

---

## Алгоритм определения городских транспортных целей с применением метода мультимодального слияния на БПЛА

*Ван Цзяньюань, Чэнь Сяотун, Чжан Юэ, Ши Донхао, Чэнь Цзинбао*

*Институт космонавтики, Нанкинский университет авиации и астронавтики*

**Аннотация:** Использование небольших беспилотных летательных аппаратов (БПЛА) для таких городских целей, как автомобили, на малых высотах, со временем стало основным методом распознавания. Для решения таких назревших проблем, как невозможность работы монитора видимого диапазона в ночное время под влиянием освещения, низкая точность однорежимных сетей из-за нечёткости рамки целевых объектов в инфракрасном излучении при эксплуатации дрона, в данной статье предлагается мультимодальный алгоритм определения городских целей на основе объединения изображений и сети глубокого обучения на дроне. Во-первых, был создан набор данных мультимодального слияния на основе набора регистрационных данных из видимого - инфракрасного диапазона DUT-VTUAV и алгоритма слияния изображений TIF; во-вторых, была выбрана наиболее подходящая облегченная сеть YOLO v5n путём сравнения точности обнаружения, скорости и количества параметров существующих сетей серии YOLO для БПЛА; наконец, был создан алгоритм определения целей посредством объединения алгоритма слияния изображений и модели определения объектов. Сравнительные эксперименты, проведённые на основе наборов данных об автомобилях, показывают, что алгоритм, предложенный в этой статье, значительно повышает точность работы по сравнению с одномодальным методом, достигая mAP 99,6%, позволяя завершить слияние изображений, полученных, исходя из диапазона визуального наблюдения и инфракрасного излучения, а также улучшает процесс определения за 0,3 секунды, демонстрируя высокую производительность в режиме реального времени.

**Ключевые слова:** определение объектов; YOLO; мультимодальное слияние; слияние данных; алгоритм TIF.

### Введение

С непрерывным развитием социальной экономики использование транспортных средств в современной экономической деятельности становится всё более частым. Заторы и непредвиденные ситуации на городских дорогах значительно влияют на жизнь людей, особенно в ночное время: когда свет слаб и видимость уменьшена, вероятность дорожно-транспортных происшествий возрастает. Поэтому важно мониторить транспортные потоки в городских дорожных системах. Однако сложность дорожных реальностей вызывает определенные трудности для реального мониторинга городских транспортных средств на низких высотах над

городом [1]. В последнее время, благодаря своей высокой гибкости, малому размеру, легкости в эксплуатации и большому функциональному радиусу действия, БПЛА широко используются в гражданской сфере [2], например, в определении городских транспортных средств, в непрерывном их отслеживании в реальное время с помощью камеры БПЛА, что позволяет выполнять более гибкое и удобное низковысотное отслеживание целевых объектов.

Цель заключается в распознании интересующих объектов на изображениях или в видео, а также в измерении их позиций и размеров [3]. Популярны алгоритмы определения объектов разделяются на двухэтапный и одноэтапный [4]. С развитием технологий, одноэтапный алгоритм постепенно становится основным за счёт своей высокой скорости и легковесной модели без значительной потери точности [5]. Серия YOLO (You Only Look Once) в качестве аппарата одноэтапного определения имеет очевидные преимущества при отслеживании целей с помощью БПЛА, благодаря чему широко используется на практике.

YOLO рассматривает задачу определения объектов как задачу регрессии, что позволяет напрямую, конец в конец выводить данные о положении и виде целевого объекта для достижения более высокой скорости работы [6]. Серия YOLO уже сейчас породила новейший вариант YOLO v8. А в предыдущем, YOLO v2 [7] на основе YOLO, был добавлен начальный анкерный ящик (anchor boxes), с помощью метода k-means извлекается соответствующая информация, такая, как информация о положении объекта, для достижения большей точности. В YOLO v3 [8] была использована структура остаточной сети, включены многомасштабные функции, углублен анализ сети и улучшена способность модели измерения объектов. В YOLO v4 [9] применяется функция активации Mish и структура пространственного пирамидального пуллинга (SPP) на основе предыдущей сетевой структуры,

---

улучшая производительность измерения сети путём оптимизации структуры основной сети. YOLO v5 [10] в основном использует метод адаптивного пулинга признаков, чтобы повысить точность прогнозирования сети путем автоматической настройки размеров анкерных рамок. YOLO v7 [11] вводит репараметризацию модели в структуру своей сети на основе YOLO v5 и предлагает метод обучения для вспомогательных головок; с помощью увеличения затраты на обучение сети улучшает общую точность измерения сети без влияния на время обучения. YOLO v8 [12] заменяет все модули C3 в YOLO v5 на модули C2f для достижения большего облегчения веса, отказываясь от стратегии Anchor Base в предыдущих сериях, применяет концепцию Anchor Free [13], путём определения центральной точки объекта и его четырех границ, повышает скорость детектора.

Существующие модели определения объектов на основе глубокого анализа продемонстрировали отличную производительность на различных международных общедоступных наборах данных, но среды, откуда собраны эти данные, в основном представляют собой помещение или же открытую среду с хорошими условиями освещения, большинство этих данных - изображения в видимом свете, на которых края целевых объектов чёткие и цвета насыщенные, что приводит к высокой точности распознавания и измерения. В реальном процессе использования, на работу камеры видимого света часто влияют сильные лучи, отраженный свет и т.п. Особенно ночью, при условиях слабого освещения, камера практически не может делать чёткие фото [14]. А инфракрасная тепловизионная технология имеет преимущества ночного видения и нечувствительности к свету, однако низкое разрешение приводит к размытым контурам целевых объектов в изображениях и низкой точности распознавания разных видов объектов, что ограничивает сферу её применения.

---

В ответ на вышеуказанные проблемы, чтобы уменьшить влияние факторов окружающей среды на работу определения городских транспортных средств и повысить способность к анализу, в данной работе разработан алгоритм мультимодального слияния с использованием БПЛА. При данном алгоритме, с помощью взаимодополняемости мультимодальных данных, в диапазоне визуального наблюдения и инфракрасного излучения улучшается точность и устойчивость определения целевых объектов [15]. Во-первых, построен мультиспектральный датасет с визуально-инфракрасными изображениями на основе набора регистрационных данных из визуального и инфракрасного диапазона DUT-VTUAUV и алгоритма слияния изображений TIF; во-вторых, проведено всестороннее сравнение производительности моделей серии YOLO при анализе набора данных транспортных средств в данной статье, была выбрана одна из них, как наиболее подходящая модель сети – YOLO v5n с учётом комплексных факторов, таких, как размер модели, необходимая вычислительная мощность, точность распознавания и скорости; наконец, был разработан end-to-end алгоритм мультимодального слияния для распознавания и измерения объектов, основанный на алгоритме слияния изображений и YOLO v5n. Экспериментальные результаты показывают, что предложенный в настоящей статье алгоритм слияния мультимодальных данных эффективно улучшает помехоустойчивость и точность системы распознавания и измерения городских транспортных целей.

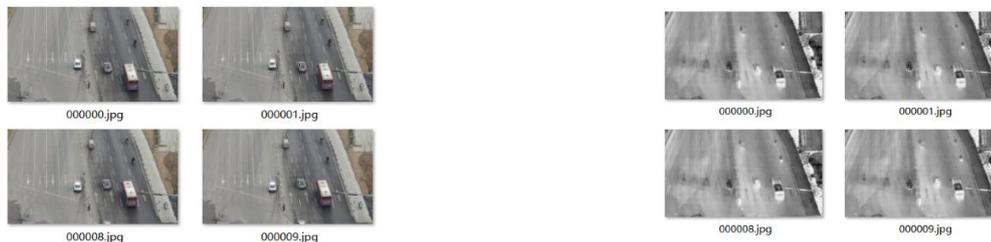
## **1. Создание мультимодального датасета**

### **1.1 Датасет DUT-VTUAUV**

В данной работе используется мультимодальный датасет DUT-VTUAUV, опубликованный Даляньским политехническим университетом [16]. В датасете собраны фото, сделанные с помощью дронов DJI, оснащённых камерами Zenmuse H20T. Часть из этих данных представлена на рисунке 1. В

---

данной работе для экспериментов по распознаванию объектов используются только изображения с транспортными средствами из датасета.



(a) Датасет, получен в диапазоне визуального наблюдения

(b) Датасет, получен в диапазоне инфракрасного излучения

Рис. 1. Часть изображений из датасета DUT-VTUA V

## 1.2 Мультимодальное слияние данных

Для улучшения чёткости характеристик аэрофотоснимков в данной статье используется метод мультимодального слияния данных, полученных в диапазоне визуального наблюдения и в диапазоне инфракрасного излучения в условиях дневного и ночного времени суток. Преимущества обоих типов данных позволяют улучшить четкость контуров краев объектов в полученных изображениях. Для слияния датасетов в настоящей работе используется алгоритм слияния изображений TIF (Two Scale Image Fusion).

### 1.2.1 Алгоритм слияния изображений TIF

Алгоритм TIF в основном осуществляет слияние изображений по трём этапам: разложение изображений, слияние изображений и восстановление изображений [17]. Конкретная схема слияния показана на рисунке 2.

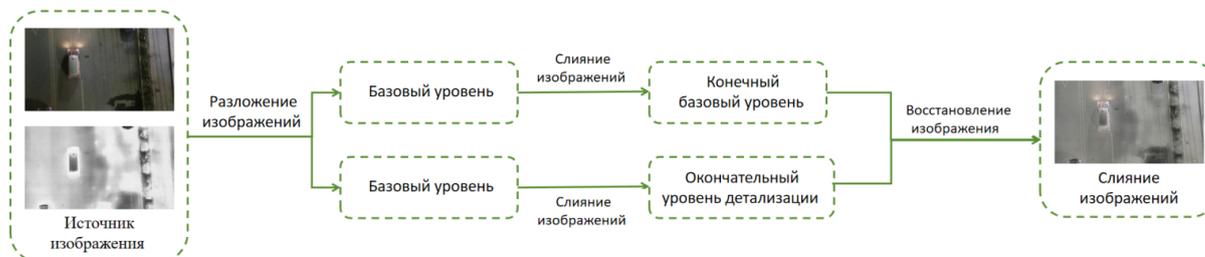


Рис. 2. Алгоритм слияния изображений TIF

(1) Разложение изображения: сначала проводится декомпозиция двух входных изображений с использованием среднего волнового фильтра для получения изображения базового слоя, после этого, для приобретения слоя деталей, выполняется операция разности между исходным входным изображением и изображением базового слоя.

(2) Слияние изображений: употребляются различные алгоритмы слияния для базового и детализированного слоев. При базовом слое используется усредненная стратегия объединения для получения конечного базового слоя; при слое деталей используется средневзвешенная стратегия объединения с целью получения конечного слоя деталей.

(3) Восстановление изображения: после получения конечного базового слоя и окончательного слоя детализации, каналы R, G и B базовых уровней и уровней детализации объединяются с помощью простых операций сложения. Процесс объединения изображений окончательного канала показан на рисунке 3.

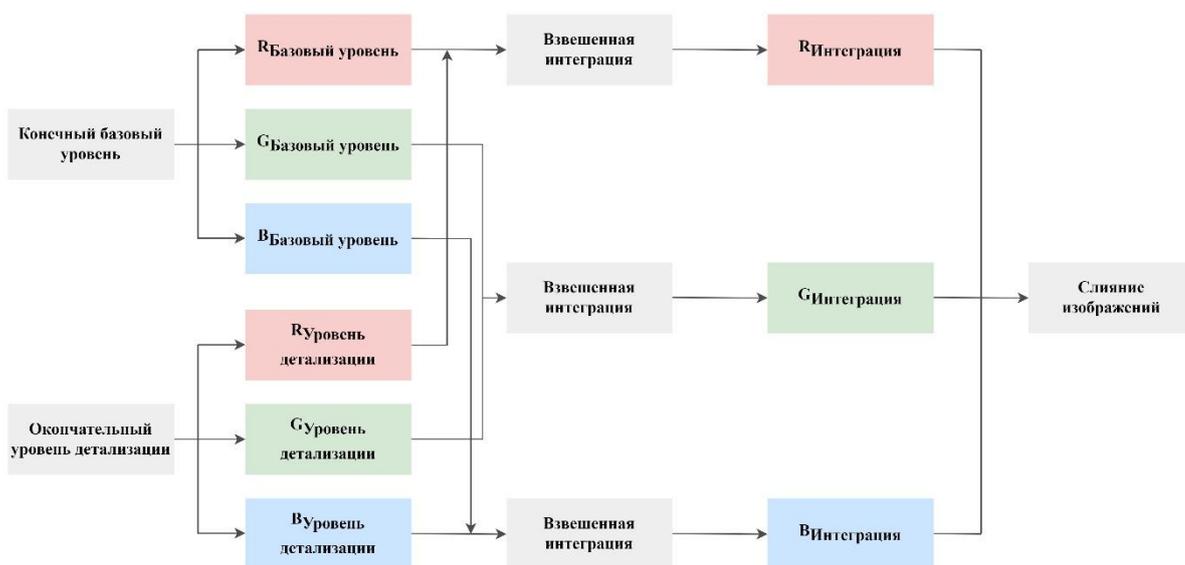


Рис. 3. Процесс слияния многоканальных данных

### 1.2.2 Слияние данных изображений из диапазона визуального наблюдения и инфракрасного излучения

Результаты слияния датасетов с использованием алгоритма TIF показаны на рисунке 4. Исходя из этого, видно, что в ночной среде посредством слияния данных эффективно соединяются преимущества изображений двух режимов, образуя взаимодополняющие преимущества и делая контуры краев изображения более чёткими. Этот алгоритм легко реализуем, работает быстро и не требует времязатратных операций, что делает его удовлетворяющим требованиям к распознаванию и измерению городских дорожных целей в реальном времени при применении БПЛА.

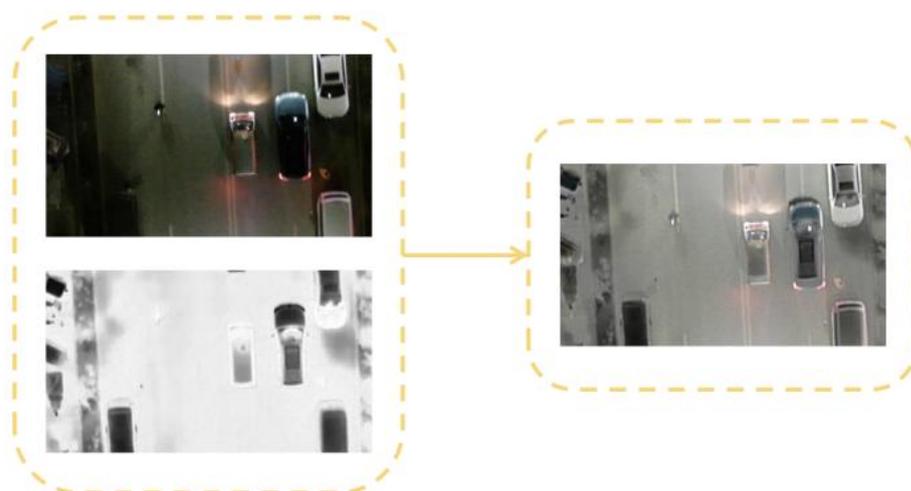


Рис. 4. Эффект слияния изображений на основе алгоритма TIF

### 1.3 Создание и аннотирование датасета

Учитывая, что датасет DUT-VTUAUV относится к датасету для отслеживания целей и не имеет меток в формате VOC, необходимых для данного исследования, из датасета было выбрано 2050 пар соответствующих изображений, полученных в диапазоне визуального наблюдения и в диапазоне инфракрасного излучения по отдельности, сделанных в дневных и ночных дорожных условиях, включающих в себя фото с видами движения автомобилей на городских дорогах и скоростных шоссе, снятые

---

беспилотниками с разных ракурсов. На основе программы Labelimg было проведено аннотирование датасета [18], как показано на рисунке 5. В результате были построены унимодальный набор данных из диапазона визуального наблюдения, унимодальный набор данных из диапазона инфракрасного излучения и мультимодальный набор данных для обучения и тестирования задачи распознавания целей, представленные в данной статье.



Рис. 5. Пример аннотации с помощью инструмента Labelimg

## **2. Проектирование алгоритма мультимодального слияния для определения целевых объектов**

### **2.1 Сравнение производительности моделей из серии YOLO**

В данной статье было проведено сравнение производительности существующих моделей серии YOLO v3-YOLO v8 на созданном наборе данных о городских автомобилях в диапазоне визуального наблюдения, представленном выше. В ходе обучения, количество итерации было ограничено 300 раз и размер входного изображения составлял  $640 \times 640$ . Результаты сравнения производительности моделей серии YOLO показаны в таблице 1. С учётом результатов сравнения и таких показателей, как размер модели, мощность алгоритма, точность и скорость распознавания, выбирается наиболее подходящая модель для распознавания и измерения городских автомобильных целей с участием БПЛА. Результаты сравнения также служат основой для выполнения задачи по распознаванию целей других типов.

Можно видеть, что при наборе данных городских автомобилей в данной статье, точность распознавания объектов различных моделей серии YOLO относительно высока и различия между моделями незначительны, с mAP50,

достигающим 99%. Однако существует значительная разница в количестве параметров, требованиях к вычислительной мощности и скорости вывода между моделями. Учитывая то, что конечной платформой системы распознавания объектов в данной статье является БПЛА, объём памяти и вычислительные ресурсы бортовых компьютеров ограничены [19], сеть распознавания объектов должна обладать характеристиками высокой точности, легковесности и простоты мобильной дислокации. Поэтому, в данном исследовании была выбрана самая легковесная модель серии YOLO – YOLO v5n, которая имеет небольшую разницу в точности распознавания объектов по сравнению с другими моделями и занимает второе место по скорости получения вывода. Она подходит для периферийных устройств и обеспечивает точное и быстрое распознавание объектов.

Таблица № 1

Сводная таблица сравнения YOLO

Method	Params(M)	GFLOPs	mAP@0.5:0.95	mAP@0.5
Yolov3	61.5	154.5	0.892	0.997
Yolov4	63.9	42.5	0.92	0.998
<b>Yolov5n</b>	<b>1.76</b>	<b>4.2</b>	<b>0.897</b>	<b>0.992</b>
Yolov7	36.5	103.2	0.902	0.999
Yolov8n	3.0	8.1	0.918	0.995

## 2.2 Мульти模альная сеть слияния изображений из диапазона визуального наблюдения и инфракрасного излучения

Структура мульти模альной сети слияния изображений из диапазона визуального наблюдения и инфракрасного излучения в этой статье показана на рисунке 6. Сначала, с употреблением алгоритма слияния изображений TIF достигается слияние данных изображений из диапазона визуального наблюдения и инфракрасного излучения. Затем, на основе модели YOLO v5n, производится распознавание и измерение целей на сливаемом изображении, и

выводится информация о местоположении и виде целей. Благодаря комбинации данных изображений из диапазона визуального наблюдения и инфракрасного излучения на уровне пикселей, данная сеть эффективно повышает помехоустойчивость и точность системы беспилотного летательного аппарата для распознавания и измерения целей.

В частности, структура сети YOLO v5n в основном состоит из «позвоночника», «шеи» и «головы». «Позвоночник» в основном выполняет задачу по извлечению информации из входных данных для последующего использования. «Шея» используется для сбора признаков карт с разных этапов, её функция заключается в лучшем слиянии и извлечении признаков, предоставленных «позвоночником», что позволяет улучшить производительность сети. Задачей «головы» является определение типа и местоположения целей во входном изображении.

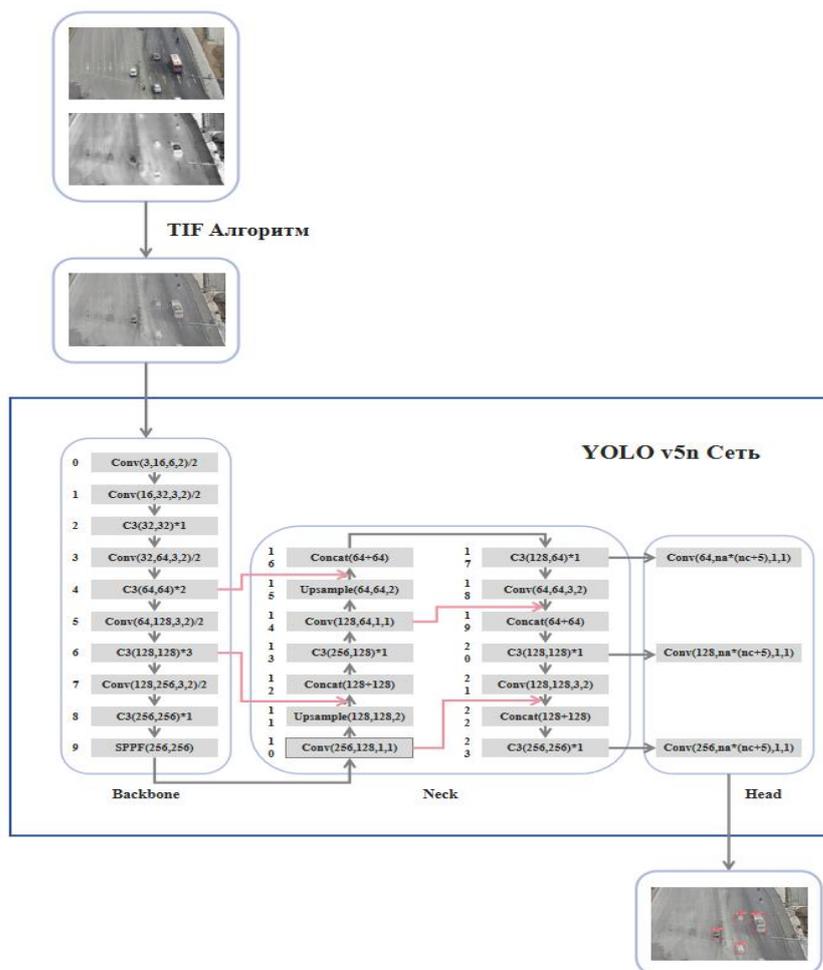


Рис. 6. Схема сети распознавания целей с мультимодальным слиянием

### 3. Экспериментальные результаты и анализ

#### 3.1 Экспериментальная среда

На экспериментальном этапе данного исследования был использован фреймворк глубокого обучения, Python. Экспериментальная среда для данного исследования следующая:

Центральный процессор (CPU) – 11-ое поколение Intel (R) Core (TM) i7, графический процессор (GPU) – RTX 3060, операционная система – win11, и CUDA – 11.1.

#### 3.2 Сравнение и анализ экспериментальных результатов

В этом эксперименте набор данных был разделен на обучающий набор, валидационный набор и тестовый набор в соотношении 1621:181:201, и

подан на вход сети YOLO v5n для проведения симуляционных экспериментов. Сравнение производительности между алгоритмом мультимодального слияния и унимодальным алгоритмом показано в таблице 2.

Таблица № 2

Сравнение производительности моделей алгоритмов

	<b>mAP@0.5:0.95</b>	<b>mAP@0.5</b>
Диапазон визуального наблюдения	0.894	0.994
Диапазон инфракрасного излучения	0.913	0.995
В мультимодальной сети	<b>0.915</b>	<b>0.996</b>

По результатам сравнения можно сделать вывод, что сеть YOLO v5n имеет лучшую производительность при распознании и измерении объектов в изображениях из мультимодального набора данных, mAP@0.5 которого достигает до 99.6%, а mAP@0.5:0.95 составляет 91.5%, и её точность на 2.1% выше показателя одномодальной сети распознания объектов в диапазоне визуального наблюдения и на 0.2% выше точности одномодальной сети распознания объектов в диапазоне инфракрасного излучения. Это подтверждает, что мультимодальный алгоритм TIF в данной статье может эффективно повысить точность работы при выполнении задач определения объектов в реальном применении.

Кроме того, в средах с недостаточным освещением ночью, преимущества алгоритма мультимодального слияния будут более очевидны. В данной статье было выбрано случайное изображение для тестирования с целью проверки эффективности предложенного метода. Как показано на рисунке 7, с употреблением обученной в данной статье одномодальной сети с использованием диапазона визуального наблюдения, с использованием диапазона инфракрасного излучения и мультимодальной сети, было проведено распознавание одного и того же изображения с ночными видами.

В связи с тем, что ночная среда темна, одномодальная сеть с использованием диапазона визуального наблюдения успешно выявила цели, но уровень уверенности составил всего лишь 91%. Хотя тепловые характеристики объектов явно выражены в инфракрасном изображении, всё же наблюдаются ложные выводы о них, сделанные сетью с использованием диапазона инфракрасного излучения. Мультимодальный алгоритм определения объектов, благодаря интеграции данных изображений из диапазона визуального наблюдения и изображений из диапазона инфракрасного излучения, не только исключает ложные выводы, но и имеет уровень уверенности в определении целей на 94%, что на 3% превышает уверенность в результатах одномодальной сети с использованием диапазона визуального наблюдения.



(a) Эффект сети в одном режиме визуального наблюдения



(b) Эффект сети в одном режиме инфракрасного излучения



(с) Эффект сети слияния  
многомодальных данных

Рис. 7. Результаты моделей в ночной среде

### Заключение

В данной статье представлен алгоритм определения целевых объектов с применением метода мультимодального слияния end-to-end, который осуществляет точное и эффективное распознавание и измерение городских транспортных средств. Путём объединения данных изображений, полученных с помощью диапазона визуального наблюдения и инфракрасного излучения, был создан мультимодальный набор данных, который имеет преимущества перед вышеуказанными двумя типами одномодальных изображений, и способен исключить недостатки одномодальных изображений, такие, как недостаточная чёткость и низкая яркость. В сочетании с самой сбалансированной сетью YOLO v5n, по результатам сравнения производительности серии YOLO, проводится анализ для более точного определения целей в реальной среде. Многогрупповые сравнительные эксперименты показывают, что предложенный в этой статье алгоритм определения целей с использованием метода мультимодального слияния изображений имеет лучший эффект анализа набора данных о транспортных средствах по сравнению с моделью сети одиночного режима визуального наблюдения/инфракрасного излучения, и точность определения составляет до 99.6%. Предложенный в этой статье алгоритм имеет большие

преимущества в условиях ночи. Однако данный алгоритм всё ещё нуждается в улучшении, поэтому ближайшими перспективами являются:

(1) Дальнейшая разработка мультимодального набора данных. Используемый в настоящем эксперименте набор данных включает в себя только изображения с целевыми объектами транспортных средств, что приводит к узкой категории образцов. Для повышения надежности эксперимента и универсальности метода можно добавить в набор данных сведения о других городских объектах;

(2) Дальнейшее слияние информации из нескольких источников. Принятый в этой статье метод мультимодального слияния осуществляется на уровне набора данных, которые относятся к фронтальной интеграции. Даже если он лучше сочетает преимущества данных, исходя из диапазона визуального наблюдения и инфракрасного излучения по отдельности, всё ещё сложно объединить характеристики на глубоком уровне, и это даже может привести к избыточности характеристик [20]. Поэтому мы можем продолжить изучение мультимодальной интеграции на уровне признаков и решений в последующей работе и применить её в процессе эксперимента, чтобы проверить, имеет ли она лучший функциональный эффект.

### **Литература(References)**

1. Yu C, Jiang X, Wu F, Fu Y, Zhang Y, Li X, Fu T, Pei J. Research on Vehicle Detection in Infrared Aerial Images in Complex Urban and Road Backgrounds. Electronics 2024, 13, 319 p.

2. Micheal A.A., Vani K, Sanjeevi S, Lin C.H. Object detection and tracking with UAV data using deep learning. Journal of the Indian Society of Remote Sensing, 2021, 49: pp. 463-469.

3. Nasrabadi N.M. Hyperspectral target detection: An overview of current and future challenges. IEEE Signal Processing Magazine, 2013, 31(1): pp. 34-44.

4.Zhang Yangting, Huang Deqi, Wang Dongwei, He Jiajia. Review on Research and Application of Deep Learning-Based Target Detection Algorithms. *Computer Engineering and Applications*, 2023, 59(18): pp. 1-13.

5.Zou Jun, Zhang Shiyi, Li Jun. Study on Improved Ant Colony Path Planning Algorithm. *Sensor World*, 2023, 29(08): pp. 9-15.

6.Liu C, Tao Y, Liang J, Li K, Chen Y. Object detection based on YOLO network//2018 IEEE 4th information technology and mechatronics engineering conference (ITOEC). IEEE, 2018: pp. 799-803.

7.Lyu Tao, Song Minmin, Hong Wenpeng, Chen Jing, Fu Xiaogang, Ji Yaping, Sang Xueyi. Infrared Target Recognition Method Based on Improved YOLOv2. *Aerospace Shanghai (Chinese & English)*, 2023, 40(02): pp. 126-133.

8.Zhao L, Li S. Object detection algorithm based on improved YOLOv3. *Electronics*, 2020, 9(3): 537 p.

9.Singha S, Aydin B. Automated drone detection using YOLOv4. *Drones*, 2021, 5(3): 95 p.

10.Luo X, Wu Y, Wang F. Target detection method of UAV aerial imagery based on improved YOLOv5. *Remote Sensing*, 2022, 14(19): 5063 p.

11.Cheng Huanxin, Xu Haotian, Luo Xiaoling. Automatic driving target detection method based on improved YOLOv7. *Laser Journal*: pp. 1-8 [2023-10-27].

12.Wang G, Chen Y, An P, Hong H, Hu J, Huang T. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors*, 23(16), 7190 p.

13.He A, Wang X. Research on Object Detection Algorithm Based on Anchor-free//2023 International Conference on Advances in Electrical Engineering and Computer Applications (AEECA). IEEE, 2023: pp. 712-717.

14.Sridevi M, Sugirtha T, Rashed H, Kiran B.R., Yogamani S. Object Detection, Tracking and Trajectory Prediction for Autonomous Driving. In

---

Autonomous Driving and Advanced Driver-Assistance Systems (ADAS) (pp. 105-134). CRC Press.

15.Cheng X, Geng K, Wang Z, Wang J, Sun Y, Ding P. SLBAF-Net: Super-Lightweight bimodal adaptive fusion network for UAV detection in low recognition environment. *Multimedia Tools and Applications*, 82(30), pp. 47773-47792.

16.Zhang P, Zhao J, Wang D, Lu H, Ruan X. Visible-thermal UAV tracking: A large-scale benchmark and new baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8886-8895).

17.Bavirisetti D.P., Dhuli R. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Physics & Technology*, 2016, 76: pp. 52-64.

18.Илани М.А., Амини Л, Карими Н, Кушшери М.С. CNN-based Labelled Crack Detection for Image Annotation. 2024.

19.Zhou Jinwei, Wang Jianping. Review of YOLO Object Detection Algorithms. *Journal of Changzhou Institute of Technology*, 2023, 36(01): pp. 18-23+88.

20.Li Yang, Zhao Ming, Xu Mengyao, Liu Yunfei, Qian Yuchen. A survey of research on multi-source information fusion technology. *Intelligent Computer and Applications*, 2019, 9(05): pp. 186-189.

**Дата поступления: 27.07.2024**

**Дата публикации: 6.09.2024**