

# СИСТЕМА КОМПЬЮТЕРНОГО ЗРЕНИЯ РОБОТА-ГУМАНОИДА ДЛЯ ЗАДАЧИ РОБОФУТБОЛА

**С.С. Панов**

*Московский физико-технический институт*  
Россия, 141700, Долгопрудный, Институтский пер., 9  
E-mail: panov.ss@phystech.edu

**Ключевые слова:** задачи управления, ВСПУ-2024, компьютерное зрение, сегментация изображений, задача локализации.

**Аннотация:** Исследование посвящено доработке системы компьютерного зрения робота-гуманоида SAHR. Были выявлены недостатки текущей системы компьютерного зрения робота SAHR: распознавание разметки неустойчиво к изменению освещенности. Были описаны нейросетевые подходы компьютерного зрения, способствующие повышению устойчивости алгоритмов к динамически изменяющемуся освещению. Был проведен обзор и сравнение нейросетевых моделей YoloV8 и YOEО. Было выявлено, что YoloV8 быстрее и качественнее решает задачу семантической сегментации. Было показано, что предложенное решение более устойчиво к изменяющимся условиям освещения, нежели текущая версия системы компьютерного зрения SAHR.

## 1. Введение

Работа была написана на базе лаборатории волновых процессов и систем управления МФТИ. Один из проектов лаборатории - роботы-футболисты (SAHR - Starkit Autonomous Humanoid Robot). Их алгоритмы тестируются на соревнованиях по робофутболу, которые проводит всемирная организация под названием RoboCup. Данная организация является одной из основных драйверов развития гуманоидной робототехники в мире.

Соревнование запрещает использовать любые сенсоры, которыми в том или ином виде не обладает человек. Поэтому информацию об окружающей среде робот черпает только от камеры, гироскопа, акселерометра и датчиков давления. Обработкой данных занимается бортовой ПК - NUC. Тем самым, присутствуют ограничения по вычислительной способности и качеству получаемого кадра.

Текущая система компьютерного зрения распознает мячи, линии и прямые углы разметки. Часть информации об окружающей среде пропадает при изменении освещения. Пропадают именно линии и углы разметки, которые распознаются классическими подходами. А мяч, который детектируется посредством нейросетевой модели остается видимым.

В связи с этим, появилась гипотеза о том, что нейросетевые алгоритмы более устойчивы к изменению освещения. Их применение позволит улучшить локализацию робота на поле.

Таким образом, целью данной работы является повышение устойчивости робота к изменениям внешней среды (рис. 1 отображает информацию об окружающей среде, которую получает робот, в зависимости от уровня освещенности).

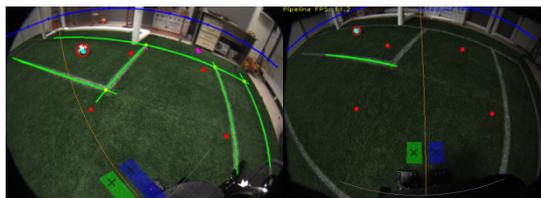


Рис. 1. Качество работы модуля зрения SAHR при разном уровне освещенности.

## 2. Основная часть

### 2.1 Архитектура системы компьютерного зрения SAHR

Архитектуру системы компьютерного зрения SAHR можно разделить на определенное количество этапов. Каждый этап называют «фильтром». Данные сущности являются обработчиками изображений. На вход каждому из них подается изображение, которое в процессе прохождения через фильтр видоизменяется и подается на выход.

Часть модуля зрения SAHR, отвечающая за распознавание линий и углов разметки, включает в себя 6 фильтров. При выделении линий и углов разметки алгоритм опирается на заданное подмножество цветового пространства HSV, которое описывает зеленый цвет поля. В случае правильной настройки данного подмножества алгоритм работает. Однако, при изменении освещения границы видимости зеленого поля в цветовом пространстве HSV сдвигаются, а используемые в алгоритме нет. В связи с чем, начиная с определенного сдвига, алгоритм перестает выделять зеленое поле. Это способствует потере информации о местоположении робота на поле, что, в свою очередь, критически снижает качество игры.

В текущей реализации распознавание разметки возможно только в условиях постоянной освещенности. В связи с этим, необходимо найти методы компьютерного зрения, которые способны работать в условиях изменяемого уровня освещенности. Такие методы, что и при изменении освещения будут продолжать распознавать паттерны, на которые настроены. Пусть и с меньшей точностью.

### 2.2. Гипотеза об устойчивости нейросетевых подходов к изменению уровня освещения

Для достижения поставленной цели было необходимо исследовать возможность применения нейросетевых подходов в данной области. Согласно статье [1], направленной на выявление устойчивости глубоких нейронных сетей к изменению освещения, была обнаружена положительная корреляция между уровнями яркости и потерей точности в моделях. Сравнение происходило по метрике Top-1 Accuracy. Сравнивались качества трех моделей сегментаторов, обученных непосредственно для проведения этого исследования (на рис. 2 изображен результат сравнение нескольких глубоких нейронных сетей).

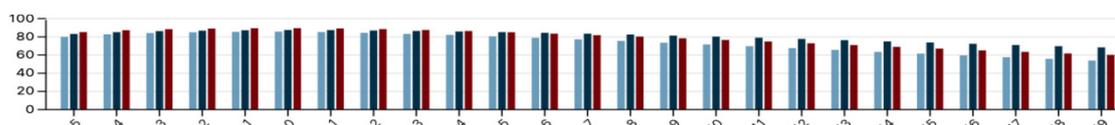


Рис. 2. График зависимости метрики Top-1 Accuracy от различных 24 уровней яркости изображений. Приведены результаты сравнения трех моделей: ResNet-50 [5], ResNet-RS-50 [3], ResNet-50+Augmix. Они описаны в статье [1].

Несмотря на потери точности в зависимости от уровня освещенности данные модели продолжали сегментировать изображения с высокими значениями метрики. В связи с этим, было решено использовать именно нейросетевые подходы.

Распознавание разметки можно свести к задаче сегментации изображения. Выделение пикселей, описывающих разметку, и обнуление остальных. То есть результатом обработки изображения является бинарная маска с выделенными полигонами разметки.

### 2.3. Поиск и обучение подходящих моделей, способных сегментировать изображения

Для достижения поставленной цели был произведен поиск подходящих моделей, способных сегментировать изображение с необходимой скоростью и качеством.

Первой стала нейросетевая модель YOEO [2]. Ее уже использовали в сфере робофутбола. Кроме того, качество масок со скоростью обработки набора изображений, полученные авторами статьи были весьма высокими.

Вторым кандидатом стала модель YoloV8 [4] для сегментации. Она достигла наилучших результатов в тестах на обнаружение объектов и на выявление масок семантической сегментации в начале 2023 году, сохраняя при этом высокую скорость и эффективность.

Обучающая выборка представляет собой сами изображения и их маски, с выделенными полигонами разметки. Был выбран датасет Torso, который включает в себя 10 000 изображений поля. На нем в дальнейшем и проводилось обучение моделей сегментации.

Обучение YOEO оценивалось по метрике IOU.

Обучение YoloV8 оценивалось по метрикам Precision, mAP50 и Recall.

### 2.4. Выбор метрик и оценка моделей

Изучение модуля зрения SAHR определило итоговые метрики оценки моделей. В связи с наличием ограничений по качеству получаемого с камеры кадра, вычислительной способности бортового ПК SAHR и скорости реакции роботов на изменения внешней среды был введен перечень характеристик, которыми должны обладать модели:

- Скорость обработки изображений модулем зрения не ниже 10 FPS. Будет произведена количественная оценка.
- Качество маски на выходе сегментатора лучше, нежели у фильтра зеленого. Будет произведена качественная оценка.

Были написаны инференсы данных моделей. И в дальнейшем определена наилучшая из них по введенному перечню метрик (в таблице 1 отражены результаты сравнения двух моделей).

**Таблица 1.** Сравнение скорости обработки изображений подходами YOEO и YoloV8. (i5 9-го поколения).

Нейросетевая модель	FPS с учетом обработки	FPS без учета обработки	Лучший FPS выборки	Худший FPS выборки
YOEO	11.0	13.8	15.6	7.8
YoloV8	16.7	24.1	27.8	13.9

На одном и том же наборе изображений YoloV8 показала скорость выше скорости YOEO в 1.5 раза.

Также была проведена качественная оценка правдоподобия полученных масок. По итогам сравнения, маска YoloV8 точнее отображает разметку на изображении (на рис. 3 показаны полученные маски YoloV8 и YOEO).

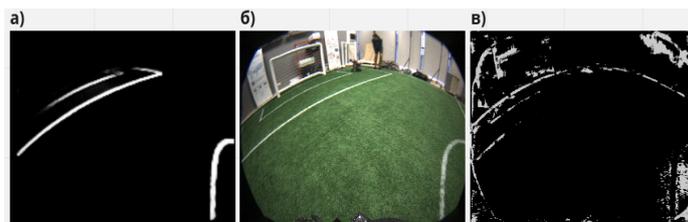


Рис. 3. а) Маска YoloV8; б) исходное изображение с камеры робота; в) Маска YOEO.

Таким образом, YoloV8 была более предпочтительна, чем YOEO. Она удовлетворяла поставленным требованиям, поэтому была встроена в проект SAHR в виде фильтра сегментации, принимающего исходное изображение с камеры робота и отправляющего на выход бинарную маску с выделенной разметкой.

## 2.5. Сравнение двух версий системы компьютерного зрения SAHR

Сравнение масок производилось попиксельно (0 - пиксель не принадлежит классу разметки, 1 – пиксель принадлежит классу разметки). Для оценки были выбраны метрики: precision, recall, accuracy.

Была собрана тестовая выборка на разных уровнях освещенности. Всего использовалось 5 уровней освещенности: 10-20 лк, 90-120 лк, 400-500 лк, 550-750 лк, 900-1200 лк. По 8 изображений в каждой из областей. Датасет из 45 изображений был пропущен через текущую реализацию получения маски поля с разметкой и новую реализацию получения маски поля с разметкой (реализации показаны на рис. 4).

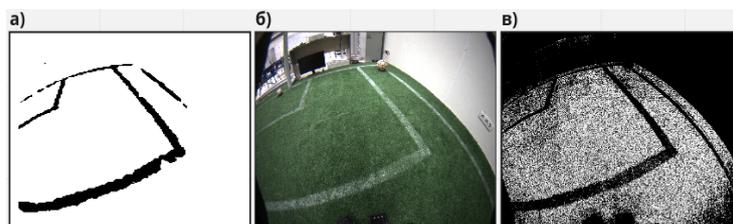


Рис. 4. а) маска из фильтра сегментации; б) исходное изображение; в) маска из Фильтра зеленого.

По итогам сравнения было выявлено, что модель видит не все линии разметки, но количество ложной информации о ней много меньше. Это позволяет модулю зрения снизить вероятность ошибки и, тем самым, работать с много меньшей погрешностью. Кроме того, качество модели сегментации улучшается сбором большего количества данных. В то время как текущая реализация уже на своем пике.

Качественная оценка нового подхода к распознаванию разметки показала, что прорисованные линии располагаются не посередине разметки, а на краю. Это говорит о том, что погрешность новой реализации при постоянном уровне освещенности больше погрешности текущей реализации. Это плохо скажется на локализации робота. Несмотря на присутствующие недостатки, линии разметки действительно распознаются при разных уровнях освещения. Что позволяет говорить о большей обобщающей способности, а также о перспективах развития нового подхода. Кроме того, скорость работы обновленного модуля зрения составила 15.6 FPS, что является достаточным для текущей динамики игры. (на рис. 5 изображены результаты работы системы компьютерного зрения после внедрения фильтра сегментации).

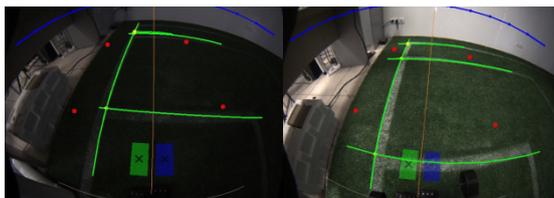


Рис. 5. Качественная оценка нового подхода к распознаванию разметки.

### 3. Заключение

В работе были описаны методы, применяющиеся в решении задач компьютерного зрения. Рассмотрены нейросетевые модели YOEO и YoloV8, решающие задачи детекции и сегментации. Проведено их сравнение и выбор YoloV8 на основе определенного перечня метрик. Она встроена в качестве фильтра модуля зрения SAHR и добавлена в последовательность фильтров, которая распознает разметку.

Встроенная новая последовательность включает в себя всего четыре фильтра. Уменьшение количества потоков, выделяемых под каждый фильтр, приводит к более эффективной работе модуля зрения SAHR.

Кроме того, по результатам сравнения работы системы компьютерного зрения текущей версии и новой с внедренным фильтром сегментации устойчивость SAHR к изменению параметров освещения действительно удалось повысить. Однако, новая последовательность фильтров все еще нуждается в доработке. Она пока не способна заменить предыдущее решение.

Требуется применить дополнительную фильтрацию маски сегментации. Обучить нейросетевую модель YoloV8 на большем количестве данных, что поможет увеличить распознаваемость разметки. Кроме того, будет включена автоподстройка яркости на камере, поэтому возможный диапазон яркости будет динамическим. Тогда модуль зрения SAHR будет готов к игре в условиях динамически меняющейся освещенности.

Таким образом, реализованный в данной работе переход на нейросетевые методы улучшил обобщающую способность модуля зрения SAHR и сократил количество используемых фильтров.

### Список литературы

1. Chengyin Hu, Weiwen Shi, Chao Li, Jialiang Sun, Donghua Wang, Junqi Wu and Guijian Tang // Impact of Light and Shadow on Robustness of Deep Neural Networks. 2023. <https://arxiv.org/abs/2305.14165>.
2. Vahl F. , Gutsche J. , Bestmann M. , Zhang J. YOEO–You Only Encode Once: A CNN for Embedded Object Detection and Semantic Segmentation // 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO). Sanya, China, 2021. P. 619-624.
3. Irwan Bello, William Fedus, Xianzhi Du, Ekin D. Cubuk, Aravind Srinivas, Tsung-Yi Lin, Jonathon Shlens, Barret Zoph // Revisiting ResNets: Improved Training and Scaling Strategies, 2021. <https://arxiv.org/abs/2103.07579>.
4. Juan R. Terven, Diana M. Cordova-Esparaza // A COMPREHENSIVE REVIEW OF YOLO: FROM YOLOV1 AND BEYOND, 2023. <https://arxiv.org/abs/2304.00501>.
5. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun // Deep Residual Learning for Image Recognition, 2015. <https://arxiv.org/abs/1512.03385>.