УЛУЧШЕНИЕ АЛГОРИТМА YOLO-V3 ДЛЯ РАСПОЗНАВАНИЯ НЕБОЛЬШИХ ОБЪЕКТОВ

IMPROVING THE YOLO-V3 ALGORITHM FOR RECOGNIZING SMALL OBJECTS



УДК 339.54.012+338.001.36

Лоу Гуанпин, магистрант,
Московский государственный технический университет, имени Н.Э. Баумана

Лоу Гуанпин aliksandal@gmail.com

Аннотация

Сельскохозяйственные беспилотные летательные аппараты используются в различных странах мира и имеют большой потенциал для развития. В Японии, которая является одной из крупнейших стран по производству сельскохозяйственных беспилотных летательных аппаратов, более 60% сельскохозяйственных беспилотных аппаратов, используемых в области защиты растений, и более 2000 сельскохозяйственных беспилотных аппаратов, распыляющих пестициды на рисовых полях каждый год, что обеспечивает высокую эффективности в области защиты растений в Японии. Также в Корее сельскохозяйственные беспилотники играют важную роль в области защиты растений. В Китае использование сельскохозяйственных беспилотных летательных аппаратов еще не достигло зрелого уровня и находится на стадии развития. Однако, пахотные земли Китая обширны и сельскохозяйственные беспилотные летательные аппараты имеют весьма многообещающие перспективы развития.

В настоящее время сельскохозяйственные беспилотники внедряются на различных посевных полях для защиты растений, перекрестного опыления, сельскохозяйственного дистанционного зондирования, сбора наземной информации, обнаружения и диагностики вредителей и многих других областях. Кроме того, они имеют большой потенциал применений, перспективы на рынке. Если взять в качестве примера задачу распыления пестицидов, то исследования показывают, что ежедневная работа по ручному распылению пестицидов может достигать 2-3 квадратных километров, в то время как работа сельскохозяйственных беспилотников может достигать 200-300 квадратных километров, что примерно в 100 раз эффективнее, чем ручное распыление.

Использование беспилотных летательных аппаратов в сельском хозяйстве, например, для обнаружения и диагностики вредителей и болезней сельскохозяйственных культур, неотделимо от обнаружения и распознавания целей, а в сельском хозяйстве многократно необходимо распознавать цели небольшого размера. Однако широко используемые алгоритмы обнаружения целей приводят к плохому распознаванию и высокой частоте промахов из-за низкого разрешения цели и особенностей при обнаружении мелких целей. Поэтому очень важно повысить способность БПЛА обнаруживать мелкие цели.

Целью данной работы является улучшение способности YOLO-V3 распознавать цели маленького размера.

Annotation

Agricultural drones are used in various countries around the world and have great potential for development. In Japan, which is one of the largest agricultural drone countries, more than 60% of agricultural drones used in crop protection, and more than 2,000 agricultural drones spraying pesticides in rice fields every year, ensuring high efficiency in crop protection in Japan. Also in Korea, agricultural drones play an important role in crop protection. In China, the use of agricultural drones has not yet reached a mature level and is in the development stage. However, China's arable land is vast and agricultural drones have very promising development prospects.

At present, agricultural drones are being introduced in various crop fields for crop protection, cross-pollination, agricultural remote sensing, ground-based information gathering, pest detection and diagnosis, and many other fields. In addition, they have great potential applications, market prospects. Taking pesticide spraying as an example, studies show that a daily manual pesticide spraying operation can reach 2-3 square kilometers, while agricultural drones can reach 200-300 square kilometers, which is about 100 times more efficient than manual spraying.

The use of drones in agriculture, for example to detect and diagnose crop pests and diseases, is inseparable from target detection and recognition, and many times in agriculture small size targets must be recognized. However, widely used target detection algorithms result in poor recognition and high miss rate due to low target resolution and peculiarities in detecting small targets. Therefore, it is very important to improve the UAV's ability to detect small targets.

The purpose of this work is to improve the ability of YOLO-V3 to detect small size targets.

Ключевые слова: глубокое обучение, нейронные сети, системы компьютерного зрения, распознавание объектов, улучшение алгоритмов.

Keywords: deep learning, neural networks, computer vision systems, object recognition, algorithm improvement.

Нейронный сеть YOLO-V3

Для задачи распознавания и классификации целей существует множество видов сетей, которые могут это выполнить, таких как RCNN, SSD, YOLO и др. Однако, в практическом использовании, нам необходимо достичь распознавания в реальном времени, поэтому скорость обработки и точность сети требуются очень высокой, поэтому мы выбираем YOLO-V3 для выполнения задачи распознавания и классификации. У нейронной сети YOLO есть следующие преимущества.

1) Распознавание YOLO быстро, сеть решает проблему распознавания объектов как регрессионную, а структура всей сети распознавания не сложна.

При гарантированной точности распознавания (63,4% mAP, тестовый набор VOC 2007) скорость распознавания может достигать 45 кадров в секунду.

2) Сеть понимает обобщенное представление объектов (это позволило обучить сеть на реальных изображениях, и прогнозы все еще были довольно точными).

3) YOLO может избежать фоновых ошибок, YOLO может избежать контекстных ошибок, в отличие от других систем обнаружения объектов, классификатор которых может получить только локальную информацию о изображении. YOLO сможет видеть информацию о целом изображении как во время обучения, так и во время тестирования, поэтому YOLO может хорошо использовать контекстную информацию при обнаружении объектов и таким образом может избежать прогнозирования неправильной информации об объекте на фоне.

Но, несмотря на многие преимущества YOLO в задачах многоцелевого распознавания, ему еще многое предстоит усовершенствовать, чтобы соответствовать нашим целям.

Улучшение структуры сети

Так как нашей задачей является распознавание целей с помощью системы технического зрения, установленной на БПЛА, из-за информации об изображении, собранной БПЛА в воздухе, размер распознаваемых целей невелик. Но многие сети признания не очень хорошо распознают маленькие цели, поэтому мы хотим улучшить нашу сеть YOLO для маленькие целей.

Основой YOLO-сети распознавания является darknet53, признаки изображения извлекаются darknet53, Затем входные изображения были подвергнуты пониженной дискретизации (downsampled) еще 5 раз. Последние 3 пониженных дискретизации (downsampled) содержат 3-масштабные карты признаков обнаружения цели. Размеры этих трех выходов с пониженной дискретностью являются 32x, 16x и 8x соответственно. С помощью этих трех различных размеров карт признаков выполняется распознавание целей различного размера.

Три распознавания, каждая из которых соответствует разному полю восприятия. 32-кратная пониженная дискретизация имеет наибольшее поле восприятия и подходит для обнаружения больших целей. 16-кратная подходит для объектов среднего размера. 8-кратная имеет наименьшее поле восприятия и подходит для обнаружения маленьких целей.

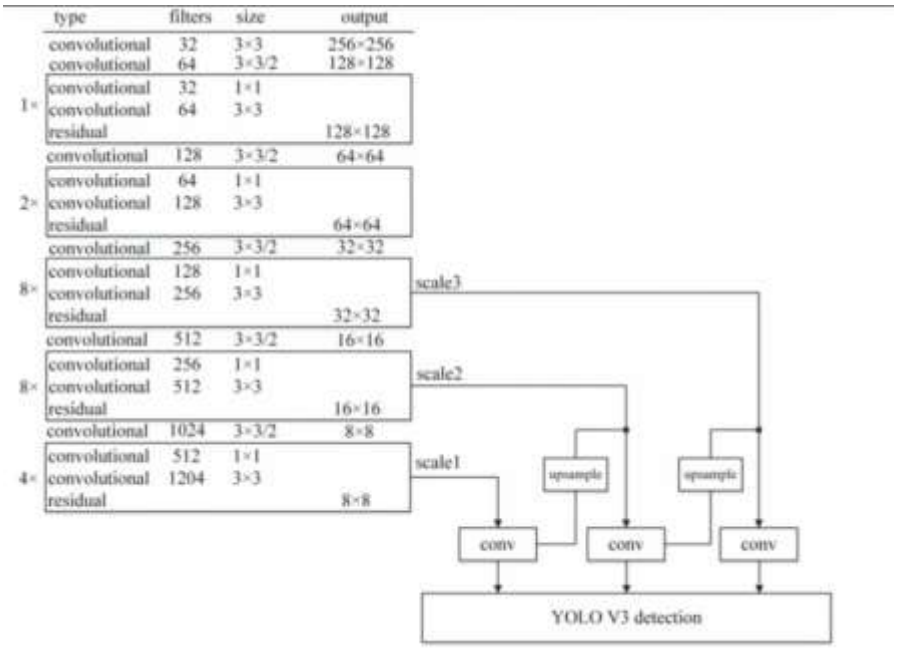


Рисунок 1 – Исходная архитектура YOLO-V3

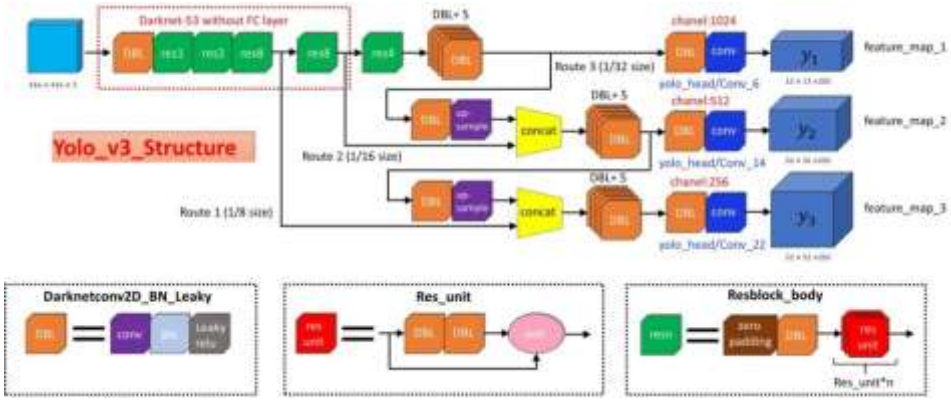


Рисунок 2 – Описание исходной архитектуры YOLO-V3

Сеть YOLO V3 распознает малые цели с помощью 8-кратной дискретизации (downsampled) карты признаков, что означает, что когда цель меньше 8 пикселей × 8 пикселей, то сети трудно прогнозировать цель, Таким образом, целевой уровень распознавания с помощью 8-кратной пониженная дискретизация (downsampled) имеет ограниченную возможность распознавания информации о небольшом целевом расположении. Для того, чтобы получить

больше информаций признаков малых целей и улучшить точность распознавания маленьких целей, цель распознается с помощью выхода 4-кратной и 8-кратной пониженной дискретизации карты признаков в исходной сети, так как она содержит больше информации о местоположении малых целей.

Сетевая архитектура YOLO V3 для 4-кратного распознавания целей с пониженной дискретизацией показана ниже.

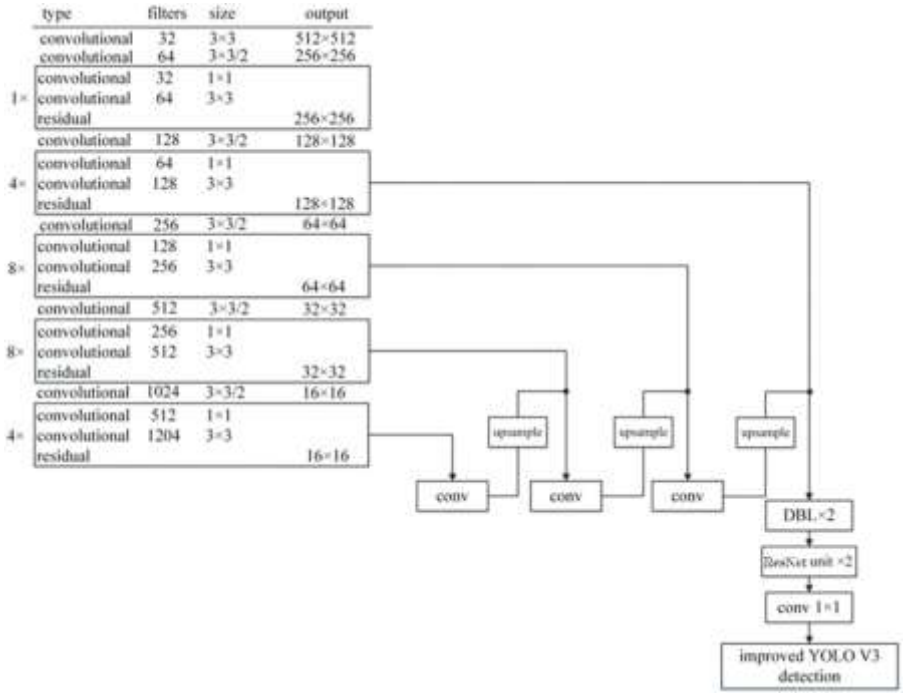


Рисунок 3 – Сетевая архитектура YOLO V3 для 4-кратного распознавания целей пониженной дискретизацией

Вывод карты признаков с 8-кратный пониженной дискретизацией из YOLO V3 увеличивается(upsampled) вдвое, а 2-кратный вывод карты признаков с пониженной дискретизацией соединяется с 4-кратным выводом карты признаков из 2-го остаточного блока в Darknet53 для построения слоя распознавания цели объединения слоя с 4-кратным выводом с пониженной дискретизацией для распознавания малых целей. В то же время, два остаточных блока добавляются ко второму остаточному блоку исходной сети, чтобы получить больше информации о расположении малых целей.

Из-за ограниченных вычислительных возможностей бортовой системы технического зрения БПЛА нам необходимо уменьшить объем вычислений, удовлетворяя при этом точность, поэтому для распознавания препятствий мы

будем использовать только выходов с 4-кратной и 8-кратной пониженной дискретностью.

Улучшения алгоритм кластеризации

Позаимствовав у алгоритма Faster R-CNN, YOLO V2 вводит Anchor Box и устанавливает ее на 5. В алгоритме YOLO V3 Anchor Box изменяется с 5 на 9, а размер Anchor Box получается алгоритмом кластеризации K-Means путем кластеризации длины и ширины всех реальных целевых Коробок. При прогнозировании смещения вывода сети относительно Якорного Коробки составляют tx,ty,tw,th, а отношение между Bounding Box и Anchor Box показано на рисунке 4.

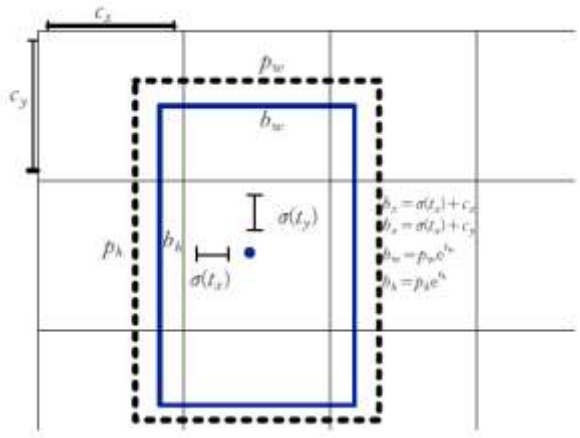


Рисунок 4 – Отношение между Bounding Box и Anchor Box

Координаты Bounding Box вычисляются как:

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases}$$

где

bx,by,bw,bh - координаты центра, а также ширина и высота Bounding Box;

cx,cy - смещения текущей сетки ячеек относительно верхнего левого угла изображения;

pw,ph - ширина и высота соответствующей Anchor Box;

σ - функция активации Sigmoid, которая отображает сетевые прогнозы tx,ty до 0~1.

Алгоритм кластеризации K-Means

Алгоритм кластеризации K-Means разбивает множество элементов векторного пространства на заранее известное число кластеров k. Действие алгоритма таково, что он стремится минимизировать среднеквадратичное отклонение на точках каждого кластера. Основная идея заключается в том, что на каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения кластеров.

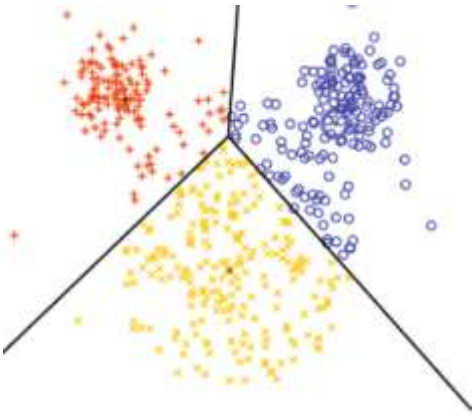


Рисунок 5 – Схема алгоритм кластеризации K-Means

Алгоритм кластеризации K-Means используется в YOLO V3 для генерация Anchor Box. Алгоритм кластеризации k-Means использует расстояние в качестве метрики сходства при итерации кластеризации для нахождения К-классов в заданном наборе данных, а центр каждого класса получается из средней величины всех точек данных в этом классе. Данный метод рассматривает каждую точку признака в формуле расстояния как одинаково взвешенную и не учитывает влияние различных точек признака на результаты кластеризации; Если в классе имеются шумные или изолированные точки, которые находятся далеко от пространства выборки данных, то они производят большие колебания при расчете центра класса, что должно сильно влиять на расчет среднего значения, или даже заставляют центр кластеризации серьезно отклоняться от

плотной области выборки класса, что приводит к серьезным отклонениям в результатах кластеризации .

Для решения проблем алгоритма кластеризации K-Means, в данной работе мы используем алгоритм кластеризации DBSCAN + K-Means для кластеризации и анализа набора обучающих данных, чтобы выбрать Anchor Box, которая больше подходит для обнаружения малых целей.

Алгоритм кластеризации DBSCAN

Алгоритм DBSCAN - это алгоритм кластеризации на основе плотности, который обычно предполагает, что категории могут быть определены по близости распределения выборки. Образцы одной и той же категории тесно связаны друг с другом, т.е. образцы одной и той же категории должны присутствовать недалеко от любой пробы этой категории.

Рассмотрим набор точек в некотором пространстве, требующий кластеризации. Для выполнения кластеризации DBSCAN точки делятся на основные точки, достижимые по плотности точки и выпадающие следующим образом:

1)Точка p является основной точкой, если по меньшей мере minPts точек находятся на расстоянии, не превосходящем ϵ (ϵ является максимальным радиусом соседства от p), включая саму точку p. Говорят, что эти точки *достижимы прямо* из p.

2)Точка q прямо достижима из p, если точка q находится на расстоянии, не большем ϵ , от точки p и p должна быть основной точкой.

3)Точка A q *достижима* из p, если имеется путь $p_1, \dots, p_2$ с $p_1 = p$ и $p_n = q$, где каждая точка p_{i+1} достижима прямо из p_i (все точки на пути должны быть основными, за исключением q).

4)Все точки, не достижимые из основных точек, считаются *выбросами*.

Теперь, если p является основной точкой, то она формирует *кластер* вместе со всеми точками (основными или неосновными), достижимые из этой точки. Каждый кластер содержит по меньшей мере одну основную точку. Неосновные точки могут быть частью кластера, но они

формируют его «край», поскольку не могут быть использованы для достижения других точек.

Улучшение нейронной сети

В нашей сети, алгоритм кластеризации DBSCAN используется для анализа количества кластеров и точек, содержащихся в кластерах, а также снятия вопроса о влиянии шумовых точек или изолированных точек на результаты кластеризации, и для получения точного значения K для алгоритма кластеризации K-Means. Затем результаты кластеризации DBSCAN анализируются алгоритмом кластеризации K-Means для определения центра кластеризации, и далее выбирается более подходящее Anchor Box в соответствии со степенью перекрытия IOU. Формула расстояния алгоритма кластеризации K-Means для процесса кластеризации определяется как:

$$D(B,C) = 1 - IOU(B,C)$$

$$IOU = \frac{B \cap A}{B \cup A}$$

где B обозначает размер коробки маркировки, A - размер кластерной коробки, C - кластерные центры, созданные алгоритмом кластеризации K-Means, а IOU - пересечение относительно объединения.

В данной работе мы анализируем кластеризацию набора данных с помощью улучшенного метода кластеризации. После анализа с помощью алгоритма кластеризации DBSCAN значение K определяется как 9. Ширина и высота соответствующих Anchor Box, выбранных после итерации алгоритма кластеризации K-Means, равны (4, 10), (9, 24), (9, 44), (17, 30), (29, 58), (55, 130), (76, 167), (132, 243), (192 326). В таблице представлено сравнение результатов обнаружения Anchor Box на наборе данных до и после оптимизации метода кластеризации.

Таблица: Сравнение результатов распознавания с использованием различных Anchor Box

нейронный сеть	Anchor Box(якорные коробки)	Средняя точность (AP)/%	Скорость распознавания (fps) /(кадр·с-1)

YOLO V3	(4，10)、(9，24)、(9，44)、(17，30)、(29，58)、(55，130)、(76，167)、(132、243)、(192，326)	90.07	42.17
YOLO V3 с улучшенным алгоритмом кластеризации	(8，14)、(16，38)、(17，40)、(20，32)、(36，83)、(46，101)、(82，177)、(162、240)、(201，352)	90.83	43.31

Обучение сети

Улучшенный YOLO V3 был обучен с начальной скоростью обучения 0,001 и коэффициентом затухания 0,0005 на этапе обучения, а скорость обучения была снижена до 0,0001 и 0,00001, когда количество обучающих итераций составляло соответственно 20 000 и 25 000, для дальнейшего сближения функции потерь.

Кривая сходимости потерь при обучении усовершенствованной сети YOLO V3 показана на рисунке.

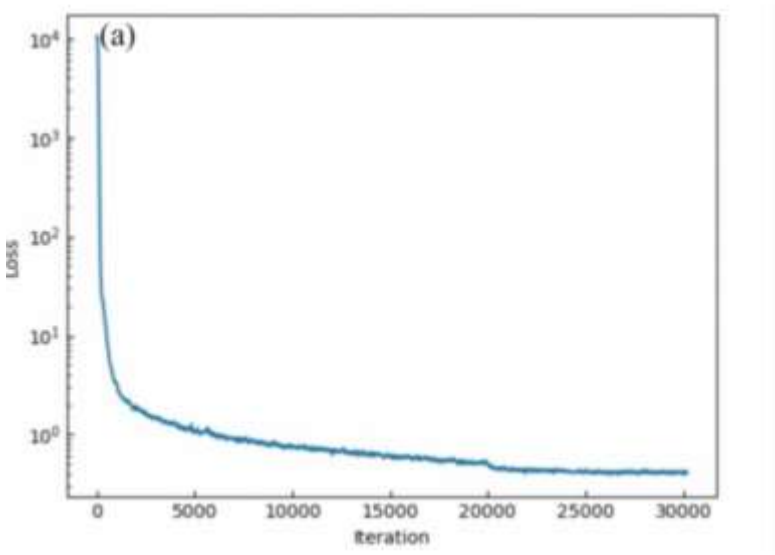


Рисунок 6 – Кривая сходимости потерь при обучении

После примерно 30 000 итераций параметры в основном стабилизировались, а конечное значение потери упало примерно до 0,2, и IOU Avg постепенно приблизился к 1, а в итоге стабилизировался примерно на 0,85. Из анализа сходимости этих параметров видно, что результаты обучения по усовершенствованной сети YOLO V3 являются более качественными.

Заключение

В данной работе мы выбрали сеть YOLO-V3 для выполнения задачи распознавания целей. Для улучшения СТЗ беспилотных летательных аппаратов для сельского хозяйства мы внесли следующие улучшения в сеть YOLO-V3: 1) Улучшили структуру сети за счет добавления слоя пониженной дискретности для распознавания малоразмерных объектов., в результате чего увеличивается эффективность распознавания. 2) улучшили метод кластеризации для получения Anchor Box. Исходная сеть YOLO-V3 использует только k-means метод кластеризации. Этот метод не исключает влияния на шум и изолированные точки. Поэтому в данной работе мы используем алгоритм кластеризации DBSCAN + K-Means для кластеризации и анализа набора обучающих данных. Этот алгоритм позволяет выбрать Anchor Box, которая больше подходит для обнаружения малых целей. Экспериментальные результаты показывают, что наша сеть имеет хорошую производительность в задачах распознавания и классификации целей.

Список литературы

1. He Y. et al. Research progress on intelligent obstacle detection methods for vehicles and their agricultural applications // Transactions of the Chinese Society of Agricultural Engineering. - 2018. - Т. 34. - №. 9.
2. Zheng et al. Improvement of YOLO V3 remote sensing image aircraft identification application // Electro-Optics and Control. - 2019. - Т. 26. - №. 4. - С. 28r32.
3. Jin Yao et al. Small pixel target detection for urban road video // Optoelectronic Engineering. - 2019. - Т. 46. - №. 9. - С. 190053.
4. Zhao, D.A. et al. YOLO-based deep convolutional neural network for robot apple picking localization in complex context // Journal of Agricultural Engineering. - 2019. - Т. 35. - №. 3. - С. 164-173.
5. Ju Muran et al. Improved YOLO V3 algorithm and its application to small target detection // Journal of Optics. - 2019. - Т. 39. - №. 7. - С. 0715004.

6. Jiang, J. J.. Research on small-target pedestrian detection based on convolutional neural networks : дис. - Lanzhou University of Technology, 2018.
7. Yang, Juanjuan. Research on agricultural UAV obstacle avoidance system based on deep learning : дис. - Gansu Agricultural University, 2019.
8. Liu, Xiaolong. Research on image acquisition and processing system based on UAV remote sensing platform : дис. - Zhejiang University, 2013.
9. Wang Chenyu. Research and application of air-ground cooperative system based on intelligent cart and UAV : дис. - Ningxia University, 2019.
10. Zhao, D.A. et al. YOLO-based deep convolutional neural network for robot apple picking localization in complex context // Journal of Agricultural Engineering. - 2019. - Т. 35. - №. 3. - С. 164-173.
11. Li Chuanpeng. Research on target recognition and grasping localization based on machine vision and deep learning : дис. - North Central University, 2017.
12. Li Chuanpeng. Research on target recognition and grasping localization based on machine vision and deep learning : дис. - North Central University, 2017.
13. Xue, Shukun. Active-passive vision-based obstacle detection and local path planning for robots : дис. - Nanjing University of Technology, 2013.