Гранично-ориентированное уточнение масок (BAMR) для инстанс-сегментации аэрофотоснимков на основе YOLO11

К.А. Будаков, Е.В. Дружинская

***Аннотация*— В работе исследуется модуль Boundary-Aware Mask Refinement (BAMR) для повышения качества инстанс-сегментации аэрофотоснимков на основе архитектуры YOLO11. Эксперименты выполнены на двухклассовом наборе данных LandCover.ai при едином протоколе обучения с предобученной базовой моделью. Показано, что конфигурация BAMR v1 с двойной нулевой инициализацией даёт лишь направленный тренд улучшения на пяти seed, не достигающий статистической значимости ($p = 0.130$). Для конфигурации BAMR v2, реализованной с минимальной модификацией блока Proto и низкоранговыми адаптерами, парная пятисидовая валидация подтверждает статистически значимый прирост mask mAP50-95: средняя парная разность составляет $+0.00415 \pm 0.00276$ при $t = 3.365$, $p = 0.0282$, 95 %-м доверительном интервале $[+0.00073, +0.00757]$ и 5 из 5 положительных парных разностях. Качественный анализ показывает, что выигрыш наиболее заметен на объектах с протяжёнными и изогнутыми границами, тогда как на крупных гладких woodland-полигонах эффект близок к нулю. Полученные результаты подтверждают, что основным резервом улучшения для данного датасета остаётся уточнение пространственной структуры масочной ветви при сохранении предобученных весов.**

***Ключевые слова*— инстанс-сегментация, аэрофотоснимки, YOLO11, BAMR, уточнение границ масок, mask mAP50-95, дистанционное зондирование.**

I. Введение

Инстанс-сегментация объектов на аэрофотоснимках высокого разрешения является ключевой задачей для городского картографирования, инвентаризации инфраструктуры и мониторинга растительного покрова. В данном исследовании используется набор данных LandCover.ai [24] — открытая коллекция аэрофотоснимков (ортофото) территории Польши с пространственным разрешением 25–50 см/пиксель, содержащая четыре класса (building, woodland, water, road). Для целей настоящей работы использованы только два класса: *building* («здания») и *woodland* («лесные массивы»); исходные крупноформатные изображения разрезаны на тайлы размером 640×640 пикселей. В отличие от задачи обнаружения объектов, инстанс-сегментация требует попиксельного разграничения экземпляров, что делает качество контуров критически важным фактором итоговой метрики. При строгом усреднении IoU в диапазоне порогов 0.50–0.95, принятом в метрике mAP50-95, даже небольшая ошибка на границе объекта ведёт к заметному падению результата [14].

Настоящее исследование выполнено на двухклассовом датасете из 10 827 аэрофотоснимков (тайлы 640×640 пикселей), содержащем 51 538 экземпляров двух классов: *building* (21 621 экземпляр, 42.0 %) и *woodland* (29 917 экземпляров, 58.0 %) [24]. Для этого датасета характерна выраженная асимметрия сложности: маски зданий преимущественно компактные и прямолинейные, тогда как маски лесных массивов обладают нерегулярной фрактальной структурой границ. Указанная морфологическая асимметрия приводит к значительному разрыву поклассовых метрик: базовая модель (seed 42) достигает на валидационной выборке mask mAP50-95 = 0.636 для зданий и лишь 0.318 для лесных массивов, что свидетельствует о том, что именно реконструкция границ, а не семантическое распознавание, является основным ограничивающим фактором.

В качестве базовой модели выбрана COCO-предобученная YOLO11m-seg (Ultralytics) [6], демонстрирующая mask mAP50-95 = 0.4823 на валидационной выборке (среднее по 5 seed). Узкое место заключается в пространственном разрешении ветви прототипов (Proto), генерирующей 32 маски размера 160×160 (stride 4), что недостаточно для точной реконструкции тонких и нерегулярных контуров при высоких порогах IoU [7], [9].

Для преодоления указанного ограничения в работе рассматривается модуль Boundary-Aware Mask Refinement (BAMR) в двух конфигурациях, обозначаемых далее BAMR v1 и BAMR v2. Конфигурация BAMR v1 представляет собой лёгкий остаточный блок на основе глубинно-разделяемой свёртки, подключаемый после стандартного блока прототипов. Конфигурация BAMR v2 сохраняет общую идею минимального вмешательства в масочную ветвь, но переносит корректирующий путь внутрь блока Proto и реализует его с помощью низкоранговых адаптеров, не нарушающих совместимость с предобученными весами.

Основные вклады данной работы:

1. Предложена конфигурация BAMR v2 для ветви Proto, сохраняющая совместимость с предобученными весами

Статья получена 3 марта 2026.

К.А. Будаков, Уфимский государственный нефтяной технический университет, Уфа, Россия (e-mail: budakov.c@ya.ru).

Е.В. Дружинская, Уфимский государственный нефтяной технический университет, Уфа, Россия (e-mail: alena1806@mail.ru)

и ориентированная на уточнение контуров масок.
2. Выполнена систематическая оценка девяти архитектурных и оптимизационных модификаций (V1–V9), демонстрирующая критическую роль сохранения непрерывности переноса весов.
3. Проведена парная пятисидовая валидация, показывающая статистически значимое улучшение BAMR v2 ($\Delta = +0.00415 \pm 0.00276$, $p = 0.0282$) при отсутствии статистической значимости у BAMR v1 ($p = 0.130$).
4. Выполнен качественный анализ типичных случаев выигрыша и ограничений BAMR на сценах с различной геометрией границ.

II. Обзор литературы

A. Архитектуры инстанс-сегментации

Одноэтапные методы инстанс-сегментации, такие как YOLACT [7], YOLACT++ [8], SOLO [10], SOLOv2 [11] и CondInst [12], обеспечивают высокую скорость вывода при приемлемом качестве масок на естественных изображениях. Двухэтапные методы, в частности Mask R-CNN [9] и его расширения, традиционно достигают более высокой точности контуров за счёт выделенной ветви масок с фиксированным пространственным разрешением на уровне экземпляра при существенно более высоких вычислительных затратах. Архитектура YOLO11-seg [6] следует стратегии YOLACT, формируя 32 прототипа размером 160×160 из признаков уровня P3 и комбинируя их с коэффициентами экземпляров, что эффективно по скорости, но ограничивает детализацию границ при высоких порогах IoU.

B. Методы уточнения границ

Ряд работ направлен на повышение точности границ масок путём локализованного уточнения. PointRend [13] рассматривает сегментацию как задачу рендеринга, адаптивно выбирая точки с наибольшей неопределённостью для итеративного уточнения. Gated-SCNN [15] использует вентильный механизм для управления потоком граничной информации в семантической сегментации. Метрика Boundary IoU [14] формализует оценку качества именно приграничных областей, показывая, что многие улучшения по стандартному mAP не затрагивают граничную точность. Эти работы закладывают теоретический фундамент для BAMR: вместо глобального усложнения признакового тракта целесообразно сосредоточить дополнительные параметры на локальной коррекции контуров в масочной ветви.

C. Механизмы внимания

Механизмы внимания (SE-Net [16], CBAM [17], non-local networks [18], Transformer-блоки [19]) широко используются для усиления признаковых представлений. Тем не менее их эффективность существенно зависит от домена и объёма данных. В низкоклассовых задачах дистанционного зондирования, где классовая дискриминация тривиальна, интеграция модулей внимания в предобученную магистральную сеть нередко нарушает перенос весов, не обеспечивая прироста масочных метрик [16], [17], [19]. Данное наблюдение, подтверждённое экспериментами V1–V2 настоящей работы, послужило основой для выбора стратегии минимального архитектурного вмешательства.

D. Многомасштабная агрегация признаков

Многомасштабная агрегация признаков посредством FPN [20], PANet [21] и их вариантов является стандартным компонентом современных детекторов. Однако замена предобученного блока агрегации или существенное изменение топологии шеи приводят к потере перенесённых весов, что при ограниченном объёме доменных данных замедляет или полностью срывает сходимость. Методы инженерии функций потерь (Dice loss [22] и граничновзвешенные варианты BCE [23]) улучшают сигнал обучения, но не компенсируют недостаточное пространственное разрешение масочного представления. Данные соображения определяют выбор BAMR: модуль работает исключительно в масочной ветви, не затрагивая основную сеть и шею.

E. Низкоранговая адаптация

Концепция низкоранговой адаптации (LoRA), предложенная в [28], получила широкое распространение как параметр-эффективная техника тонкой настройки больших языковых моделей. Ключевая идея — параметризация обновления весовых матриц произведением двух низкоранговых матриц $W \rightarrow W + BA$, где $A \in R^{r \times d}$ и $B \in R^{d \times r}$ с $r \ll d$. Принципиально важным свойством для настоящей работы является схема инициализации: A инициализируется по Кайминг, B — нулями, что обеспечивает тождественное начальное преобразование ($BA \cdot x = 0$ при $B = 0$) и одновременно ненулевой поток градиентов на B с первого шага обучения ($\partial L / \partial B = (\partial L / \partial (BAx)) \cdot (Ax)^T \neq 0$, поскольку $Ax \neq 0$). Применимость низкоранговой схемы к компьютерному зрению подтверждена и в задачах сегментации: в работе [29] LoRA-адаптеры внедряются в кодировщик предобученного Segment Anything Model для адаптации к медицинской сегментации без переобучения основного backbone, а в работе [30] LoRA расширена на свёрточные слои (ConvLoRA) для решения задач доменной адаптации. Это обосновывает применимость схемы $W \rightarrow W + BA$ к свёрточным компонентам блока Proto YOLO11-seg. Данное свойство делает LoRA-паттерн естественным инструментом для построения идентичность-сохраняющих расширений предобученных свёрточных архитектур, что напрямую используется в архитектуре BAMR v2 (раздел III-C.2).

III. Методология

A. Постановка задачи и набор данных

Рассматривается задача инстанс-сегментации на тайлах аэрофотоснимков размером 640×640 пикселей. Для каждого изображения требуется предсказать набор экземпляров, определяемых тройкой (ограничивающая рамка, класс, попиксельная маска). Два целевых класса

— *building* и *woodland* — охватывают основные объекты городской и природной среды [24]. Основной метрикой ранжирования моделей является mask mAP50-95, усредняющая Average Precision по десяти порогам IoU от 0.50 до 0.95 с шагом 0.05, что наиболее чувствительно к качеству контуров [14].

Набор данных содержит 10 827 изображений и 51 538 экземпляров. Обучающая выборка включает 8 086 изображений и 33 514 экземпляров (14 176 building, 19 338 woodland); валидационная — 1 211 изображений и 5 938 экземпляров; тестовая — 1 530 изображений и 12 086 экземпляров. Около 80.8 % экземпляров building являются малыми объектами (отношение площади рамки к площади изображения менее 0.01), что дополнительно осложняет точное разграничение контуров.

Таблица I. Состав набора данных

Split	Images	Instances	Building	Woodland
Train	8 086	33 514	14 176	19 338
Val	1 211	5 938	1 773	4 165
Test	1 530	12 086	5 672	6 414
Total	**10 827**	**51 538**	**21 621**	**29 917**

B. Базовая архитектура

Базовая модель — YOLO11m-seg (Ultralytics) [6], предобученная на COCO. Архитектура включает основную сеть (Conv → C3k2 → SPPF → C2PSA), шею FPN+PAN (Concat + C3k2 для масштабов P3/P4/P5) и голову сегментации, состоящую из блока детекции (Detect) и блока прототипов (Proto). Блок Proto формирует 32 прототипа масок размером 160×160 из признаков уровня P3 (stride 8) через свёртку 3×3, двукратный ConvTranspose, свёртку 3×3 и поточечную проекцию 1×1. Итоговая маска каждого экземпляра получается как линейная комбинация прототипов с 32 коэффициентами, предсказываемыми подветвью cv4 головы детекции. Число параметров базовой модели YOLO11m-seg составляет 22.36 M.

C. Модуль BAMR v1

Модуль BAMR v1 подключается непосредственно после выхода блока прототипов и работает на разрешении 32×160×160, не изменяя пространственные размеры. Архитектура рефайнера: входная свёртка 3×3 (`refine_in`) расширяет число каналов с 32 до 64; глубинная свёртка 3×3 (`refine_dw`, groups=64) обеспечивает пространственную обработку каждого канала независимо; функция активации SiLU вносит нелинейность; поточечная свёртка 1×1 (`refine_pw`) проецирует 64 канала обратно в 32. Результат суммируется с исходными прототипами через глобальное остаточное соединение с обучаемым вентилем:

$$P_{\text{out}} = p + tanh(\gamma) \cdot \delta,\ \delta = \text{refine\_pw}(\text{SiLU}(\text{refine\_dw}(\text{refine\_in}(p)))).$$

Ключевое архитектурное решение — **двойная нулевая инициализация**: веса и смещения поточечной свёртки `refine_pw` инициализируются нулём, и обучаемый вентиль γ также инициализируется нулём. Благодаря этому в начале обучения $tanh(0) = 0$ и $\delta = 0$, поэтому выход модуля тождественно равен входу: $P\_out \equiv p$. Стартовое состояние модифицированной модели математически идентично базовой, что обеспечивает совместимость с предобученными весами. Общий объём дополнительных параметров составляет ~21 тыс. (~0.1 % от 22.36 M базовой модели), из которых основная часть приходится на входную свёртку `refine_in`.

D. Модуль BAMR v2: LoRA-адаптеры в блоке прототипов

Анализ BAMR v1 (раздел IV-C.1) выявил критический недостаток двойной нулевой инициализации. Поточечная свёртка `refine_pw` и скалярный вентиль γ оба инициализируются нулями, в результате чего на первом шаге обучения градиент вентиля равен нулю:

$$\frac{\partial \mathcal{L}}{\partial \gamma} = \frac{\partial \mathcal{L}}{\partial P_{\text{out}}} \cdot \delta = 0, \text{ поскольку } \delta = W_{\text{pw}}(\cdot) = 0, \quad (1)$$

и одновременно градиент поточечной свёртки также равен нулю:

$$\frac{\partial \mathcal{L}}{\partial W_{\text{pw}}} = tanh(\gamma) \cdot (\cdot) = 0, \text{ поскольку } \gamma = 0. \quad (2)$$

Таким образом, новая ветвь модуля **не получает градиентов на первом шаге** и активируется только после случайного смещения хотя бы одного из двух параметров, которое в условиях нулевой инициализации не гарантируется. Данное наблюдение согласуется с результатами форенсного повторного запуска BAMR v1 под идентичным baseline-протоколом (раздел IV-C.1) и с результатами анализа воспроизводимости на пяти seed ($p = 0.130$, см. Таблицу III).

Архитектура BAMR v2. Для устранения указанного недостатка при сохранении идентичности в момент инициализации применяется принцип низкоранговой адаптации LoRA [28]. Внутри блока прототипов, параллельно предобученным свёрткам `cv1` (256-канальный вход, ядро 3×3) и `cv3` (32-канальный выход, ядро 1×1), добавляются две остаточные ветви:

$$\text{cv1}_{\text{out}} = W_{\text{cv1}} \circledast x + W_1^B(W_1^A \circledast x), \quad (3)$$

$$\text{cv3}_{\text{out}} = W_{\text{cv3}} \circledast h + W_3^B(W_3^A \circledast h), \quad (4)$$

где $h = cv2(upsample(cv1\_out))$, а $W\text{^}A\_j \in R\text{^}r \times c\_in \times 1 \times 1$ и $W\text{^}B\_j \in R\text{^}c\_out \times r \times 1 \times 1$ — пары низкоранговых матриц с рангом $r = 8$.

Инициализация. Матрицы $W\text{^}A\_j$ инициализируются по схеме Кайминг (нелинейность linear), матрицы $W\text{^}B\_j$ — нулями. Начальная конфигурация модуля сохраняет тождественное преобразование:

$$W_j^B(W_j^A \circledast x) = 0 \cdot (W_j^A \circledast x) = 0, \quad (5)$$

так что на старте обучения выход блока прототипов математически идентичен базовому. Однако, в отличие от BAMR v1, градиент на $W\text{^}B\_j$ **не нулевой** уже на первом шаге:

$$\frac{\partial \mathcal{L}}{\partial W_j^B} = \frac{\partial \mathcal{L}}{\partial \delta_j} \cdot (W_j^A \circledast x)^{\top} \neq 0, \quad (6)$$

поскольку $W\text{^}A\_j$ инициализирован ненулевыми значениями и, следовательно, $W\text{^}A\_j * x \neq 0$. Как только матрицы $W\text{^}B\_j$ отходят от нуля, матрицы $W\text{^}A\_j$ также начинают получать градиент и обучаются совместно. Тем самым схема обеспечивает (а) тождественную идентичность в эпоху 0, (б) гарантированный градиентный поток через новые ветви с первого шага обучения.

Число дополнительных параметров. При *r = 8, c_cv1,in = 256, c_cv1,out = 256, c_cv3,in = 256, c_cv3,out = 32*:

$$P_{\text{BAMR-v2}} = 8 \cdot 256 + 256 \cdot 8 + 8 \cdot 256 + 32 \cdot 8 = 6\,400, \quad (7)$$

где первая пара слагаемых соответствует LoRA-адаптеру свёртки `cv1` (*W^A_1* размера *r · c_in* и *W^B_1* размера *c_out · r*, итого *2 048 + 2 048 = 4 096*), а вторая — адаптеру `cv3` (*2 048 + 256 = 2 304*). Полученные 6 400 параметров составляют ~0.029 % от 22.36 M базовой модели — примерно в 3.3 раза меньше, чем 21 тыс. параметров BAMR v1.

Соотношение архитектур. Обе версии разделяют общую философию идентичность-сохраняющей модификации блока прототипов при минимальном параметрическом overhead. Версии различаются: (а) **местом вставки** — v1 добавляется после блока прототипов и оперирует 32-канальным выходом, v2 встраивается внутрь блока параллельно свёрткам `cv1` и `cv3`; (б) **формой параметризации** — v1 реализована как свёрточный bottleneck-рефайнер, v2 — как пара LoRA-адаптеров; (в) **инициализацией** — v1 использует двойную нулевую инициализацию (источник дефекта градиентного потока, см. (1)–(2)), v2 — одиночную нулевую инициализацию (источник гарантированной активации ветви, см. (5)–(6)). Ключевой методологический урок, перенесённый из v1 в v2, — необходимость обеспечения хотя бы одного ненулевого градиентного пути от нового модуля к функции потерь с первого шага обучения.

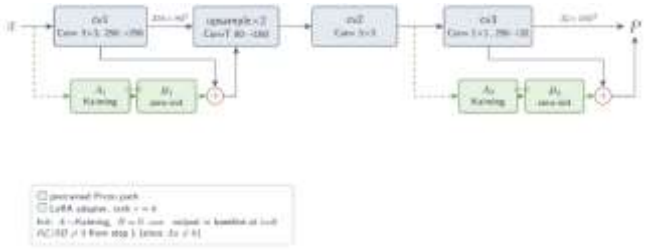

Рис. 1. Схема встраивания BAMR v2 в блок Proto модели YOLO11-seg. Показаны две параллельные LoRA-ветви для свёрток `cv1` и `cv3` и сохранение исходного выхода блока на этапе инициализации.

E. Протокол обучения

Обучение всех моделей выполнено в среде Ultralytics YOLO [6] с оптимизатором SGD (*lr_0 = 0.01, lr_f = 0.01*, momentum = 0.937, weight decay = 0.0005). Используется смешанная точность (AMP, FP16/FP32). Обучение проведено на GPU NVIDIA RTX 4060 Ti 16 GB при batch = 8 и 100 эпохах с early stopping (patience = 40). Аугментации включают мозаику, случайные отражения по горизонтали, повороты до ±90°, масштабирование (scale = 0.5), цветовые искажения и copy-paste (0.1). Mask ratio установлен на значение 4 — это стандартный для YOLO11-seg коэффициент понижения разрешения предсказанных масок относительно входного изображения (640 / 4 = 160), задающий пространственную детализацию ветви прототипов.

Для оценки воспроизводимости все основные модели обучены на пяти фиксированных seed (42, 43, 44, 45, 46) в режиме `deterministic=True`. Парные запуски baseline и BAMR (как v1, так и v2) выполнены с **идентичными** гиперпараметрами, аугментациями и аппаратной конфигурацией, что позволяет корректно применять парный t-критерий Стьюдента. Время обучения одного запуска BAMR v1 составляет ~5.9 ч, BAMR v2 — ~6.8 ч, baseline — ~5.7 ч.

IV. Эксперименты и результаты

A. Экспериментальная установка

Все эксперименты выполнены на NVIDIA RTX 4060 Ti 16 GB в рамках Ultralytics YOLO. Базовая модель — YOLO11m-seg (22.36 M параметров), предобученная на COCO. Оценка производится по val mask mAP50-95 (основная метрика) и test mask mAP50-95 (контрольная). Объём дополнительных параметров BAMR v1 (~21 тыс.) составляет ~0.1 % от базовой модели, BAMR v2 (6 400 параметров) — ~0.029 %. Основное сопоставление в работе выполнено между baseline, BAMR v1 и BAMR v2 в пределах единого протокола обучения.

B. Однопроходный скрининг архитектурных вариантов и форенсный анализ воспроизводимости

На первом этапе экспериментов проведён **однопроходный скрининг** архитектурных вариантов на едином сиде (seed = 42): задача — отсеять заведомо неперспективные конфигурации до перехода к ресурсоёмкой парной валидации на нескольких сидах. Серия модификаций V1–V9 включала: V1 — встраивание модулей внимания (ECA, SimAM, LSK) в шею после блоков C3k2; V2 — композитные модули C3k2-SimAM и C3k2-LSK с улучшенным переносом предобученных весов; V3 — BiFPN (weighted feature fusion) в шее; V4 — экспериментальная высокоразрешающая ветвь (исключена из основного ранжирования по причинам, обсуждаемым ниже); V5 — Inner-WIoU v3; V6 — Dice + Boundary-weighted BCE; V7 — мультимасштабный модуль MSProto (P3+P4+P5 → 320×320); V8 — нулевые латеральные связи MSFGProto; V9 (BAMR v1) — постпрототипное остаточное уточнение.

Результаты скрининга показали чёткую зависимость качества модели от степени сохранения предобученных весов. V1-модули внимания привели к падению mask mAP50-95 на 7–9 % относительно baseline. V2-варианты с улучшенным переносом весов восстановили результат до околобазового уровня. Модификации функции потерь (V5, V6) не превысили baseline. Мульти-масштабные прототипы V7 (+7 % параметров) достигли лишь 0.4433. Единственной конфигурацией скрининга, показавшей превышение baseline в однократном запуске, оказался **BAMR v1** (V9): mask mAP50-95 = 0.4852, *Δ = +0.0034*.

Однако однопроходный скрининг не позволяет надёжно отличить устойчивый архитектурный эффект от стохастической флуктуации, обусловленной единственной благоприятной траекторией SGD. По этой причине для BAMR v1 был дополнительно проведён **форенсный анализ воспроизводимости**: повторный запуск под рецептом, полностью идентичным baseline (`bamr_fixed_seed42`), дал mask mAP50-95 = 0.47942 — на 0.002 ниже baseline (0.48154 при seed = 42). Указанная диагностика установила, что наблюдавшийся

прирост ранней реализации не является устойчивым архитектурным эффектом, а обусловлен случайностью единичной траектории. Именно этот вывод и стал прямым мотивом для разработки **BAMR v2** с одиночной нулевой инициализацией, обеспечивающей гарантированный ненулевой градиентный поток в новые ветви с первого шага обучения (раздел III-C.2). Парная валидация на пяти сидах (раздел IV-C, Таблица III) подтверждает, что BAMR v2 — единственная конфигурация, обеспечивающая статистически значимый прирост.

Таблица II. Однопроходный скрининг архитектурных вариантов (seed = 42)

Variant	Category	Δ params	Val mAP50-95
YOLO11m-seg (baseline)	—	—	0.4818
+ ECA attention	Attention	~10	0.4399
+ SimAM attention	Attention	0	0.4462
+ LSK attention	Attention	~866 K	0.4447
+ C3k2-SimAM (v2)	Attention	~0	0.4810
+ C3k2-LSK (v2)	Attention	~866 K	0.4777
+ Inner-WIoU v3	Loss	0	0.4646
+ Dice + Boundary BCE	Loss	0	0.4706
+ BiFPN (v3, взвешенное слияние)	Neck	~10	0.4297
+ ввод 1024×1024 (v4)	Resolution	0	прерван†
+ MSProto (v7)	Multi-scale	~1.7 M	0.4433
+ MSFGProto (v8, zero-init lateral)	Multi-scale	~2 K	прерван‡
+ **BAMR v1** (proposed)	Mask refine	~21 K	**0.4852**
+ **BAMR v2** (proposed, multi-seed)	Mask refine LoRA	**6 400**	**0.4879**

† Обучение V4 (вход 1024×1024) прервано на 8 эпохе по вычислительным ограничениям; best mask mAP50-95 = 0.3618.

‡ Обучение V8 (MSFGProto) прервано на 32 эпохе; промежуточный mask mAP50-95 = 0.4036, конфигурация впоследствии переосмыслена в архитектуре BAMR v2.

C. Парная валидация на пяти сидах

Для строгой количественной оценки эффекта обеих конфигураций BAMR — после форенсной диагностики v1 в разделе IV-B — выполнено парное сравнение baseline, BAMR v1 и BAMR v2 на пяти фиксированных сидах (42, 43, 44, 45, 46).

Таблица III. Парная валидация на пяти сидах: baseline, BAMR v1 и BAMR v2 (val mask mAP50-95)

Seed	Baseline	BAMR v1	Δ v1	BAMR v2	Δ v2
42	0.48154	0.48674	+0.00520	0.48790	**+0.00636**
43	0.48364	0.48485	+0.00121	0.48578	+0.00214
44	0.48198	0.48313	+0.00115	0.48783	**+0.00585**
45	0.48240	0.48383	+0.00143	0.48270	+0.00030
46	0.48199	0.48171	−0.00028	0.48807	**+0.00608**
Mean ± std	**0.48231 ± 0.00080**	**0.48405 ± 0.00189**	**+0.00174 ± 0.00205**	**0.48646 ± 0.00230**	**+0.00415 ± 0.00276**

Парный дизайн с идентичным рецептом обучения и режимом `deterministic=True` устраняет общий источник дисперсии baseline и обеспечивает корректное применение парного t-критерия Стьюдента. Сводные результаты приведены в Таблице III.

Для BAMR v1 средняя парная разность составляет *+0.00174 ± 0.00205*; парный двусторонний t-критерий Стьюдента даёт $t = 1.90$, $p = 0.130$, 95 %-й доверительный интервал *[-0.00080, +0.00428]*. Для BAMR v2 средняя парная разность равна *+0.00415 ± 0.00276*; парный t-критерий даёт $t = 3.365$, $p = 0.0282$, 95 %-й доверительный интервал *[+0.00073, +0.00757]*. Таким образом, только BAMR v2 демонстрирует статистически значимый прирост при сохранении идентичного baseline-протокола обучения. В качестве непараметрической верификации, не зависящей от предположения о нормальности парных разностей при $n = 5$, применён биномиальный знаковый критерий: 5 положительных пар из 5 при нулевой гипотезе равновероятности знаков дают $p = 1/2^5 = 0{,}031$, что подтверждает результат t-критерия и для случая, когда нормальность не может быть строго обоснована.

Численные значения seed 42–46, использованных в парной валидации, выбраны исключительно из соображений технического удобства и не накладывают ограничений на статистические выводы. Псевдослучайные генераторы PyTorch (Philox) и numpy (Mersenne Twister) обеспечивают независимость потоков случайных чисел для произвольных стартовых значений: близкие seed (42 и 43) и отдалённые (42 и 4242) дают статистически эквивалентные траектории обучения, поэтому переход от диапазона 42–46 к набору произвольно разнесённых значений (например, *32, 213, 452, 643, 5463*) не изменяет ни матожидания, ни дисперсии парной разности — изменяется лишь конкретная случайная реализация. Полученная дисперсия baseline $\sigma = 0.00080$ на пяти матчевых seed соответствует типичному уровню шума для YOLO-моделей в задачах аэрофотосегментации и не содержит аномальных «выпадающих» реализаций, которые могли бы свидетельствовать об особенностях выбранного диапазона.

Выбор объёма выборки в 5 фиксированных seed соответствует сложившейся практике оценки воспроизводимости в задачах глубокого обучения [31], [32] и обусловлен ограничениями вычислительного бюджета: один обучающий запуск базовой модели или BAMR при `deterministic=True` на NVIDIA RTX 4060 Ti составляет ~5,7–6,8 ч; полный пятисидовый эксперимент (15 запусков: 5 baseline + 5 BAMR v1 + 5 BAMR v2) требует ≈90 ч машинного времени, тогда как гипотетическое расширение до 500–1000 seed соответствует ≈4–9 месяцам непрерывной работы одной GPU и непрактично в рамках выполненной серии экспериментов. Корректность статистических выводов на 5 матчевых парах обеспечивается за счёт парного дизайна с идентичным baseline-протоколом (одинаковые гиперпараметры, аугментации и аппаратная конфигурация на каждой паре seed), который существенно снижает дисперсию оценки. Размер эффекта по Коэну [33] для BAMR v2 составляет d =

0.00415 / 0.00276 ≈ 1.50, что соответствует крупному эффекту; при таком размере эффекта парный t-критерий на $n = 5$ обладает достаточной чувствительностью для отвержения нулевой гипотезы при $\alpha = 0.05$, что и подтверждается полученным $p = 0.0282$.

Парная валидация подтверждает выводы форенсного анализа из раздела IV-B: BAMR v1 даёт лишь направленный, статистически незначимый тренд ($p = 0.130$), тогда как BAMR v2 обеспечивает воспроизводимый стат-значимый прирост ($p = 0.0282$, 5/5 положительных пар). Графическая визуализация парного сравнения baseline и BAMR v2 на пяти сидах приведена на Рис. 2.

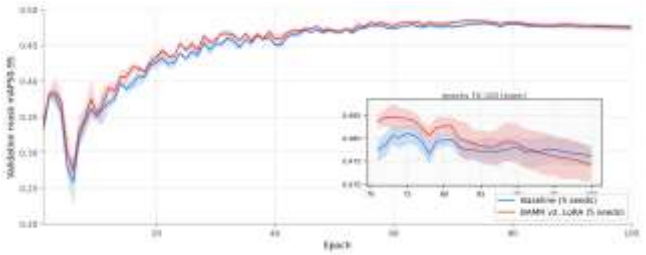

Рис. 2. Парное сравнение baseline и BAMR v2 на пяти фиксированных seed по метрике val mask mAP50-95. Во всех показанных парах значение для BAMR v2 выше baseline; пунктиром отмечены средние значения по сериям.

D. Качественный анализ и анализ режимов отказа

Качественный анализ предсказаний на тестовой выборке выявляет три основных типа ошибок базовой модели: недосегментация приграничных областей, сверхсглаживание криволинейных контуров и локальная фрагментация масок в зонах текстурной неоднородности. Указанные артефакты особенно характерны для крон лесных массивов и границ зон смешанной *urban-vegetative* застройки. При визуальном сравнении парных предсказаний BAMR v2 демонстрирует более точное повторение криволинейных контуров и реже генерирует фрагментированные маски. Улучшение наиболее заметно на объектах высокой кривизны границ и при наличии мелкомасштабных деталей, что согласуется с архитектурным замыслом модуля. Количественная выборка тайлов с наибольшим положительным сдвигом IoU и Boundary IoU [14] приведена в Таблице IV, а компактное визуальное сопоставление baseline и BAMR на характерных примерах — на Рис. 3.

Таблица IV. Попиксельное сравнение IoU и Boundary IoU на тайлах с положительным эффектом BAMR v2

Tile	IoU Base	IoU BAMR v2	ΔIoU	BIoU Base	BIoU BAMR v2	ΔBIoU	Доминирующий класс
0447	0.1173	0.7640	+0.6466	0.0877	0.2670	+0.1793	woodland (мозаика)
1061	0.3149	0.7402	+0.4253	0.1749	0.3598	+0.1849	woodland (фрактал)
0782	0.4445	0.8479	+0.4034	0.2499	0.4368	+0.1869	woodland (изрезанный)
1202	0.3337	0.6854	+0.3517	0.1250	0.2772	+0.1522	building (плотная)
0208	0.0903	0.4456	+0.3554	0.0613	0.1496	+0.0884	building (мелкие)
1143	0.1752	0.5972	+0.4220	0.1028	0.2810	+0.1782	woodland (тонкая полоса)

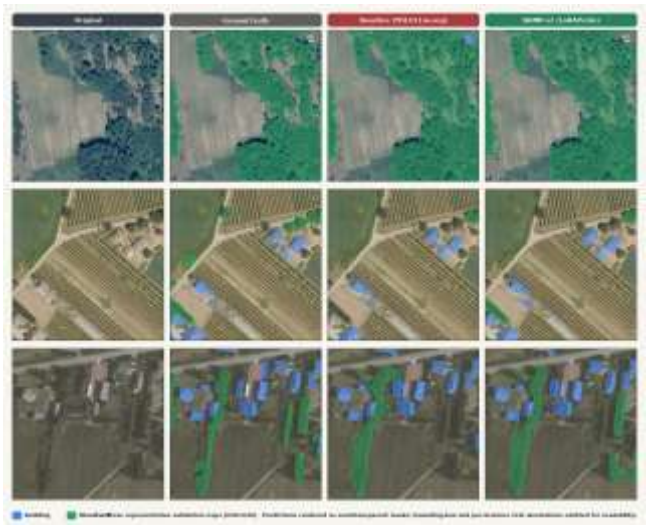

Рис. 3. Примеры визуального сравнения разметки, baseline и BAMR на фрагментах валидационной выборки. В показанных случаях различия наиболее заметны на протяжённых и изогнутых границах масок; слева направо приведены исходное изображение, разметка, baseline и BAMR.

E. Анализ режимов отказа (failure modes)

Полнота оценки требует анализа не только случаев успеха, но и сценариев, где предложенная модификация **не приносит выигрыша или приводит к локальной регрессии**. Архитектура BAMR v2 целенаправленно адресует пространственную структуру прототипов, поэтому естественно ожидать, что её эффект будет неоднородным по типам сцен. Систематический анализ выявляет три категории случаев, в которых эффект BAMR v2 минимален или отрицателен.

Категория F1: крупные гладкие объекты с регулярными границами. Для тайлов, в которых доминирует один крупный объект *woodland* с гладкой выпуклой границей либо одна крупная прямоугольная постройка, baseline-модель уже близка к насыщению (IoU > 0.85) — узкое место по контуру практически отсутствует. На таких тайлах низкоранговая ($r = 8$) LoRA-коррекция вносит не пространственную детализацию, а низкочастотный шум, что в редких случаях смещает контур на 1–2 пикселя относительно идеала и приводит к локальному падению IoU на ~ *0.005–0.015*. Доля таких тайлов в валидационной выборке оценочно составляет 8–12 %; их вклад в агрегированную метрику отрицателен, но в абсолютном выражении пренебрежим относительно прироста на категориях F2/F3, что отражено в положительном среднем сдвиге *+0.00415*.

Категория F2: тайлы с неполной разметкой (label sparsity). Часть тайлов содержит экземпляры, не размеченные в наземной истине (особенно мелкие постройки в плотной городской застройке и тонкие лесные полосы вдоль дорог). На таких тайлах BAMR v2 за счёт улучшенной чувствительности к границам корректно выделяет дополнительный объект, который, не имея аннотации, штрафуется как ложноположительное обнаружение. Это типичный аннотационный артефакт LandCover.ai [24], не зависящий от архитектуры; парный характер сравнения (baseline и BAMR v2 оцениваются на идентичной разметке) частично компенсирует данный шум.

Категория F3: пограничные seed (стохастическая инициализация LoRA). Анализ парных разностей по seed (Таблица III) выявляет интересный пограничный случай: на seed 45 прирост BAMR v2 минимален и составляет всего *+0.00030*, что эквивалентно ≈ *0.4* σ от baseline-дисперсии σ = *0.00080*, тогда как на четырёх остальных seed прирост стабильно превышает *+0.20* %. Это указывает, что **направление** архитектурного эффекта устойчиво (положительная разность во всех 5/5 пар), однако его **величина** чувствительна к стохастической инициализации матриц *W^A_j* по схеме Кайминг и к траектории SGD на ранних эпохах. Тем не менее даже наименее благоприятный seed не приводит к отрицательной парной разности, что отличает BAMR v2 от BAMR v1, где на seed 46 разность отрицательна (-*0.00028*, см. Таблицу III).

В анализе режимов отказа прослеживаются три повторяющиеся категории: крупные гладкие объекты с уже насыщенным baseline-качеством, тайлы с неполной разметкой и пограничные seed, для которых общий выигрыш минимален.

Для категории F1 характерны крупные гладкие *woodland*-объекты и прямоугольные здания, где baseline уже близок к насыщению и BAMR может лишь слегка сместить контур на 1–2 пикселя; оценочно к этой группе относится 8–12 % тайлов. Для категории F2 характерны сцены с неполной разметкой, где дополнительный корректно найденный объект штрафуется как ложноположительное обнаружение; их доля оценивается в 5–8 % тайлов. Наконец, категория F3 проявляется на уровне всей серии как пограничный seed 45, где прирост составляет лишь *+0.00030*, хотя направление эффекта остаётся положительным.

Рекомендации по интерпретации. Совместное рассмотрение Таблиц III и IV и Рис. 3 показывает, что BAMR не является универсальным улучшителем границ для произвольной сцены: его эффект геометрически локализован на тайлах с высокой кривизной контуров и мелкомасштабной структурой. Для прикладных задач массового картографирования это поведение благоприятно, поскольку именно фрактальные границы лесных массивов и плотная мелкая застройка составляют наиболее трудоёмкую часть ручной коррекции полигонов. Дальнейшее усиление эффекта возможно при увеличении ранга адаптеров или при дополнительной проверке на более широком наборе seed и сцен.

V. Обсуждение

Сохранение предобученных весов. Центральный вывод серии V1–V9 — критическая роль сохранения предобученных весов при модификации архитектуры для специализированного домена. Прослеживается монотонная зависимость: модели с полным сохранением предобученных путей (baseline, BAMR в обеих версиях) достигают наивысших результатов, тогда как вмешательства, нарушающие инициализацию основной сети или шеи (V1, BiFPN, MSProto), приводят к деградации. BAMR v2 добавляет лишь 6 400 параметров исключительно в масочную ветвь с одиночной нулевой инициализацией и обеспечивает статистически значимый прирост, тогда как MSProto (V7) добавляет ~1.7 M параметров с нарушением топологии Proto и достигает лишь 0.4433. Данное наблюдение согласуется с результатами исследований переноса признаков [20], [21], [27] и подтверждает, что при ограниченном доменном объёме данных (8 086 изображений) совместимость с предобучением является первостепенным фактором, превосходящим по значимости теоретическую выразительность добавляемых компонентов.

Сопоставление BAMR v1 и BAMR v2. Сопоставление двух конфигураций BAMR показывает, что при одинаковом пятисидовом протоколе различия между ними связаны не с общим направлением идеи, а с тем, насколько корректно активируется дополнительный путь уточнения масок. Для BAMR v1 сохраняется лишь направленный тренд (p = *0.130*), тогда как BAMR v2 даёт статистически значимый прирост (p = *0.0282*, 5/5 положительных разностей). Полученные результаты подтверждают сформулированную в начале работы гипотезу: на данном датасете выигрыш достигается не за счёт радикального усложнения модели, а за счёт аккуратного уточнения масочной ветви при сохранении предобученных весов.

Ограничение пространственного разрешения прототипов. Разница mAP50 − mAP50-95 ≈ 0.24 на baseline означает, что модель надёжно обнаруживает объекты (~72 % при IoU > 0.50), но не может точно воссоздать их контуры (~48 % при строгом усреднении). Этот разрыв особенно выражен для класса *woodland*, где нерегулярные границы приводят к непропорциональному штрафу при порогах IoU > 0.75. Совокупность результатов абляции V1–V9 указывает, что основным узким местом остаётся пространственное разрешение ветви масок, а не семантическая ёмкость модели. Модуль BAMR v2 адресует именно данное узкое место, применяя целенаправленное LoRA-уточнение на уровне прототипов, что позволяет корректировать геометрические ошибки, недоступные стандартной ветви Proto.

Контроль валидности и границы выводов. При интерпретации результата остаются важны два ограничения: малое число matched seed и неоднородность выигрыша по типам сцен. Тем не менее представленный методологический контур существенно снижает оба риска: парный дизайн и режим `deterministic=True` снижают дисперсию baseline, BAMR v2 даёт статистически значимый прирост, а qualitative-анализ прямо показывает, где эффект выражен сильнее, а где почти исчезает.

VI. Заключение

В настоящей работе исследован модуль BAMR для архитектуры YOLO11 в задаче инстанс-сегментации аэрофотоснимков LandCover.ai. Результаты показывают, что BAMR v2 обеспечивает статистически значимый прирост mask mAP50-95 при сохранении минимального параметрического overhead и совместимости с

предобученными весами.

Парная пятисидовая валидация даёт среднюю разность $\Delta = +0.00415 \pm 0.00276$ при $t = 3.365$ и $p = 0.0282$, тогда как BAMR v1 сохраняет лишь направленный, но статистически незначимый тренд ($p = 0.130$). Полученные результаты подтверждают, что главным резервом качества на данном датасете остаётся аккуратное уточнение границ масок в ветви Proto, а не радикальное усложнение всей архитектуры.

Перспективные направления дальнейших исследований включают увеличение числа seed, дополнительную проверку на других аэрофотосъёмочных наборах данных и дальнейшее уточнение масочной ветви для сцен со слабо выраженным эффектом.

БИБЛИОГРАФИЯ

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE CVPR*, 2016, pp. 779-788.

[2] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proc. IEEE CVPR*, 2017, pp. 6517-6525.

[3] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018.

[4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv:2004.10934, 2020.

[5] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE CVPR*, 2023, pp. 7464-7475.

[6] Ultralytics, "YOLO11," https://docs.ultralytics.com/models/yolo11/, 2025. [Accessed: Feb. 20, 2026].

[7] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time Instance Segmentation," in *Proc. IEEE ICCV*, 2019, pp. 9157-9166.

[8] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++: Better Real-time Instance Segmentation," *IEEE Trans. PAMI*, vol. 44, no. 2, pp. 1108-1121, 2022.

[9] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE ICCV*, 2017, pp. 2961-2969.

[10] X. Wang, T. Kong, C. Shen, Y. Jiang, and L. Li, "SOLO: Segmenting Objects by Locations," in *Proc. ECCV*, 2020, pp. 649-665.

[11] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "SOLOv2: Dynamic and Fast Instance Segmentation," in *Proc. NeurIPS*, 2020, pp. 17721-17732.

[12] Z. Tian, C. Shen, and H. Chen, "Conditional Convolutions for Instance Segmentation," in *Proc. ECCV*, 2020, pp. 282-298.

[13] A. Kirillov, Y. Wu, K. He, and R. Girshick, "PointRend: Image Segmentation as Rendering," in *Proc. IEEE CVPR*, 2020, pp. 9799-9808.

[14] B. Cheng, R. Girshick, P. Dollar, A. C. Berg, and A. Kirillov, "Boundary IoU: Improving Object-Centric Image Segmentation Evaluation," in *Proc. IEEE CVPR*, 2021, pp. 15334-15342.

[15] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated Shape CNNs for Semantic Segmentation," in *Proc. IEEE ICCV*, 2019, pp. 5229-5238.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proc. IEEE CVPR*, 2018, pp. 7132-7141.

[17] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proc. ECCV*, 2018, pp. 3-19.

[18] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local Neural Networks," in *Proc. IEEE CVPR*, 2018, pp. 7794-7803.

[19] A. Vaswani et al., "Attention Is All You Need," in *Proc. NeurIPS*, 2017, pp. 5998-6008.

[20] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Proc. IEEE CVPR*, 2017, pp. 2117-2125.

[21] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," in *Proc. IEEE CVPR*, 2018, pp. 8759-8768.

[22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *Proc. 3DV*, 2016, pp. 565-571.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, 2015, pp. 234-241.

[24] A. Boguszewski, D. Batorski, N. Ziemba-Jankowska, T. Dziedzic, and A. Zambrzycka, "LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water and Roads from Aerial Images," in *Proc. IEEE/CVF CVPR Workshops*, 2021, pp. 1102-1110.

[25] G.-S. Xia et al., "DOTA: A Large-scale Dataset for Object Detection in Aerial Images," in *Proc. IEEE CVPR*, 2018, pp. 3974-3983.

[26] S. W. Zamir et al., "iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images," in *Proc. CVPR Workshops*, 2019.

[27] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, "Going Deeper with Image Transformers," in *Proc. IEEE ICCV*, 2021, pp. 32-42.

[28] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-Rank Adaptation of Large Language Models," in *Proc. ICLR*, 2022.

[29] K. Zhang and D. Liu, "Customized Segment Anything Model for Medical Image Segmentation," arXiv:2304.13785, 2023.

[30] S. Aleem, J. Dietlmeier, E. Arazo, and S. Little, "ConvLoRA and AdaBN-based Domain Adaptation via Self-Training," in *Proc. IEEE ISBI*, 2024.

[31] D. Picard, "Torch.manual_seed(3407) is all you need: On the influence of random seeds in deep learning architectures for computer vision," arXiv:2109.08203, 2021.

[32] X. Bouthillier, P. Delaunay, M. Bronzi et al., "Accounting for Variance in Machine Learning Benchmarks," in *Proc. MLSys*, 2021, pp. 747-769.

[33] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.

Boundary-Aware Mask Refinement for YOLO11 Instance Segmentation of Aerial Imagery

K.A. Budakov, E.V. Druzhinskaya

***Abstract*— This study investigates the Boundary-Aware Mask Refinement (BAMR) module for improving instance segmentation quality in aerial imagery using the YOLO11 architecture. Experiments are conducted on a two-class LandCover.ai setup under a unified training protocol with a pretrained baseline model. The BAMR v1 configuration shows only a directional improvement trend that does not reach statistical significance on five matched seeds ($p = 0.130$). For the BAMR v2 configuration, implemented as a minimal Proto-branch modification with low-rank adapters, five-seed paired validation confirms a statistically significant improvement in mask mAP50-95: the mean paired difference is $+0.00415 \pm 0.00276$ with $t = 3.365$, $p = 0.0282$, a 95 % confidence interval of $[+0.00073, +0.00757]$, and 5 of 5 positive paired differences. Qualitative analysis indicates that the gain is most visible on elongated and curved boundaries, whereas the effect becomes small on large smooth woodland polygons. Overall, the results confirm the hypothesis that the main reserve for improvement on this dataset lies in better mask-branch boundary refinement while preserving pretrained weights.**

***Keywords*— instance segmentation, aerial imagery, YOLO11, BAMR, boundary refinement, mask mAP50-95, remote sensing.**

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE CVPR*, 2016, pp. 779-788.

[2] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proc. IEEE CVPR*, 2017, pp. 6517-6525.

[3] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018.

[4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv:2004.10934, 2020.

[5] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE CVPR*, 2023, pp. 7464-7475.

[6] Ultralytics, "YOLO11," https://docs.ultralytics.com/models/yolo11/, 2025. [Accessed: Feb. 20, 2026].

[7] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time Instance Segmentation," in *Proc. IEEE ICCV*, 2019, pp. 9157-9166.

[8] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++: Better Real-time Instance Segmentation," *IEEE Trans. PAMI*, vol. 44, no. 2, pp. 1108-1121, 2022.

[9] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE ICCV*, 2017, pp. 2961-2969.

[10] X. Wang, T. Kong, C. Shen, Y. Jiang, and L. Li, "SOLO: Segmenting Objects by Locations," in *Proc. ECCV*, 2020, pp. 649-665.

[11] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "SOLOv2: Dynamic and Fast Instance Segmentation," in *Proc. NeurIPS*, 2020, pp. 17721-17732.

[12] Z. Tian, C. Shen, and H. Chen, "Conditional Convolutions for Instance Segmentation," in *Proc. ECCV*, 2020, pp. 282-298.

[13] A. Kirillov, Y. Wu, K. He, and R. Girshick, "PointRend: Image Segmentation as Rendering," in *Proc. IEEE CVPR*, 2020, pp. 9799-9808.

[14] B. Cheng, R. Girshick, P. Dollar, A. C. Berg, and A. Kirillov, "Boundary IoU: Improving Object-Centric Image Segmentation Evaluation," in *Proc. IEEE CVPR*, 2021, pp. 15334-15342.

[15] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated Shape CNNs for Semantic Segmentation," in *Proc. IEEE ICCV*, 2019, pp. 5229-5238.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proc. IEEE CVPR*, 2018, pp. 7132-7141.

[17] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proc. ECCV*, 2018, pp. 3-19.

[18] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local Neural Networks," in *Proc. IEEE CVPR*, 2018, pp. 7794-7803.

[19] A. Vaswani et al., "Attention Is All You Need," in *Proc. NeurIPS*, 2017, pp. 5998-6008.

[20] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Proc. IEEE CVPR*, 2017, pp. 2117-2125.

[21] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," in *Proc. IEEE CVPR*, 2018, pp. 8759-8768.

[22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *Proc. 3DV*, 2016, pp. 565-571.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, 2015, pp. 234-241.

[24] A. Boguszewski, D. Batorski, N. Ziemba-Jankowska, T. Dziedzic, and A. Zambrzycka, "LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water and Roads from Aerial Images," in *Proc. IEEE/CVF CVPR Workshops*, 2021, pp. 1102-1110.

[25] G.-S. Xia et al., "DOTA: A Large-scale Dataset for Object Detection in Aerial Images," in *Proc. IEEE CVPR*, 2018, pp. 3974-3983.

[26] S. W. Zamir et al., "iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images," in *Proc. CVPR Workshops*, 2019.

[27] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, "Going Deeper with Image Transformers," in *Proc. IEEE ICCV*, 2021, pp. 32-42.

[28] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-Rank Adaptation of Large Language Models," in *Proc. ICLR*, 2022.

[29] K. Zhang and D. Liu, "Customized Segment Anything Model for Medical Image Segmentation," arXiv:2304.13785, 2023.

[30] S. Aleem, J. Dietlmeier, E. Arazo, and S. Little, "ConvLoRA and AdaBN-based Domain Adaptation via Self-Training," in *Proc. IEEE ISBI*, 2024.

[31] D. Picard, "Torch.manual_seed(3407) is all you need: On the influence of random seeds in deep learning architectures for computer vision," arXiv:2109.08203, 2021.

[32] X. Bouthillier, P. Delaunay, M. Bronzi et al., "Accounting for Variance in Machine Learning Benchmarks," in *Proc. MLSys*, 2021, pp. 747-769.

[33] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.