

УДК 004.93

doi: 10.15622/rcai.2025.066

ПРИМЕНЕНИЕ МЕТОДОВ 2D СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ НА ОСНОВЕ АРХИТЕКТУРЫ ТРАНСФОРМЕРА В ЗАДАЧЕ ДИСТАНЦИОННОГО ИССЛЕДОВАНИЯ СТРУКТУРЫ АРХЕОЛОГИЧЕСКИХ ПАМЯТНИКОВ¹

А.В. Вохминцев (*vav@csu.ru*)^A

В.Р. Аббазов (*abbavar@yandex.ru*)^B

М.А. Романов (*std.romanov.ma@gmail.com*)^A

Т.С. Вохминцева (*Atska85@mail.ru*)^B

М. Хатер (*Mostafa.khater2024@xzhmu.edu.cn*)^{B,C}

А.В. Мельников (*MelnikovAV@uriit.ru*)^B

^A Челябинский государственный университет, Челябинск

^B Югорский государственный университет, Ханты-Мансийск

^C Университет Суйчжоу, Суйчжоу, Китай

Данная работа посвящена разработке методов семантической сегментации археологических памятников с применением методов машинного обучения и 2d цифровых моделей археологического ландшафта по данным аэрофотосъемки и дистанционного зондирования Земли, был проведен анализ современных моделей для детектирования объектов и сегментации изображений, рассмотрены архитектуры на основе модели трансформера, такие как InternImage, BEiT Pre-Training of Image Transformers, проведено обучение моделей с использованием коллекции данных об археологических памятниках бронзового века Южного Зауралья с оценкой на валидационном наборе, полученные результаты представлены и обсуждены.

Ключевые слова: семантическая сегментация, трансформеры, цифровая модель рельефа, археологический памятник.

¹ Работа выполнена при финансовой поддержке РНФ (проект № 23-11-20007).

Введение

Данное исследование посвящено проблеме разработки системы для обнаружения и исследования археологических памятников на основе методов машинного обучения, картографирования, ГИС, геофизики и цифровых моделей археологического ландшафта. В период 80-х – 90-х гг. XX века проведена аэрофотосъемка с использованием малой авиации территории Южного-Зауралья и Северного Казахстана, при исследовании материалов съемок группой археологов под руководством проф. Здановича Г.Б. был открыт комплекс укрепленных поселений, датируемых XXI-XVIII вв. до н.э. бронзового века, который впоследствии стали относить к синташтинской археологической культуре. Самым известным обнаруженным укрепленным поселением является Аркаим, однако для исследователей наибольший интерес представляют менее известные поселения, например, археологические памятники вблизи поселений Степное и Левобережное. В результате археологических экспедиций по проекту РНФ № 23-11-20007 проведены поисковые работы и собрана цифровая коллекция данных об утраченных памятниках археологии, открытых в XIX-XX вв. без привязки к системе координат, поврежденных памятниках со снятым верхним почвенным слоем, а также нечитаемых на местности объектов в известных археологических памятниках (грунтовых захоронениях, ритуальных и хозяйственных сооружениях без выраженных в рельефе признаков). Возле многих из укреплений поселений обнаружены отдельные неукрепленные поселения и курганные могильники. Комплекс археологических памятников синташтинской культуры, так называемая “Страна городов” является уникальным объектом для проведения междисциплинарных исследований с использованием методов геофизики, математического моделирования и искусственного интеллекта. Дистанционные методы исследования позволяют сохранить культурное наследие для будущих поколений, а также решать важные практические задачи, например, определение границ памятника и точных мест для забивки шурфов с целью получения археологического материала без раскопки поверхности памятника. В исследовании [Vokhmintcev et al., 2023a] предложена методология для картирования и съемки археологических объектов на основе методов машинного обучения, методов геофизики и картографирования. Ключевым компонентом данной методологии являются методы 2d сегментации данных, в данной статье предлагается оригинальный метод 2d сегментации археологических памятников на основе архитектуры трансформера. Согласно бенчмаркинг лидирующие позиции при решении поставленной задачи принадлежат методам сегментации на основе архитектуры трансформера, можно отметить следующие модели InternImage [Wang et al., 2023a], ONE-PEACE [Wang et al., 2023b], BEiT: BERT Pre-Training of Image Transformers [Bao et al., 2023], ViT-Adapter [Chen et al., 2023]. Давайте рас-

смотрим данные модели более подробно, InternImage – это крупномасштабная модель, основанная на свёрточных нейронных сетях и разработанная для задач компьютерного зрения, включая классификацию изображений, обнаружение объектов и семантическую сегментацию. Архитектура данной сети содержит базовый блок модели Deformable Convolution v3 (DCNv3), слой нормализации, функцию активации GELU и полносвязную сеть. Блок DCNv3 улучшает стандартную деформируемую свёртку, вводя обучаемые смещения и коэффициенты, которые позволяют ядру свёртки динамически приспосабливаться к структуре входного изображения. Это обеспечивает эффективное улавливание как локальных, так и глобальных зависимостей в данных. Процесс обработки данных в модели разделен на несколько этапов с последовательным уменьшением пространственного разрешения, что позволяет извлекать и обрабатывать признаки на разных уровнях абстракции с высоким значением метрики mask AP, модель протестирована на разных эталонных коллекциях данных, показывает выдающиеся результаты для коллекции COCO test-dev. Модель ONE-PEACE представляет собой универсальную архитектуру для обучения представлений, способную интегрировать и согласовывать данные различных модальностей: изображения, аудио и текста. Для каждой модальности модель использует специальные адаптеры, преобразующие исходные данные в последовательности признаков. Процесс обработки модальности изображений использует иерархический MLP-сте́м (hMLP) для разбивки изображения на сегменты размером 16×16 . Mask2Former применяется в качестве головы модели, обеспечивая подход к семантической сегментации с поддержкой различных задач сегментации. Модель сначала обучается на наборе данных COCO-Stuff для лучшего понимания контекстов объектов, а затем дообучается на целевом наборе данных ADE20K, достигая выдающихся результатов для метрики mIoU. BEiT – модель, которая изначально нацелена на обработку естественного языка (NLP), но затем разработчики адаптировали ее для задач в области компьютерного зрения. В основе архитектуры BEiT находится идея моделирования маскированного изображения (Masked Image Modeling MIM). Работа модели BEiT основана на случайном процессе маскирования определенной части сегментов исследуемого изображения. Далее модель обучается восстанавливать соответствующие визуальные токены, используя контекст оставшихся непрерывных сегментов. Модель BEiT основана на классической архитектуре трансформера при обработке последовательности сегментов изображения, что позволяет эффективно моделировать долгосрочные зависимости между различными частями изображения и интегрировать глобальные признаки объектов. Основным преимуществом модели BEiT является способность работать на глобальном уровне, предсказывая дискретные визуальные токены вместо предсказания пиксель-

ных значений маскированных сегментов, что приводит во многих известных методах 2d сегментации к фокусировке на локальном уровне сегментации, который затрагивает низкоуровневые детали изображения. Для получения визуальных токенов используется дискретный вариационный автоэнкодер dVAE [Vahdat et al., 2018], который осуществляет процесс токенизации изображения в последовательность дискретных представлений изображения. Данная особенность модели BEiT позволяет концентрироваться на высокоуровневых признаках объектов и семантической информации. Модель ViT-Adapter является улучшенной модификацией архитектуры трансформера Vision Transformer (ViT) [Dosovitskiy et al., 2018], которая позволяет повысить эффективность решения задачи 2d семантической сегментации. Модель ViT-Adapter состоит из трех компонентов: модуль 1 – пространственного приоритета (Spatial Prior Module, SPM), модуль 2 – инжектор пространственных признаков, модуль 3 для извлечения многоуровневых признаков. Модуль 1 содержит нескольких свёрточных слоёв и слоёв max-pooling, которые позволяют извлекать локальные пространственные признаки из изображения. Модуль 1 создает многоуровневые пространственные признаки с различным разрешением: 1/8, 1/16 и 1/32 от размера изображения. Модуль 2 интегрирует локальные пространственные признаки из SPM в одноуровневые токены ViT. Модуль 3 осуществляет реконструкцию многоуровневых признаков из исходных одноуровневых признаков ViT. Работа модели трансформера ViT-Adapter основана на концепции многоэтапного взаимодействия признаков. Слои трансформера разделяются на несколько блоков, на каждом этапе обработки осуществляется обмен информацией между локальными и глобальными признаками из SPM через Модуль 2 и Модуль 3. В данной работе согласно бенчмаркинг (<https://paperswithcode.com/task/semantic-segmentation>) для компьютерного моделирования выбраны модели InternImage и BEiT, проведено компьютерное моделирование отобранных моделей применительно к задаче 2d семантической сегментации, которое показало необходимость внесения изменений в архитектуру трансформеров. Статья организована следующим образом: в разделе 1 представлено описание коллекции данных археологических памятников бронзового века Южного Зауралья, в разделе 2 предложены оригинальные методы 2d семантической сегментации на основе архитектуры трансформера, в разделе 3 представлены результаты компьютерного моделирования.

1. Признаки дешифрирования археологических памятников бронзового века

Коллекция данных археологических памятников бронзового века Южного Зауралья создана на основе материалов экспедиций и полевых работ, проведенных специалистами учебно-научного центра изучения проблем

природы и человека ЧелГУ и историко-археологического музея-заповедника «Аркаим», начиная с 80-х годов прошлого века по настоящее время. Материалы для коллекции данных подобраны так, чтобы максимально представить типичные памятники степной зоны Зауралья: 22 укрепленных поселения эпохи бронзы, 16 неукрепленных поселений эпохи бронзы, 12 курганов и курганных могильников разных эпох, 5 средневековых курганов «с усами», 4 могильника с гантелевидными и подковообразными насыпями. Археологические памятники были обследованы с применением аэрофотосъемки, тахеометрической съемки и электроразведки: трехмерные данные были получены с использованием тахеометра сканеров глубины LiDAR в виде ортофотопланов и облаков точек, тахеометра Trimble 3300 и системы индукционного профилирования АЭМП-14 соответственно. Аэрофотоснимки были сделаны в масштабе 1:14 000 с высоким разрешением для всей территории Кизильского района Челябинской области. На основе собранных данных была построена цифровая модель археологического ландшафта опорных участков вокруг поселений Степное и Левобережное и территории археологического микрорайона в среднем течении р. Синташта, введены описания структур данных, соответствующих классам археологических объектов на языке графического описания объектного моделирования. В данной работе в качестве источников данных для обучения моделей были использованы: результаты локальной аэрофотосъемки с использованием малой авиации и квадрокоптера DJIMini 2 (с 2022-2024 гг.) и результаты дистанционного зондирования Земли с зарубежных и российских спутников Sentinel-2 (с 2015 г.), Ресурс-П, (с 2013 по 2021 гг.), Канопус-В, (с 2013 по 2023 гг.) с пространственным разрешением 10 м, от 0.7 до 1.5 м, 2.1 м соответственно. Аэрофотоснимки прошли обработку в ПО Agisoftmetashape перед включением в коллекцию данных. В 2023 г. было открыто два новых укрепленных поселения Верхнеуральское и Нижнеуспенское, что стало большим событием в российской археологии: новые памятники бронзового века заполняли собой «белое пятно» на карте так называемой «Страны городов» между поселениями Степное и Черноречье и остальными объектами. В этой работе используется следующие классы археологических объектов: жилище и могильники.

2. Методы 2d семантической сегментации на основе архитектуры трансформера

Для реализации методов 2d семантической сегментации археологических памятников бронзового века Южного Зауралья предлагается подход, в котором можно выделить следующие шаги:

Шаг 1. Формирование набора снимков территорий, содержащих археологические памятники. Источником данных является существующая база аэрофотосъемки и космоснимков из коллекции данных.

Шаг 2. Создание разметки полученного набора. Разметка включает в себя классификацию археологических объектов по установленным классам и создание масок для задачи 2d семантической сегментации.

Шаг 3. Расширение размеченного набора данных в 10 раз с использованием различных вариантов аугментаций, основанных на повороте и переносе частей изображений, а также мозаичной технологии.

Шаг 4. Разделение аугментированного на шаге 3 набора снимков на обучающую, валидационную и тестовую выборку.

Шаг 5. Выполнение компьютерного моделирования современных моделей для 2d семантической сегментации изображений InternImage и BEiT на основе реальных данных и созданного на Шаге 4 набора снимков. Обучение данных моделей на тестовом наборе данных с оценкой моделей на валидационном наборе с целью предотвращения переобучения.

Шаг 6. Оценка моделей, отобранных на шаге 5 на тестовой выборке с использованием метрики IoU.

Далее рассмотрим подробно Шаг 5 подхода, в работе для 2d сегментации были использованы модели InternImage (см. Листинг 1) и BEiT (см. Листинг 2). Для задачи семантической сегментации InternImage используется в качестве энкодера, в качестве декодера используются такие модели как UperNet или SegFormer. Модифицированные шаги методов 2d сегментации отмечены символом *, были внесены изменения в процесс маскирования сегментов исследуемого изображения и восстановления визуальных токенов на основе контекста непрерывных сегментов.

Листинг 1

1:	procedure InternImage-UPerNet*(I)
2:	Входные данные: Изображение I / Выходные данные: Карта сегментации S
3:	// Этап 1. Извлечение признаков – backbone InternImage
4:	X_feat ← Преобразовать I в тензор входных признаков (X0)
5:	StageOutputs ← новый пустой словарь // хранение {X1,X2,X3,X4}
6:	for s_idx in 1,...,4 do // Индекс стадии (s_idx соответствует X1, X2, X3, X4)
7:	X_current_stage_processing ← X_feat // Рабочий тензор для текущей стадии
8:	for l_idx in 1,...,КоличествоБазовыхБлоков[s_idx] do // Индекс базового блока в текущей стадии
9:	X_residual_connection ← X_current_stage_processing
10:	// Вычисление смещений и модульных множителей DCNv3 на основе X_current_stage_processing

11:	Offsets, Modulators \leftarrow Вычислить DCN Параметры(X_current_stage_processing)
12:	H \leftarrow DCNv3(X_current_stage_processing, Offsets, Modulators) // H - промежуточный результат
13:	*H \leftarrow LayerNorm(H); H_ffn \leftarrow FFN(H)
15:	X_current_stage_processing \leftarrow H_ffn + X_residual_connection // Обновление тензора после блока
16:	end for // конец цикла по базовым блокам
17:	StageOutputs[s_idx] \leftarrow X_current_stage_processing
18:	if s_idx < 4 then // Если это не последняя стадия извлекающего признака блока
19:	// Понижение разрешения для входа следующей стадии
20:	X_feat \leftarrow Downsample(X_current_stage_processing)
21:	else // Для последней стадии (X4) нет понижения разрешения; этот X_feat (X4) пойдет в PPM
23:	X_feat \leftarrow X_current_stage_processing
24:	end if // конец условия понижения разрешения
25:	end for // конец цикла по стадиям
26:	// Извлечение именованных признаков стадий для удобства
27:	X1 \leftarrow StageOutputs[1,..., X4 \leftarrow StageOutputs[4];
31:	// Этап 2. Декодер UPerNet
32:	// 2.1 Pyramid Pooling Module (PPM)
33:	PPM_features \leftarrow PyramidPoolingModule(X4)
34:	// 2.2 Feature Pyramid Network (FPN)
35:	// FPN: объединение признаков с разных уровней через lateral connections и upsampling
36:	FPN_features \leftarrow FPN({PPM_features, X3, X2, X1})
37:	// 2.3 Объединение признаков
38:	// FPN_— это коллекция карт признаков с разных уровней FPN
39:	*FusedFeatures \leftarrow Конкатенация&Сверточное слияние(FPN_features)
40:	// Этап 3. Классификация и сегментация
41:	SegmentationMap_logits \leftarrow Conv1x1(FusedFeatures) // Получение логитов для каждого класса
42:	*SegmentationMap_upsampled \leftarrow Upsample(SegmentationMap_logits, до разрешения входного изображения I)
43:	// Этап 4. Финальный результат
44:	*S \leftarrow argmax(SegmentationMap_upsampled, по_каналам) // Выбор класса с максимальной вероятностью для каждого пикселя
45:	return S; end procedure.

Листинг 2

1:	procedure BEIT-UPerNet*(I)
2:	// Этап 1. Представление I и предварительная обработка
3:	X_tensor ← Преобразовать I в тензор входных признаков
4:	Patches ← Разделить X_tensor на сетку из N патчей размером P×P (16×16) // N = (H_img/P) * (W_img/P)
5:	TokenSequence ← [] // Пустой список для последовательности эмбедингов
6:	for каждый Патч_i in Patches do
7:	Embedding_i ← ЛинейноеПреобразование(Патч_i) // Patch Embedding
8:	TokenSequence.Добавить(Embedding_i)
9:	end for // TokenSequence теперь имеет размерность N x D, где D – размерность эмбединга
10:	TokenSequence ← Добавить Абсолютные Позиционные Эмбединги(TokenSequence) // К каждому токenu
11:	// Этап 2. BEIT Encoder (Vision Transformer)
12:	// L – общее количество слоев трансформера
13:	// FeatureExtractionLayers – индексы слоев, после которых извлекаются признаки
14:	EncoderStageOutputs_raw ← новый пустой список // для {токены_стадии1, ..., токены_стадии4}
15:	CurrentTokenSequence ← TokenSequence // Рабочая последовательность токенов
16:	for l_idx in 1,...,L do // Индекс слоя трансформера
17:	InputToBlock ← CurrentTokenSequence
18:	// Multi-Head Self-Attention с относительным позиционным смещением (bias)
19:	AttentionOutput ← MultiHeadSelfAttention(InputToBlock, относительные позиционные bias)
20:	Residual1 ← InputToBlock + AttentionOutput // Первое skip
21:	Normalized1 ← LayerNorm(Residual1) // Первая нормализация
22:	// Feed-Forward Network
23:	*FFNOutput ← FFN(Normalized1)
24:	*Residual2 ← Normalized1 + FFNOutput // Второе skip
25:	CurrentTokenSequence ← LayerNorm(Residual2)
26:	InputToBlock_l ← CurrentTokenSequence
27:	*AttentionOut_l ← MultiHeadSelfAttention(InputToBlock_l, с относительными позиционными bias)
28:	Sum1_l ← InputToBlock_l + AttentionOut_l // Остаточное соединение
29:	Norm1_l ← LayerNorm(Sum1_l)
30:	FFNOut_l ← FFN(Norm1_l)
31:	CurrentTokenSequence ← Norm1_l + FFNOut_l // Остаточное соединение

32	if l_idx находится в FeatureExtractionLayers then
33	PatchTokens_l ← ОтброситьCLСтокенЕслиЕсть(CurrentTokenSequence)
34	EncoderStageOutputs_raw.Добавить(PatchTokens_l)
35	end if
36	end for // конец цикла по слоям трансформера
37	// Преобразование извлеченных последовательностей токенов в 2D карты признаков
38	// H_feat = H_img/P, W_feat = W_img/P
39	// D_stage_i - размерность эмбединга на выходе i-ой выбран- ной стадии
40	X1 ← ReshapeTo2D(EncoderStageOutputs_raw[0], H_feat, W_feat) // Карта признаков (H feat, W feat, D_stage 1) . . .
43	X4 ← ReshapeTo2D(EncoderStageOutputs_raw[3], H_feat, W_feat) // Карта признаков (H feat, W feat, D_stage 4)
44:	// Этап 2. Декодер UPerNet
45:	// 2.1 Pyramid Pooling Module (PPM)
46:	*PPM_features ← PyramidPoolingModule(X4)
47:	// 2.2 Feature Pyramid Network (FPN)
48:	// FPN: объединение признаков с разных уровней через lateral connections и upsampling
49:	FPN_features ← FPN({PPM_features, X3, X2, X1})
50:	// 2.3 Объединение признаков
51:	// FPN – это коллекция карт признаков с разных уровней FPN
52:	FusedFeatures ← Конкатенация&Сверточное_Слияние(FPN_features)
53:	// Этап 3. Классификация и сегментация
54:	SegmentationMap_logits ← Conv1x1(FusedFeatures) // Получение логитов для каждого класса
55:	*SegmentationMap_upsampled ← Upsample(SegmentationMap_logits, до разрешения входного изображения I)
56:	// Этап 4. Финальный результат
57:	*S ← argmax(SegmentationMap_upsampled, по_каналам) // Выбор класса с максимальной вероятностью для каждого пикселя
58:	return S; end procedure.

3. Компьютерное моделирование

В данном разделе представлены и обсуждены результаты компьютерного моделирования, для тестов применялся компьютер на базе Intel Core i7-9800X с графическим процессором NVIDIA GeForce RTX 2080 Ti 11264MiB. Оценим точность и сходимость предложенных методов на примере коллекции данных об археологических памятниках бронзового века Южного Зауралья, которая состоит из 9205 кадров в обучающей выборке и 3229 в тестовой. Результаты 2d семантической сегментации с использованием различных вариантов моделей InternImage и BEiT

для классов жилище и могильник приведены на рис. 1 и 2 соответственно. Для обучения модели InternImage использованы следующие параметры обучения: число эпох 100, оптимизатор AdamW, функция потерь Cross Entropy Loss, Learning Rate – LR, базовый learning rate – 0.00006, использовано полиномиальное затухание, был применен warmup со следующими параметрами: тип – линейный, количество итераций warmup – 5 (по эпохам), начальное значение – $1e^{-6}$ от основного LR. Для обучения были использованы модели InternImage-S (Small) и InternImage-B (Base).

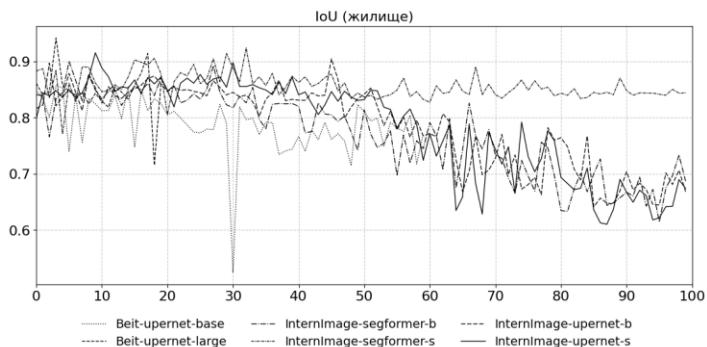


Рис. 1. Качество 2d семантической сегментации для класса жилище

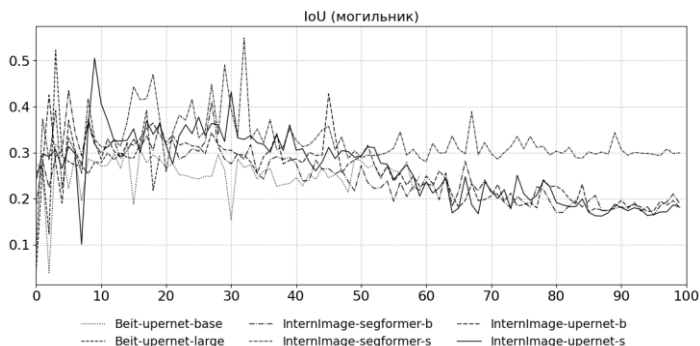


Рис. 2. Качество 2d семантической сегментации для класса могильник

Лучшие значения метрик IoU для разных моделей InternImage и BEiT представлены в табл. 1. Примеры сегментированных масок для неукрепленного поселения в среднем течении реки Синташта в Челябинской области приведены на рис. 3, масками отмечены объекты класса жилище. Для обучения модели BEiT использованы параметры: количество эпох обу-

чения для модели base – 60, для модели large – 20, использованы предобученные модели microsoft/beit-base-finetuned-ade-640-640 и microsoft/beit-large-finetuned-ade-640-640 соответственно, оптимизатор – AdamW, размер патча – 16, функция потерь – Cross Entropy Loss, learning rate – $5e^{-5}$. Дополнительно отметим, предложенные модели InternImage и BEiT производят 2d семантическую сегментацию лучше, чем базовые версии моделей: для InternImage-S разница составляет 0.15 пункта, InternImage-B – 0.09 пункта, Beit-upernet-base – 0.06 пункта, Beit-upernet-large – 0.08 пункта.

Таблица 1

Название модели	IoU, среднее (классы жилище и могильник)	IoU класс могильник	IoU класс жилище
InternImage-upernet, Small (s)	0.710	0.505	0.915
InternImage-segformer, Small (s)	0.737	0.549	0.924
InternImage-upernet, Base (b)	0.666	0.428	0.904
InternImage-segformer, Base (b)	0.668	0.436	0.900
Beit-upernet-base	0.587	0.320	0.857
Beit-upernet-large	0.732	0.522	0.942

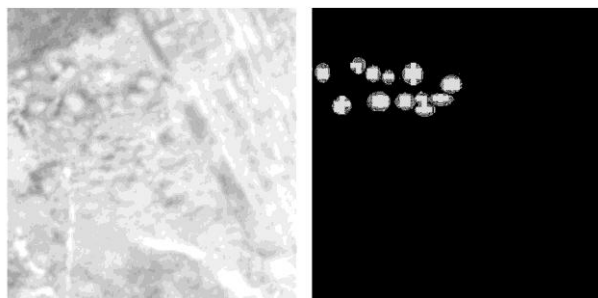


Рис. 3. Примеры сегментированных масок для неукрепленного поселения

В процессе обучения моделей для класса могильник получены невысокие значения по метрике качества IoU, с одной стороны это объясняется признаками дешифрирования данного класса [Vokhmintcev et al., 2023a], а с другой стороны связано с архитектурой трансформера BEiT. Решение данной проблемы видится в использовании для обучения моделей результатов магнитометрической съёмки, которая позволит идентифицировать могильники бронзового века с существенно более высокой точностью по сравнению с другими геофизическими методами и методами дистанционного зондирования Земли.

Заключение

В работе предложены модифицированные варианты методов InternImage и BEiT на основе архитектуры трансформера для задачи 2d семантической сегментации археологических памятников бронзового века Южного Зауралья на основе цифровой модели археологического ландшафта, была использована коллекция аэрофотоснимков и космических снимков с высоким пространственным разрешением. Полученные результаты позволяют автоматизировать процесс сегментации структуры археологического памятника, предложенные методы позволили с высокой точностью определять границы жилищ в поселениях бронзового века и как следствие определять точные места для забивки шурфов для получения археологического материала, не прибегая к раскопке поверхности всего памятника.

Список литературы

- [Bao et al., 2023] Bao H., Dong L., Piao S., Wei F. BEiT: BERT Pre-Training of Image Transformers // In: Proc. The 10-th International Conference on Learning Representations (ICLR), virtual event, 2022. – P. 1-16. – <https://arxiv.org/pdf/2106.08254v1>.
- [Chen et al., 2023] Chen Z., Duan Y., Wang W., He J., Lu T., Dai J., Qiao Y. Vision Transformer Adapter for Dense Predictions // The 11-th International Conference on Learning Representations (ICLR), Kigali, Rwanda, 2023. – P. 1-20. – arXiv preprint arXiv:2205.08534.
- [Dosovitskiy et al., 2018] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // The 9-th International Conference on Learning Representations (ICLR), virtual event, 2021. – P. 1-22. – arXiv:2010.11929v2. [8].
- [Vahdat et al., 2018] Vahdat A., Andriyash E., Macready W.G. DVAE#: Discrete Variational Autoencoders with Relaxed Boltzmann Priors // The 22th Annual Conference on Neural Information Processing Systems (NIPS), Canada 2018. – 7.
- [Vokhmintcev et al., 2023a] Vokhmintcev A.V., Melnikov A.V., Romanov M.A. and etc. Research System of Archaeological Sites Using Deep Learning // Pattern Recognition and Images Analysis. – 2023. – No. 3. – P. 304-315. – doi: 10.1134/S105466182470038X.
- [Vokhmintcev et al., 2024b] Vokhmintcev A.V., Khristodulo O.I., Melnikov A.V. and etc. Application of Dynamic Graph CNN* and FICP for Detection and Research Archaeology Sites // International Conference on Analysis of Images, Social Networks and Texts. Ed. by D.I. Ignatov, M. Khachay, A. Kutuzov and etc. Lecture Notes in Computer Science. – 2024. – No. 14486. – P. 294-308. – doi:10.1007/978-3-031-54534-4_21.
- [Wang et al., 2023a] Wang W., Dai J., Chen Z., Huang Z., Li Z., Zhu X., Hu X., Lu T., Lu L., Li H., Wang X., Qiao Y. InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 2023. – P. 14408-14419. – <https://arxiv.org/abs/2211.05778>.
- [Wang et al., 2023b] Wang P., Wang S., Lin J., Bai S., Zhou X., Zhou J., Wang X., Zhou C. ONE-PEACE: Exploring One General Representation Model Toward Unlimited Modalities [Электронный ресурс] // Computer Vision and Pattern Recognition 2023. URL: <https://arxiv.org/abs/2305.11172> (дата обращения 19.04.2025).