



Применение модели YOLOv11 и адаптированного набора данных LaDD для поиска людей в малонаселённой местности

Александр Владимирович **Смирнов**¹, Игорь Петрович **Тищенко**²,
Сергей Александрович **Лазарев**³

¹⁻³ Институт программных систем им. А. К. Айламазяна РАН, Вельского, Россия

Аннотация. Настоящее исследование направлено на разработку нейросетевого метода обнаружения/детекции людей в малонаселённой местности по снимкам, полученным с беспилотного воздушного судна (БВС). В качестве нейросетевого детектора была использована модель архитектуры YOLOv11m. В рамках проведённого исследования был разработан и применён алгоритм адаптации обучающего набора данных LaDD. Проведены эксперименты по предварительному обучению модели на оригинальном и адаптированном наборе, в результате которых удалось доказать целесообразность использования адаптированного набора данных. Итоговая точность модели при обучении достигла значения в 98.7% по mAP^{50} . Инференс модели показал точность детекции в 0.895 (89.5%) по F1 и 0.901 (90.1%) по mAP^{50} , что позволяет подтвердить работоспособность представленного метода.

Ключевые слова и фразы: Анализ изображений, поиск людей, снимки БВС, YOLOv11, нейронные сети, набор данных, адаптация

Для цитирования: Смирнов А. В., Тищенко И. П., Лазарев С. А. *Применение модели YOLOv11 и адаптированного набора данных LaDD для поиска людей в малонаселённой местности* // Программные системы: теория и приложения. 2025. Т. 16. № 4(67). С. 217–240. https://psta.psisras.ru/read/psta2025_4_217-240.pdf

Введение

Обработка данных фото- и видеосъемки с беспилотного воздушного судна (БВС) играет важную роль во многих сферах жизни, в числе которых: строительство, маркетинг, сельское хозяйство, а также другие направления, включая социально значимые. Среди социально значимых направлений можно выделить задачу поиска людей на аэрофотоснимках. Согласно *статистическим данным*^{[URL](#)}, в 2022 году было более 42 тысяч обращений в поисково-спасательные организации для поиска пропавших людей. Развитие IT-сферы позволяет использовать высокие технологии для решения данной задачи прежде всего благодаря тому, что с высоты полета БВС можно исследовать гораздо большую площадь, чем при наземном обследовании территории.

В настоящее время многие алгоритмы обнаружения людей на снимках, полученных с БВС, используют технологии искусственных нейронных сетей (ИНС). Это стало возможным благодаря появлению релевантных обучающих наборов данных. Для содействия прогрессу в исследованиях обнаружения малых объектов при помощи БВС, авторы работы [1] представили набор данных для обнаружения людей *Manipal-UAV*. Рассматриваемый набор содержит 13 462 выборочных изображения из 33 видеороликов, содержащих 1 53 112 экземпляров целевых объектов – людей. Съёмка видео была выполнена в условиях неограниченной среды со сложными сценами, охватывающими небольшие объекты различных масштабов, поз, освещённости и окклюзии. В данной статье также сравниваются особенности набора данных *Manipal-UAV*^{[URL](#)} со стандартными наборами, такими как *VisDrone* и *Okutama*, содержащими объекты, снятые с воздуха.

Поиск людей на снимках с БВС в системе *DronesDetect*^{[URL](#)} осуществляется по средствам нейросетевой обработки данных с тепловизора. Таким образом, использование тепловой карты позволяет обнаруживать людей даже при недостаточной освещённости.

Использование изображений вне видимого диапазона также рассмотрено в работе [2]. Для решения проблемы обнаружения/детекции людей в условиях недостаточной освещённости, авторы представляют решение на основе использования изображений, снятых в инфракрасном диапазоне. Предлагаемый метод использует IPD-YOLO – усовершенствованный алгоритм глубокого обучения для обнаружения людей на инфракрасных изображениях, полученных с дронов. Предложенная архитектура IPD-YOLO основана на базовой модели *YOLOv11*^{[URL](#)}. В качестве модификаций вводятся модуль извлечения признаков MASRCNet, детектор LQENet и функция потерь NWD-Inner CIoU. Эксперименты показали,

что использование предложенной комбинации модификаций базовой модели демонстрирует улучшение на 4.7, 7.4, 6.3 и 6.4% по метрике mAP50 по сравнению с основными вариантами YOLO, такими как *YOLOv5n*^{URL}, *YOLOv8n*^{URL}, *YOLOv10n*^{URL} и YOLOv11n соответственно.

В статье [3] рассмотрена задача поиска людей на аэрофотоснимках, сделанных с помощью дронов, при проведении поисково-спасательных операций. Для решения поставленной задачи исследована возможность применения метода *raycast* для геолокации и определения расстояния до человека в реалистичных условиях. Результатом проведённого исследования стала система SAR-DAG, способная точно выполнять геолокацию людей, обнаруженных на обработанных в автономном режиме изображениях, полученных в ходе поисково-спасательной операции. Разработанная система SAR-DAG включает в себя 3D-генератор рельефа и модуль *raycast*, а также детектор людей на основе глубокой нейронной сети, обученный на собственном наборе данных *SAR-DAG_raycast*^{URL}. Предложенный метод показал погрешность геолокации 0.7 м при съёмке с высоты 30 м.

В рамках исследования [4] рассмотрено использование свёрточных нейронных сетей для решения задачи обнаружения пропавших людей на снимках с БВС. Авторы предложили создание и использование так называемой карты ScoreMap для выделения участков, на которых могут находиться люди, после чего уже происходит оценка этих фрагментов снимков. Исследования показали, что при предварительном выделении областей интереса, точность нахождения целевых объектов повышается по сравнению с результатами без использования карты областей интереса.

Работа [5] описывает реальный пример успешного применения БВС для поиска пропавших людей. Поиск выполнялся с использованием программы *SARUAV*^{URL}, в основе которой лежат технологии сверточных нейронных сетей (CHC).

Исследование [6] нацелено на создание автономного БВС, способного самостоятельно выполнять обнаружение различных объектов, в том числе и людей. В качестве бортовой системы анализа изображений авторы использовали микрокомпьютер *Raspberry Pi*^{URL} со вспомогательным модулем *Coral Edge TPU*^{URL}. В результате удалось достигнуть высокой точности обнаружения автомобилей и людей, равной 96%.

Статья [7] посвящена разработке метода обнаружения препятствий на пути движения БВС. Для достижения своих целей авторы использовали нейронную сеть архитектуры YOLOv8 [8] и набор данных, содержащий несколько классов объектов: «Грузовик», «Человек», «Автобус», «Автомобиль», «Дерево», «Здание». Несмотря на то, что представленный метод должен работать непосредственно на БВС, его тестирование авторы

проводили на стационарном компьютере. В результате достигнута точность по mAP^{50} от 75% до 89% в зависимости от количества эпох обучения.

Работа [9] описывает создание метода обнаружения замаскированных объектов с использованием полуконтролируемой нейронной сети на базе SINet [10]. Обучение сети происходило на наборе данных *COD10K*^{URL}, который содержал несколько классов замаскированных объектов. Результатом стала точность обнаружения 66% по индексу комплексной оценки.

В работе [11] представлен метод детектирования взлётно-посадочной площадки (ВПП) для БВС, а также людей, находящихся в непосредственной близости от площадки. В свою очередь, используемый БВС оснащён микрокомпьютером с поддержкой *CUDA*^{URL} *Jetson Nano*^{URL} или *Jetson Xavier NX*^{URL}. В своём исследовании авторы проводят тестирование различных нейросетевых архитектур для получения точности детектирования и скорости работы, чтобы подтвердить возможность использования разработанного метода в реальном времени. В итоге с использованием сети *MobileNet V3*^{URL} получены точность определения людей в 76%, точность обнаружения ВПП в 95% и общая точность в 98% по метрике F1.

Исследование [12] рассматривает поиск целевых объектов (людей) при помощи нейронной сети YOLOv8, которая была предварительно обучена на наборе данных *COCO*^{URL}. Также авторы используют набор данных *SARD*^{URL} для оптимизации модели обнаружения людей на аэрофотоснимках, снятых при помощи БВС. В ходе исследования рассмотрены различные нейросетевые модели семейства YOLOv8, доработанные для набора SARD. В результате показано, что модель YOLOv8x достигает наивысших результатов в решении поставленных задач.

Материал [13] содержит обзор различных приложений машинного обучения для БВС, предназначенных для обнаружения и отслеживания объектов и людей. В нём также рассматриваются различные задачи, такие как построение маршрута, навигация, анализ изображений. В качестве нейросетевой базы используется архитектура семейства YOLO.

В статье [14] описана система из двух БВС, предназначенная для локализации целей в неоднородной среде. Система представляла собой платформу, которая позволяет управлять полетом двух БВС и выполнять задачи идентификации и локализации персонала на земле. Решение объединяет данные полета двух БВС с координатами пикселей. Использована модель с архитектурой YOLOv7 [15], оптимизированная для обнаружения целей.

Кроме исследований направленных на непосредственный поиск целевых объектов (людей) на изображениях, полученных с БВС, существуют работы в области улучшения быстродействия моделей семейства YOLO.

Например, статье [16] представлена модель PS-YOLO – облегченная сеть для обнаружения объектов на изображениях с БВС, основанная на актуальной версии *YOLOv11s*^{[URL](#)}. Метод содержал три ключевых компонента: использование частичной свёртки (partial convolution); новый модуль FasterBIFFPN; а также детектор GSCD с разделённой сверткой и функцией потерь NWDLoss. Эксперименты на датасете VisDrone показали, что PS-YOLO превосходит базовую модель по точности (увеличение mAP^{50} на 1,3%), быстродействию (прирост FPS на 26,7%) и уменьшению размера модели (сокращение параметров на 41,3%).

Архитектура CF-YOLO, представленная в статье [17] является усовершенствованным детектором мелких объектов на основе модели YOLOv11, адаптированным для задач обнаружения с БВС. В рамках исследования разработаны модель, сохраняющая детали мелких объектов при многомасштабном слиянии признаков и двухступенчатая стратегия слияния для пространственной калибровки и адаптивного взвешивания и интеграции признаков. Эксперименты показали улучшение точности обнаружения мелких объектов – прирост mAP^{50} до 12,7% по сравнению с базовой моделью.

Задача поиска мелких объектов решена в работе [18], где был предложен легковесный алгоритм YOLO-LSM для детекции мелких объектов на видео с БВС, учитывающий их ограничения вычислительных ресурсов. Основными нововведениями стали слой для мелких целей, модуль многомасштабной свёртки, функцию потерь Focaler-IoU [19] и механизм деформируемого внимания для оптимизации слияния признаков. Эксперименты на датасетах VisDrone и TinyPerson [20] показали сохранение точности при сокращении числа параметров примерно на 72% по сравнению с базовой моделью YOLOv5 [21].

Исследование [22] описало подход к обнаружению людей с БВС на основе параллельного использования двух трансформерных нейросетевых моделей – Detection Transformer (DETR) [23] и Vision Transformer (ViT) [24]. DETR точно локализовывал объекты в плотных городских сценах, тогда как ViT выявлял мелкие и частично скрытые объекты. Для объединения результатов обеих моделей предложен динамический механизм слияния, адаптивно взвешивающий предсказания в зависимости от контекста сцены и характеристик объектов, улучшая точность и устойчивость системы. Эксперименты на датасете VisDrone показали, что объединённая модель превосходит по mAP^{50} , точности и полноте каждую из двух моделей, демонстрируя прирост точности до 15%.

Из вышеперечисленного можно сделать вывод: использование современных нейросетевых технологий в задачах поиска людей на снимках

с БВС находит все более широкое применение в социально значимых областях и не теряет актуальности. Более того, развитие различных нейросетевых архитектур обеспечивает постепенное увеличение точности детектирования людей на снимках. Тем не менее, обнаружение людей вне типичной для них среды обитания до сих пор остаётся задачей, требующей решения.

В настоящей статье описывается возможность применения глубоких нейронных сетей для поиска людей на снимках, полученных с БВС в малозаселённой местности. Выполнена постановка задачи поиска людей на снимках с БВС. Описывается способ применения нейросетевой архитектуры YOLO в решении поставленных задач, в результате которого получена точность детектирования людей на изображениях, равная 90% по mAP^{50} .

1. Цели и задачи исследования

Использование БВС для поиска людей в малонаселённой местности может позволить оперативно получать данные с большой площади поверхности за короткие сроки. Однако задача поиска людей с высоты имеет свои особенности. К ним можно отнести ракурс съёмки для обнаружения людей. В зависимости от высоты полета, силуэт человека может иметь нетипичный вид, где невозможно выделить особые формы, такие как руки, ноги, голова. Помимо этого, человек (особенно в спящем состоянии или в состоянии истощения) может принять нетипичную в повседневности позу, например, так называемую позу «эмбриона». Другой особенностью является неоднородность фона на снимках и наличие всяческих помех, таких как ветки деревьев, различный мусор и особенности рельефа.

Цель настоящего исследования заключается в применении технологии искусственных нейронных сетей (ИНС) глубокого обучения для поиска людей. Среди задач можно выделить следующие:

- Выбор актуальной модели ИНС.
- Подготовка обучающего набора данных.
- Обучение выбранной модели ИНС.
- Инференс модели и анализ полученных результатов.

Предложенный метод поиска/детекции целевых объектов (людей) на снимках с БВС будет считаться работоспособным, а основная цель настоящего исследования достигнутой, если при интересе модели будет получена точность более 80% по метрике $F1_{URL}$ и более 75% по mAP^{50}_{URL} .

2. Используемая нейросетевая модель

Для решения поставленных задач была использована нейросетевая архитектура YOLO. Оригинальная версия YOLO (You Only Look Once) была создана Джозефом Редмоном в специальной среде под названием *Darknet*^{[URL](#)}.

YOLO представляет собой одноступенчатый детектор, выполняющий как обнаружение, так и классификацию объектов за один проход сети. Рассматривая задачу обнаружения как однократный регрессионный подход для определения габаритных прямоугольников объектов интереса, модели архитектуры YOLO работают быстрее и содержат сравнительно малое количество параметров, что ускоряет их обучение и упрощает развертывание, особенно на периферийных устройствах с ограниченными вычислительными возможностями.

Сейчас одной из наиболее актуальных версий этой архитектуры является *YOLOv11*^{[URL](#)}. Однако версия *YOLOv8*^{[URL](#)} также остаётся весьма востребованной. При *сравнении*^{[URL](#)} представленных версий архитектуры YOLO, версия 8 обладает более высокой надёжностью, развитой экосистемой и поддержкой сообщества, а также высокой универсальностью. В то же время версия 11 обеспечивает лучшую точность и производительность и не уступает в универсальности использования.

Исходя из этого, был сделан выбор в пользу версии YOLOv11. Готовая к обучению модель реализована в библиотеке *Ultralytics*^{[URL](#)} языка программирования Python. В настоящей статье использовалась версия 8.3.213 данной библиотеки.

Существует несколько вариантов архитектуры YOLOv11. В зависимости от количества гиперпараметров они подразделяются на следующие классы: n – nano, s – small, m – medium, l – large, и x – extra large. Для рассматриваемой модели также существуют авторские решения такие как: PS-YOLO, CF-YOLO и YOLO-LSM. Модель PS-YOLO основана на YOLOv11m и превосходит её по точности в среднем на 1.5%, и по быстродействию на 25% на наборе данных VisDrone. CF-YOLO показывает прирост точности от 3.5% до 12.7% в зависимости от набора данных, по сравнению с базовой моделью YOLOv11n. Вариант YOLO-LSM, основанный на YOLOv11s, превосходит базовую модель по точности на 3.5%, при этом используя примерно на 72% меньше гиперпараметров, чем оригинальная модель.

Несмотря на то, что каждая из представленных авторских моделей превосходит базовую YOLOv11, оценить качество их работы (точность, быстродействие и т.д.) в контексте решаемых задач можно лишь косвенно,

так как получить актуальный исходный код моделей не представляется возможным. Исходя из этого, выбор модели YOLOv11m был наиболее верным с точки зрения решаемых задач, так как данная версия модели обладает наиболее сбалансированными характеристиками, такими как точность, быстродействие, количество гиперпараметров и производительность.

Поскольку модели семейства YOLOv11 изначально *предварительно обучены*^(URL) на наборе данных COCO, было принято решение провести предварительный инференс предобученной модели YOLOv11m, которая ещё не была обучена на наборе LaDD. Целью его проведения было определение начальных значений точности детектирования целевых объектов, а также установление факта необходимости обучения модели. Процедура инференса модели описана в п.5 настоящей статьи. В результате было достигнуто значение 0.471 (47.1%) по метрике F1 и 0.306 (30.6%) по метрике mAP^{50} . Полученные показатели свидетельствуют о необходимости дополнительного обучения предобученной модели YOLOv11m для решения поставленных задач.

3. Обучающий набор данных

Для обучения выбранной нейросетевой модели был использован набор данных *Lacmus Drone Dataset (LaDD)*^(URL). Данный набор содержит изображения, снятые с использованием БВС, где запечатлены люди, находящиеся в малозаселённой местности. Всего в составе этого набора насчитывается 1624 изображения со средним размером 4000 на 3000 пикселей, к каждому из которых прилагается свой xml-файл с разметкой. Размета была выполнена только для одного класса объектов – Pedestrian. Все изображения набора распределены на три выборки, обучающую (train), тестовую (test) и валидационную (val) в количестве 1220, 202 и 202 изображения соответственно.

3.1. Подготовка набора данных

Используемый набор данных LaDD уже содержал необходимые файлы разметки. Однако, для обучения ИНС семейства YOLO разметка должна быть переведена в соответствующий *формат*^(URL). В процессе подготовки набора данных к каждому из изображений набора был сформирован текстовый файл, содержащий описание целевых объектов, находящихся на изображении. Описание содержало такие данные как: класс объекта в числовом эквиваленте, а также нормализованные координаты его центра, ширину и высоту.

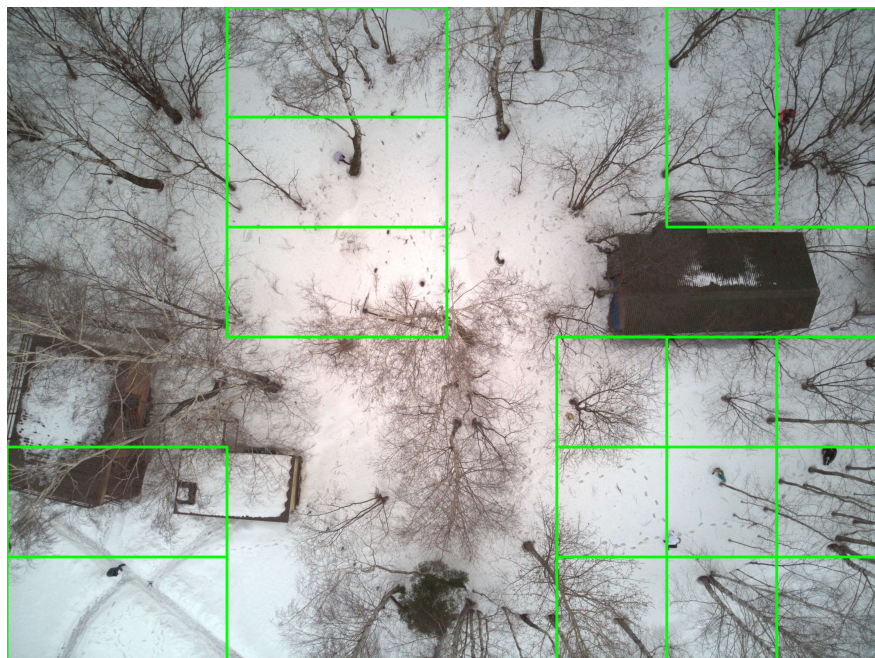
3.2. Адаптация набора данных

На основании того, что выбранная модель YOLOv11m *оптимизирована*^{URL} под входные изображения с размером 1024 на 1024 пикселя, было принято решение по адаптации используемого набора данных, целью которой являлось формирование обучающей и валидационной выборки изображений с требуемым разрешением из исходных изображений.

Процесс адаптации заключался в извлечении фрагментов размером 1024 на 1024 пикселя, содержащих целевые объекты, из исходного изображения. Процедуре адаптации подверглась только обучающая выборка исходного набора данных. Алгоритм адаптации набора данных состоял из следующих шагов:

- (1) Сбор и формирование списка из координат габаритных прямоугольников целевых объектов (люди, класс «Pedestrian») с использованием данных разметки набора.
- (2) Формирование списка из координат сегментов оригинального изображения. В данном случае сегменты имели размер 1024 на 1024 пикселя. Координаты сегментов фиксировались с применением метода скользящего окна с шагом, равным половине размера сегмента по вертикали и горизонтали.
- (3) Проверка размера сегментов. Данная проверка выполнялась чтобы исключить сегменты, находящиеся на границе изображения. Если размер сегмента не превышал $3/4$ от установленного размера (в данном случае это 768 пикселей) по горизонтали и/или вертикали, то такой сегмент не учитывался. Если же размер сегмента был меньше требуемого, но больше его $3/4$, то недостающая часть сегмента восполнялась необходимым числом пикселей по ширине и высоте методом копирования краевых пикселей.
- (4) Проверка на принадлежность целевых объектов (координат их габаритных прямоугольников) к сегментам изображения и извлечение сегментов. Если габаритный прямоугольник хотя бы одного целевого объекта полностью входил в сегмент, то этот сегмент извлекается из оригинального изображения. Затем происходил перерасчёт координат всех габаритных прямоугольников целевых объектов, принадлежащих данному сегменту, в его координатное пространство.
- (5) Формирование аннотационных файлов для извлечённых сегментов оригинального изображения. На данном этапе происходило формирование аннотационных текстовых файлов для каждого из извлечённых сегментов в формате Yolo. Для каждого целевого объекта, изображённого на извлечённом фрагменте, использовались пересчитанные координаты его габаритного прямоугольника, которые были получены на предыдущем шаге.

Весь процесс адаптации набора данных был реализован на языке Python с использованием библиотеки *OpenCV*^{URL}. На рисунке 1 показан пример исходного изображения и извлечённых из него фрагментов.



(а) исходное изображение: зелёные прямоугольники – места извлечения



(б) извлечённые фрагменты: красные прямоугольники – целевые объекты

Рисунок 1. извлечение фрагментов с целевыми объектами

В результате было получено 7360 изображений размером 1024 на 1024 пикселя, которые, в свою очередь, были разделены на обучающую и валидационную выборки в соотношении 75/25.

Процесс адаптации набора данных с последующим разделением на обучающую и валидационную выборки не исключает возможности попадания одного и того же целевого объекта одновременно в обе выборки. Тем не менее, хоть целевой объект и может быть одним и тем же, его положение на изображениях всегда будет разным, что принципиально схоже с *мозаичной аугментацией данных*^{URL}, которая применяется для того, чтобы модели обнаружения объектов лучше справлялись с различными масштабами и перемещениями объектов.

4. Обучение модели

Обучение моделей происходило посредством специально написанного скрипта на языке Python. Первоначально было проведено предварительное обучение на протяжении 100 эпох, которое позволило оценить рассматриваемые модели и выбрать лучший вариант по показателю точности по метрике mAP^{50} . Среди рассмотренных моделей была как оригинальная YOLOv11m предобученная на наборе данных COCO, так и «чистая» модель YOLOv11m (YOLOv11m_y) сформированная из специального конфигурационного yaml-файла. В качестве эксперимента для предварительного обучения использовались оба набора данных: оригинальный (LaDD_orig) и адаптированный (LaDD_adapt).

Точность предварительного обучения моделей оценивалась по метрикам mAP^{50} , $Precision$ ^{URL} и $Recall$ ^{URL}, значения которых представлены в таблице 1.

Таблица 1. Точность предварительного обучения моделей по метрикам Precision, Recall и mAP^{50}

Набор	Модель	Precision	Recall	mAP^{50}
LaDD_orig	YOLOv11m	0.938	0.884	0.936
LaDD_adapt	YOLOv11m	0.958	0.914	0.966
LaDD_orig	YOLOv11m_y	0.916	0.875	0.902
LaDD_adapt	YOLOv11m_y	0.940	0.898	0.956

Анализ результатов предварительного обучения моделей YOLOv11m и YOLOv11m_u показал прирост точности от 3 до 5% (по mAP^{50}) при использовании адаптированного варианта набора данных LaDD по сравнению с его оригинальной версией. Также стоит отметить, что используемые модели показали очень близкие значения точности при обучении на адаптированном наборе данных LaDD. Разница между самой низкой точностью в 0.956 (95.6%), полученной при использовании модели YOLOv11m_u, и самой высокой точностью в 0.966 (96.6%), достигнутой моделью YOLOv11m, составила всего 1%. Таким образом, предобученная модель с загруженными весами показала себя лучше, чем модель со случайной инициализацией весов. Учитывая вышеизложенное, было принято решение о проведении полноценного обучения предобученной модели YOLOv11m, так как она показала наибольшую точность по mAP^{50} .

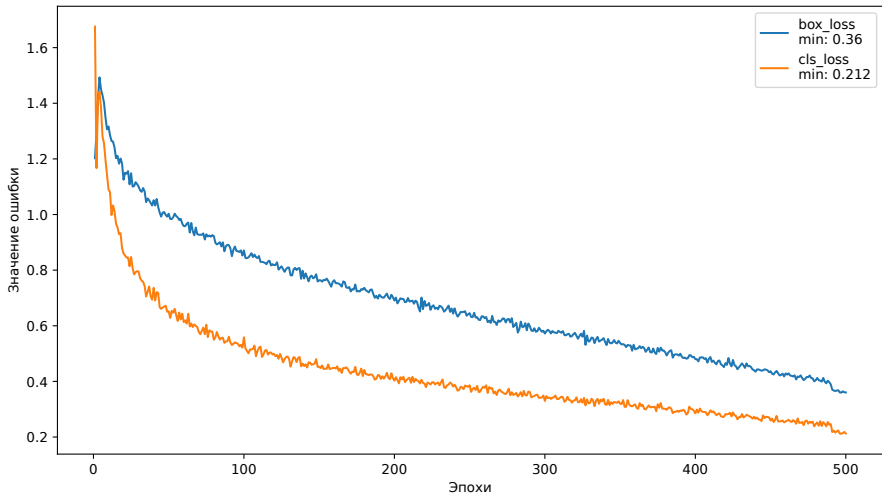
Ради чистоты эксперимента полученные ранее результаты и веса предварительного обучения не учитывались при проведении основного/полноценного обучения, то есть модель обучалась заново. При полноценном обучении модели была применена стратегия ранней остановки обучения – *Early Stopping*^{[URL](#)}. Использование данного подхода позволяет отслеживать качество валидации модели и остановить обучение, как только модель перестанет улучшаться. Это позволяет сэкономить вычислительные ресурсы и предотвратить *переобучение*^{[URL](#)}. Стратегия ранней остановки заключается в установке параметра «терпения», который определяет количество эпох, в течение которых необходимо ожидать улучшения показателей валидации, прежде чем остановить обучение.

В данном случае параметр «терпения» был установлен на 100 эпох. Параметр скорости обучения (learning rate) был установлен на 0.01. Из-за аппаратных ограничений размер пакета (batch size) был уменьшен до 4.

В процессе обучения было выяснено, что при достижении 500 эпох модель больше не показывала значимых улучшений в точности детектирования на валидационной выборке по метрике mAP^{50} , что спустя 100 эпох вызвало срабатывание механизма ранней остановки обучения.

Статистика обучения модели записывалась в специальный csv-файл. Основные параметры, такие как ошибка и точность обучения, представлены в виде графиков на рисунке 2а и рисунке 2б.

В результате обучения модели YOLOv11m с предварительно загруженными весами на адаптированном наборе данных LaDD удалось достичь точности в 0.987 (98.7%) по mAP^{50} .



(а) Графики зависимости от эпох обучения ошибок

box_loss определения габаритных прямоугольников целевых объектов и
cls_loss классификации габаритных прямоугольников целевых объектов

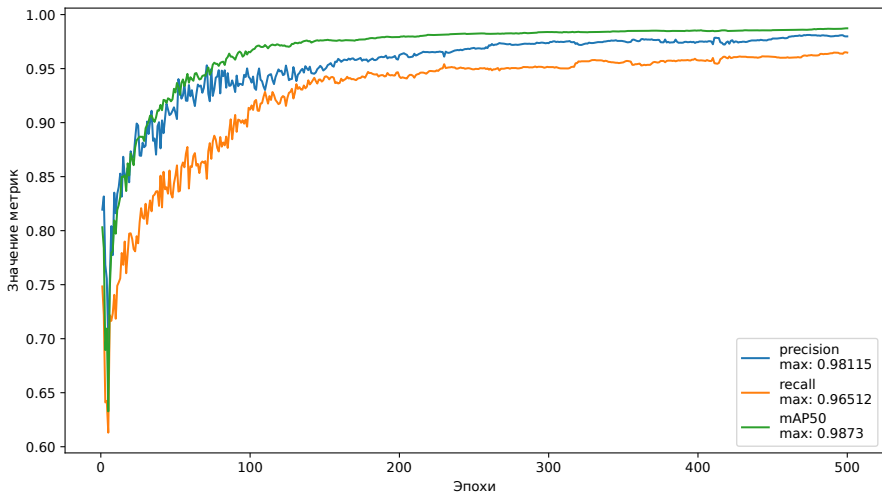
(б) График изменения точности обучения по метрикам Precision, Recall и mAP^{50}

РИСУНОК 2. График зависимости от эпох ошибки и точности обучения

5. Инференс модели и анализ результатов

Инференс модели проводился с помощью скрипта на языке Python с применением библиотеки *SAHI (Slicing Aided Hyper Inference)*^{[URL](#)}. SAHI – это специальная библиотека, разработанная для оптимизации алгоритмов обнаружения объектов для крупномасштабных изображений с высоким разрешением. Её основная функциональность заключается в разделении изображений на управляемые фрагменты, запуске обнаружения объектов на каждом фрагменте и последующем объединении результатов. За визуализацию результатов и оценку точности отвечала библиотека *Supervision*^{[URL](#)}.

Использовалась тестовая выборка оригинального набора данных LaDD, то есть та, что содержала изображения большого размера - 4000 на 3000 пикселей. Размер слайса (фрагмента оригинального изображения) составлял 1024 на 1024 пикселя. Значение перекрытия было установлено на половину размера слайса, то есть 512 пикселей. Инференс производился с использованием графического процессора (GPU). Визуализация результатов детекции целевых объектов (людей) показана на рисунках 3 и 4.

Заключение

Задача поиска людей на малонаселённой местности имеет важное социальное значение. В настоящей статье был описан подход к решению поставленной задачи с использованием нейросетевой модели YOLOv11, в частности YOLOv11m. Исследования показали, что задача поиска людей на снимках, полученных с БВС, может быть решена с использованием ИНС. Также нейросетевой анализ может стать серьёзным подспорьем для разработки и улучшения алгоритмов поиска людей при проведении поисково-спасательных операций.

Средняя точность модели, полученная в ходе инференса, составила 0.895 (89.5%) по F1 и 0.901 (90.1%) по mAP^{50} .

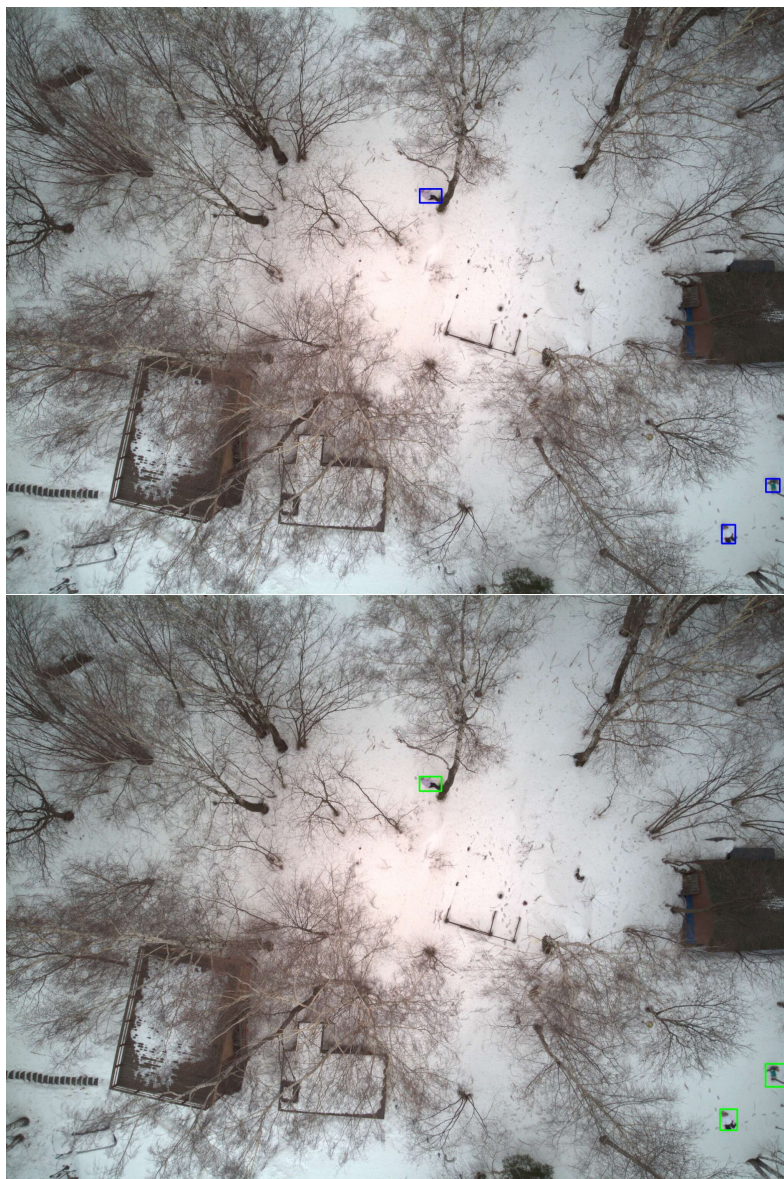


Рисунок 3. Оригинальное изображение с разметкой и визуализация предсказаний модели



Рисунок 4. Другое оригинальное изображение с разметкой и визуализация предсказаний модели

В рамках проведённого исследования продемонстрирована разница в точности обучения при использовании оригинального набора данных и адаптированного под размер входа набора. Предварительное обучение модели показало, что итоговая точность модели при использовании адаптированного набора данных выше как минимум на 3%, чем при обучении на оригинальном наборе. Таким образом, при полноценном обучении удалось достичь точности, равной 0.987 (98.7%) по mAP^{50} .

Однако используемый в настоящем исследовании алгоритм адаптации набора данных ведёт к кратному увеличению количества содержащихся в нём изображений. К примеру, в данном случае количество изображений в обучающей выборке увеличилось примерно в 4.5 раза. Эта ситуация имеет одновременно как положительные, так и отрицательные моменты. Как известно, что чем больше и разнообразней обучающая выборка, тем более адекватно и контролируемо будет происходить обучение модели. Но с увеличением объёма обучающей выборки также расчёт общее время обучения модели. Тем не менее, оптимизация скорости обучения не являлась одной из решаемых задач настоящего исследования.

В результате инференс модели показал высокие показатели точности, а именно 0.895 (89.5%) по F1 и 0.901 (90.1%) по mAP^{50} , что позволяет подтвердить работоспособность представленного в настоящем исследовании метода поиска людей в малонаселённой местности по снимкам с БВС.

Список использованных источников

- [1] Akshatha K.R., Karunakar A.K., Satish S.B., Phani P.K., Chinmay V.D., Johnson D. *Manipal-UAV person detection dataset: A step towards benchmarking dataset and algorithms for small object detection* // ISPRS Journal of Photogrammetry and Remote Sensing.– 2023.– No. 195.– Pp. 77-89.  [↑218](#)
- [2] Li M., Yan N. *IPD-YOLO: Person detection in infrared images from UAV perspective based on improved YOLO11* // Digital Signal Processing.– 2026.– No. 168.– id. 105469.  [↑218](#)
- [3] Paulin G., Sambolek S., Ivasic-Kos M. *Application of raycast method for person geolocalization and distance determination using UAV images in Real-World land search and rescue scenarios* // Expert Systems with Applications.– 2024.– No. 237.– id. 121495.  [↑219](#)
- [4] Gotovac S., Zelenika D., Marušić Ž., Božić-Štulić D. *Visual-based person detection for search-and-rescue with UAS: Humans vs. machine learning algorithm* // Remote Sensing.– 2020.– Vol. 12.– No. 20.– id. 3295.– 25 pp.  [↑219](#)

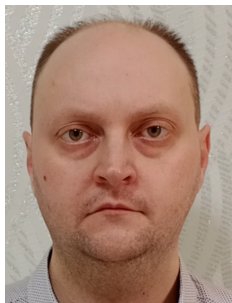
- [5] Niedzielski T., Jurecka M., Miziński B., Pawul W., Motyl T. *First successful rescue of a lost person using the human detection system: A case study from Beskid Niski (SE Poland)* // Remote Sensing.– 2021.– Vol. **13**.– No. 23.– id. 4903.– 18 pp.  ^{↑219}
- [6] Kirac E., Ozbek S. *Deep learning based object detection with unmanned aerial vehicle equipped with embedded system* // Journal of Aviation.– 2024.– Vol. **8**.– No. 1.– Pp. 15–25.  ^{↑219}
- [7] Rahman S., Rony J. H., Uddin J., Samad M. A. *Real-time obstacle detection with YOLOv8 in a WSN using UAV aerial photography* // Journal of Imaging.– 2023.– Vol. **9**.– No. 10.– id. 216.– 17 pp.  ^{↑219}
- [8] Yaseen M. *What is YOLOv8: an in-depth exploration of the internal features of the next-generation object detector*.– 2024.– 10 pp.  2408.15857  ^{↑219}
- [9] Yang L., Cong-qing W., Yong-jun Zh. *Camouflaged people detection based on a semi-supervised search identification network* // Defence Technology.– March 2023.– Vol. **21**.– Pp. 176–183.  ^{↑220}
- [10] Fan D. P., Ji G. P., Sun G., Cheng M. M., Shen J., Shao L. *Camouflaged object detection* // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Seattle, WA, USA, 13–19 June 2020).– IEEE.– 2020.– ISBN 978-1-7281-7168-5.– Pp. 2774–2784.  ^{↑220}
- [11] Pieczyński D., Ptak B., Kraft M., Piechocki M., Aszkowski P. *A fast, lightweight deep learning vision pipeline for autonomous UAV landing support with added robustness* // Engineering Applications of Artificial Intelligence.– May 2024.– Vol. **131**.– id. 107864.– 13 pp.  ^{↑220}
- [12] Sambolek S., Ivasic-Kos M. *Person detection and geolocation estimation in UAV aerial images: an experimental approach* // Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods – ICPRAM.– V. 1 (Rome, Italy, February 24–26, 2024), Lecture Notes in Computer Science.– vol. **15568**, Cham: Springer.– 2024.– ISBN 978-989-758-684-2.– Pp. 785–792.  ^{↑220}
- [13] Saeed R. A., Nahla A. N. E. *Applications ML in UAVs-based detecting and tracking objects and people* // Applications of Machine Learning in UAV Networks.– IGI Global.– 2024.– ISBN 9789369305782.– Pp. 79–96.  ^{↑220}
- [14] Chang R., Pan A., Yu K., Zhou C., Yang Y. *A dual UAV cooperative positioning system with advanced target detection and localization* // IEEE Access.– 2024.– Vol. **12**.– Pp. 43235–43244.  ^{↑220}
- [15] Wang C., Bochkovskiy A., Liao H.-Y. M. *YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors* // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Vancouver, BC, Canada, 17–24 June 2023).– IEEE.– ISBN 979-8-3503-0129-8.– Pp. 7464–7475; – 2022.   ^{↑220}
- [16] Zhong H., Zhang Y., Shi Z., Zhang Y., Zhao L. *PS-YOLO: a lighter and faster network for UAV object detection* // Remote Sensing.– 2025.– Vol. **17**.– No. 9.– id. 1641.– 28 pp.  ^{↑221}

- [17] Wang C., Han Y., Yang Ch., Wu M., Chen Z., Yun L., Jin X. *CF-YOLO for small target detection in drone imagery based on YOLOv11 algorithm* // Scientific Reports.– 2025.– Vol. **15**.– id. 16741.– 18 pp.  [↑221](#)
- [18] Wu C., Cai C., Xiao F., Wang J., Guo Y., Ma L. *YOLO-LSM: a lightweight UAV target detection algorithm based on shallow and multiscale information learning* // Information.– 2025.– Vol. **16**.– No. 5.– id. 393.– 21 pp.  [↑221](#)
- [19] Zhang H., Zhang S. *Focaler-IoU: more focused intersection over union loss*.– 2024.– 4 pp. [arXiv](#)  2401.10525  [↑221](#)
- [20] Yu X., Gong Y., Jiang N., Ye Q., Han Z. *Scale match for tiny person detection* // 2020 IEEE Winter Conference on Applications of Computer Vision (WACV) (Snowmass, CO, USA, 01–05 March 2020).– IEEE.– 2020.– ISBN 978-1-7281-6553-0.– Pp. 1246–1254.  [↑221](#)
- [21] Khanam R., Hussain M. *What is YOLOv5: A deep look into the internal features of the popular object detector*.– 2024.– 8 pp. [arXiv](#)  2407.20892  [↑221](#)
- [22] Stan A.-S., Popescu D., Ichim L. *Person detection from UAV based on a dual transformer approach* // GISTAM 2025 — 11th International Conference on Geographical Information Systems Theory, Applications and Management.– V. 1 (Porto, Portugal, 1–3 April 2025).– SciTePress.– 2025.– ISBN 978-989-758-741-2.– Pp. 95–102.  [↑221](#)
- [23] Zhao Y., Lv W., Xu S., Wei J., Wang G., Dang Q., Liu Y., Chen J. *DETRs beat YOLOs on real-time object detection* // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Seattle, WA, USA, 16–22 June 2024).– IEEE.– 2023.– ISBN 979-8-3503-5300-6.– Pp. 16965–16974.– 14 pp.   [arXiv](#)  2304.08069 [↑221](#)
- [24] Fu Z. *Vision transformer: Vit and its derivatives*.– 2022.– 10 pp. [arXiv](#)  2205.11239  [↑221](#)

Поступила в редакцию	28.10.2025;
одобрена после рецензирования	09.11.2025;
принята к публикации	09.11.2025;
опубликована онлайн	12.11.2025.

Рекомендовал к публикации

к.т.н. В. П. Фраленко

Информация об авторах:**Александр Владимирович Смирнов**

Младший научный сотрудник Лаборатории методов обработки и анализа изображений, Институт Программных Систем имени А. К. Айламазяна РАН. Научные интересы: компьютерное зрение; нейронные сети; робототехника; автоматизация и управление



0000-0002-7104-1462

e-mail: asmirnov_1991@mail.ru**Игорь Петрович Тищенко**

Кандидат технических наук, ИО директора, Институт Программных Систем имени А. К. Айламазяна РАН. Научные интересы: компьютерное зрение; нейронные сети; робототехника; автоматизация и управление



0000-0002-0369-0524

e-mail: igor.p.tishchenko@yandex.ru**Сергей Александрович Лазарев**

Младший научный сотрудник Исследовательского центра искусственного интеллекта (ИЦИИ), Институт Программных Систем имени А. К. Айламазяна РАН. Научные интересы: искусственный интеллект; беспилотные системы



0009-0007-0668-289X

e-mail: LazarevSerg76@yandex.ru

Вклад авторов: *А. В. Смирнов* – 80% (методология, программное обеспечение, валидация, формальный анализ, расследование, сбор материала, написание черновой версии, доработка и редактирование, визуализация); *И. П. Тищенко* – 10% (идея, курирование данных, администрирование, финансирование); *С. А. Лазарев* – 10% (доработка и редактирование).

Декларация об отсутствии личной заинтересованности: *благополучие авторов не зависит от результатов исследования.*



Applying the YOLOv11 model and the adapted LaDD dataset to people detection in sparsely populated areas

Alexander Vladimirovich **Smirnov**¹, Igor Petrovich **Tishchenko**²,
Sergey Alexandrovich **Lazarev**³

¹⁻³ Ailamazyan Program Systems Institute of RAS, Ves'kovo, Russia

Abstract. This study is aimed at developing a neural network method for detecting people in sparsely populated areas using images obtained from an unmanned aerial vehicle (UAV). The YOLOv11m architecture model was used as a neural network detector. As part of the study, an adaptation algorithm for the LaDD training dataset was developed and applied. Experiments were conducted to preliminary train the model on the original and adapted datasets, which demonstrated the advisability of using the adapted dataset. The final accuracy of the model during training reached 98.7% by metric mAP^{50} . Model inference showed a detection accuracy of 0.895 (89.5%) by metric F1 and 0.901 (90.1%) by metric mAP^{50} , which confirms the workability of the presented method. (*In Russian*).

Key words and phrases: Image analysis, people detection, UAV imagery, YOLOv11, neural networks, dataset, adaptation

2020 *Mathematics Subject Classification:* 68T10; 68T45

For citation: Alexander V. Smirnov, Igor P. Tishchenko, Sergey A. Lazarev. *Applying the YOLOv11 model and the adapted LaDD dataset to people detection in sparsely populated areas.* Program Systems: Theory and Applications, 2025, **16**:4(67), pp. 217–240. (*In Russ.*) https://psta.psir.ru/read/psta2025_4_217-240.pdf

References

- [1] K.R. Akshatha, A.K. Karunakar, S.B. Satish, P.K. Phani, V.D. Chinmay, D. Johnson. “Manipal-UAV person detection dataset: A step towards benchmarking dataset and algorithms for small object detection”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, no. 195, pp. 77-89. [doi](#)
- [2] M. Li, N. Yan. “IPD-YOLO: Person detection in infrared images from UAV perspective based on improved YOLO11”, *Digital Signal Processing*, 2026, no. 168, id. 105469. [doi](#)
- [3] G. Paulin, S. Sambolek, M. Ivasic-Kos. “Application of raycast method for person geolocalization and distance determination using UAV images in Real-World land search and rescue scenarios”, *Expert Systems with Applications*, 2024, no. 237, id. 121495. [doi](#)
- [4] S. Gotovac, D. Zelenika, Ž. Marušić, D. Božić-Štulić. “Visual-based person detection for search-and-rescue with UAS: Humans vs. machine learning algorithm”, *Remote Sensing*, **12**:20 (2020), id. 3295, 25 pp. [doi](#)
- [5] T. Niedzielski, M. Jurecka, B. Miziński, W. Pawul, T. Motyl. “First successful rescue of a lost person using the human detection system: A case study from Beskid Niski (SE Poland)”, *Remote Sensing*, **13**:23 (2021), id. 4903, 18 pp. [doi](#)
- [6] E. Kirac, S. Ozbek. “Deep learning based object detection with unmanned aerial vehicle equipped with embedded system”, *Journal of Aviation*, **8**:1 (2024), pp. 15–25. [doi](#)
- [7] S. Rahman, J. H. Rony, J. Uddin, M. A. Samad. “Real-time obstacle detection with YOLOv8 in a WSN using UAV aerial photography”, *Journal of Imaging*, **9**:10 (2023), id. 216, 17 pp. [doi](#)
- [8] M. Yaseen. *What is YOLOv8: an in-depth exploration of the internal features of the next-generation object detector*, 2024, 10 pp. [arXiv:2408.15857](#) [doi](#)
- [9] L. Yang, W. Cong-qing, Zh. Yong-jun. “Camouflaged people detection based on a semi-supervised search identification network”, *Defence Technology*, **21** (March 2023), pp. 176–183. [doi](#)
- [10] D. P. Fan, G. P. Ji, G. Sun, M. M. Cheng, J. Shen, L. Shao. “Camouflaged object detection”, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA, 13–19 June 2020), IEEE, 2020, ISBN 978-1-7281-7168-5, pp. 2774–2784. [doi](#)
- [11] D. Pieczyński, B. Ptak, M. Kraft, M. Piechocki, P. Aszkowski. “A fast, lightweight deep learning vision pipeline for autonomous UAV landing support with added robustness”, *Engineering Applications of Artificial Intelligence*, **131** (May 2024), id. 107864, 13 pp. [doi](#)

- [12] S. Sambolek, M. Ivasic-Kos. “Person detection and geolocation estimation in UAV aerial images: an experimental approach”, *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods – ICPRAM*. V. 1 (Rome, Italy, February 24–26, 2024), Lecture Notes in Computer Science, vol. **15568**, Springer, Cham, 2024, ISBN 978-989-758-684-2, pp. 785-792. 
- [13] R. A. Saeed, A. N. E. Nahla. “Applications ML in UAVs-based detecting and tracking objects and people”, *Applications of Machine Learning in UAV Networks*, IGI Global, 2024, ISBN 9798369305782, pp. 79–96. 
- [14] R. Chang, A. Pan, K. Yu, C. Zhou, Y. Yang. “A dual UAV cooperative positioning system with advanced target detection and localization”, *IEEE Access*, **12** (2024), pp. 43235–43244. 
- [15] C. Wang, A. Bochkovskiy, H-Y. M. Liao. “YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”, *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Vancouver, BC, Canada, 17–24 June 2023), IEEE, 2023, ISBN 979-8-3503-0129-8, pp. 7464–7475. 
- [16] H. Zhong, Y. Zhang, Z. Shi, Y. Zhang, L. Zhao. “PS-YOLO: a lighter and faster network for UAV object detection”, *Remote Sensing*, **17**:9 (2025), id. 1641, 28 pp. 
- [17] C. Wang, Y. Han, Ch. Yang, M. Wu, Z. Chen, L. Yun, X. Jin. “CF-YOLO for small target detection in drone imagery based on YOLOv11 algorithm”, *Scientific Reports*, **15** (2025), id. 16741, 18 pp. 
- [18] C. Wu, C. Cai, F. Xiao, J. Wang, Y. Guo, L. Ma. “YOLO-LSM: a lightweight UAV target detection algorithm based on shallow and multiscale information learning”, *Information*, **16**:5 (2025), id. 393, 21 pp. 
- [19] H. Zhang, S. Zhang. *Focaler-IoU: more focused intersection over union loss*, 2024, 4 pp. [arXiv:2401.10525](https://arxiv.org/abs/2401.10525) 
- [20] X. Yu, Y. Gong, N. Jiang, Q. Ye, Z. Han. “Scale match for tiny person detection”, *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Snowmass, CO, USA, 01–05 March 2020), IEEE, 2020, ISBN 978-1-7281-6553-0, pp. 1246–1254. 
- [21] R. Khanam, M. Hussain. *What is YOLOv5: A deep look into the internal features of the popular object detector*, 2024, 8 pp. [arXiv:2407.20892](https://arxiv.org/abs/2407.20892) 
- [22] A.-S. Stan, D. Popescu, L. Ichim. “Person detection from UAV based on a dual transformer approach”, *GISTAM 2025 — 11th International Conference on Geographical Information Systems Theory, Applications and Management*. V. 1 (Porto, Portugal, 1–3 April 2025), SciTePress, 2025, ISBN 978-989-758-741-2, pp. 95–102. 

- [23] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, J. Chen. “DETRs beat YOLOs on real-time object detection”, *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA, 16–22 June 2024), IEEE, 2023, ISBN 979-8-3503-5300-6, pp. 16965–16974, 14 pp.  [doi](https://doi.org/10.1109/CVPR4617.2024.2304.08069) [arXiv !\[\]\(1c108ad8fe9e61d7178ebe04ab721b96_img.jpg\) 2304.08069](https://arxiv.org/abs/2304.08069)
- [24] Z. Fu. *Vision transformer: Vit and its derivatives*, 2022, 10 pp. [arXiv !\[\]\(7dcf48fddd351e3ddacf30e5bc93fcd2_img.jpg\) 2205.11239](https://arxiv.org/abs/2205.11239) 