. URL: https://arxiv.org/pdf/1810.10438 (дата обращения: 16.11.2025).
15. Song A. Deep Learning-Based Semantic Segmentation of Urban Areas Using Heterogeneous Unmanned Aerial Vehicle Datasets [Электронный ресурс] // Aerospace. 2023. С. 880. URL: https://www.mdpi.com/2226-4310/10/10/880#:~:text=areas,be%20used%20on%20diverse%20data (дата обращения: 16.11.2025).
16. Zhang Z., Li G. UAV Imagery Real-Time Semantic Segmentation with Global–Local Information Attention [Электронный ресурс] // Sensors. 2025. С. 1786. URL: https://www.mdpi.com/1424-8220/25/6/1786#:~:text=semantic%20segmentation%20of%20drones%20that,branch%20compresses%20and%20extracts%20global (дата обращения: 16.11.2025).
17. Zhang Y., David P., Gong B. Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes [Электронный ресурс]. 2018. URL: https://arxiv.org/pdf/1707.09465 (дата обращения: 16.11.2025).
18. Choi J., Kim Y. Colorful Cutout: Enhancing Image Data Augmentation with Curriculum Learning [Электронный ресурс]. 2024. URL: https://arxiv.org/pdf/2403.20012 (дата обращения: 16.11.2025).
19. Huang S. и др. DEIM: DETR with Improved Matching for Fast Convergence [Электронный ресурс]. 2025. URL: https://www.themoonlight.io/en/review/deim-detr-with-improved-matching-for-fast-convergence (дата обращения: 16.11.2025).

20. Зуев, В. М. Подготовка данных для обучения нейронной сети, управляющей движением механизма / В.М. Зуев, О.А. Бутов, А.А. Никитина, С.И. Уланов [Электронный ресурс] // Материалы Донецкого международного круглого стола «Искусственный интеллект: теоретические аспекты и практическое применение. ИИ – 2021». – 2021. – С. 92-95. ГУ «ИПИИ». URL: http://paijournal.guiaidn.ru/download_pai/2021_2/2_%D0%97%D1%83%D0%B5%D0%B2_%D0%91%D1%83%D1%82%D0%BE%D0%B2_%D0%98%D0%B2%D0%B0%D0%BD%D0%BE%D0%B2%D0%B0_%D0%9D%D0%B8%D0%BA%D1%82%D0%B8%D0%BD%D0%B0_%D0%A3%D0%BB%D0%B0%D0%BD%D0%BE%D0%B2.pdf (дата обращения: 17.11.2025).
21. Пикалёв, Я. С. О нейронных архитектурах извлечения признаков для задачи распознавания объектов на устройствах с ограниченной вычислительной мощностью / Я.С. Пикалёв [Электронный ресурс] // Проблемы искусственного интеллекта. – 2023. – № 2413-7383. – С. 44-54. – DOI: 10.34757/2413-7383.2023.30.3.004. – ISSN: 2413-7383. URL: http://paijournal.guiaidn.ru/download_pai/2023_3/4_%D0%9F%D0%B8%D0%BA%D0%B0%D0%BB%D0%B5%D0%B2_%D0%95%D1%80%D0%BC%D0%BE%D0%BB%D0%B5%D0%BD%D0%BA%D0%BE.pdf (дата обращения: 17.11.2025).
22. Tsai Y.-H. и др. Learning to Adapt Structured Output Space for Semantic Segmentation [Электронный ресурс]. 2020. URL: https://arxiv.org/pdf/1802.10349 (дата обращения: 16.11.2025).
23. Hoffman J. и др. CyCADA: Cycle-Consistent Adversarial Domain Adaptation [Электронный ресурс]. 2017. URL: https://arxiv.org/pdf/1711.03213 (дата обращения: 16.11.2025).
24. Zhong Z. и др. Adversarial Style Augmentation for Domain Generalized Urban-Scene Segmentation [Электронный ресурс]. 2022. URL: https://arxiv.org/pdf/2207.04892 (дата обращения: 16.11.2025).
25. Li Y. и др. LSKNet: A Foundation Lightweight Backbone for Remote Sensing [Электронный ресурс]. 2025. URL: https://arxiv.org/pdf/2403.11735 (дата обращения: 16.11.2025).
26. Szántó M. и др. Building Maps Using Monocular Image-feeds from Windshield-mounted Cameras in a Simulator Environment [Электронный ресурс] // Periodica Polytechnica Civil Engineering. 2023. URL: https://doi.org/10.3311/PPci.21500 (дата обращения: 10.11.2025).

RESUME

V. Yu. Ustenko, Blizno M.V.

Data Augmentation Method for Segmenting Key Objects in UAV Aerial Imagery

Semantic segmentation of urban UAV scenes is complicated by the domain gap between real and synthetic SynDrone data and by the high computational demands of training. This study examines the combination of a pre-trained StripNet backbone and an adaptive FPN decoder with the authors' proposed AugUAV data augmentation methodology for semantic segmentation of five key classes: Road, Nature, Construction, Obstacle, and Water.

The methodology enables a staged training regime: the model is first exposed to mild transformations and then progressively to more complex object-level perturbations. This increases robustness to seasonal and illumination shifts. Experimental results show that AugUAV stabilizes early training epochs and achieves mIoU = 0.6264 and FWIoU = 0.9565 with moderate resource consumption, improving quantitative metrics compared to standard augmentation pipelines.

The findings support the underlying hypothesis: harmonizing augmentations effectively mitigates the synthetic-to-real gap and makes AugUAV suitable for training semantic segmentation systems targeting key objects in UAV aerial imagery.

РЕЗЮМЕ

В.Ю. Устенко, М.В. Близно

Методика расширения данных для сегментации ключевых объектов на аэрофотоснимках БПЛА

Семантическая сегментация городских сцен с БПЛА осложняется доменным разрывом между реальными и синтетическими данными SynDrone и высокими требованиями к вычислительным ресурсам. В работе исследуется сочетание предобученного StripNet-backbone и адаптивного FPN-декодера с предложенной авторами методики расширения данных AugUAV для задачи семантической сегментации (для пяти ключевых классов: Road, Nature, Construction, Obstacle, Water). Данная методика позволяет сначала мягко обучить модель, затем постепенно вводить сложные объектные трансформации, что повышает устойчивость к сезонным и световым сдвигам. Эксперименты показывают, что AugUAV стабилизирует ранние эпохи и обеспечивает mIoU=0.6264, FWIoU=0.9565 при умеренном потреблении ресурсов, демонстрируя улучшение количественных показателей по сравнению со стандартной методикой расширения данных. Результаты подтверждают гипотезу: гармонизация аугментаций действительно помогает сгладить синтетико-реальный разрыв и делает AugUAV пригодной для использования при обучении систем семантической сегментации ключевых объектов на аэрофототснимках.

Устенко Владимир Юрьевич – младший научный сотрудник, лаборатория интеллектуальных систем и анализа данных, аспирант, Федеральное государственное бюджетное научное учреждение «Институт проблем искусственного интеллекта».
Область научных интересов: системы распознавания образов, эл. почта ustencko.vova@ya.ru, адрес: 283048, Российская Федерация, г. Донецк, ул. Артема, д. 118 б.

Близно Максим Витальевич – младший научный сотрудник, лаборатория интеллектуальных систем и анализа данных, Федеральное государственное бюджетное научное учреждение «Институт проблем искусственного интеллекта».
Область научных интересов: системы распознавания образов,
адрес: 283048, Российская Федерация, г. Донецк, ул. Артема, д. 118 б.

Статья поступила в редакцию 23.10.2025.