

Исследование нейросетевых алгоритмов распознавания объектов железнодорожной инфраструктуры на видеоизображениях

Е.В. Медведева¹, А.А. Перовошикова¹

¹ Вятский государственный университет, 610000, Россия, г. Киров, ул. Московская, д. 36

Аннотация

В статье разработаны два нейросетевых алгоритма распознавания объектов железнодорожной инфраструктуры на видеоизображениях. Оба алгоритма направлены на повышение безопасности движения железнодорожного транспорта. Первый алгоритм – это детектирование объектов, создающих препятствия, и объектов железнодорожной инфраструктуры. Второй алгоритм – семантическая сегментация главных и вспомогательных железнодорожных путей, а также поездов в пределах видимости следования локомотива. Алгоритмы реализованы на основе сверточных нейронных сетей YOLO и U-Net. Для обучения и тестирования СНС использована база изображений НИИ информатизации, автоматизации и связи на железнодорожном транспорте. Проведены экспериментальные исследования, направленные на повышение эффективности алгоритмов обнаружения и сегментации объектов за счет методов аугментации данных, дополнительной предобработки, выбора архитектуры и наилучших гиперпараметров сети. Алгоритм обнаружения работает в режиме реального времени, достигаемая средняя точность для 11 классов объектов по метрике mAP – 64 %. Скорость работы алгоритма семантической сегментации составляет 5 кадр/с, средняя точность для трех классов объектов по метрике IoU – 92 %.

Ключевые слова: обнаружение объектов, семантическая сегментация, объекты железнодорожной инфраструктуры, безопасность движения железнодорожного транспорта, системы технического зрения, нейросетевые алгоритмы.

Цитирование: Медведева, Е.В. Исследование нейросетевых алгоритмов распознавания объектов железнодорожной инфраструктуры на видеоизображениях / Е.В. Медведева, А.А. Перовошикова // Компьютерная оптика. – 2025. – Т. 49, № 3. – С. 443-450. – DOI: 10.18287/2412-6179-CO-1563.

Citation: Medvedeva EV, Perevoshchikova AA. Research of neural network algorithms for recognizing railway infrastructure objects in video images. Computer Optics 2025; 49(3): 443-450. DOI: 10.18287/2412-6179-CO-1563.

Введение

В настоящее время как в России, так и за рубежом активно ведутся работы по разработке и совершенствованию интеллектуальных систем для повышения безопасности движения железнодорожного транспорта.

Мониторинг железнодорожной обстановки осуществляется на основе анализа данных видеоизображений, поступающих от различных сенсоров (оптических с разным фокусным расстоянием, инфракрасных, LiDAR, радаров миллиметрового диапазона) комплекса систем технического зрения (СТЗ), устанавливаемых в голове поезда и при необходимости на самих путях в особо опасных местах [1–4].

К задачам, направленным на повышение безопасности движения поездов, относятся: контроль свободности рельсового пути, обнаружение препятствий и расстояний до них, обнаружение цветковых сигналов светофоров и определение их показаний, распознавание стрелочных переводов и их положений с высокой точностью вне зависимости от времени суток, погодных условий, скорости движения поезда и т.п. [1, 2].

Решение этих задач актуально для автоматизации движения поездов в части реализации технического зрения локомотива [1].

Основными требованиями к решению данных задач являются высокая точность распознавания (обнаружения, сегментации) объектов и скорость работы алгоритмов в режиме реального времени. Также следует отметить, что для поездов характерен значительный тормозной путь, поэтому препятствия требуется обнаруживать на большем расстоянии [1].

Данная работа направлена на решение двух задач, связанных с применением СТЗ для повышения безопасности движения поездов. Первая задача – это обнаружение элементов железнодорожной инфраструктуры, определяющих движение поездов (рельсы, стрелочные переводы, светофоры), а также посторонних объектов, создающих препятствия (людей, машин, вагонов). Вторая задача связана с семантической сегментацией главных и вспомогательных железнодорожных путей, а также определением местоположения поездов в пределах видимости следования локомотива.

Принимая во внимание, что данные задачи должны решаться при разных погодных условиях, уровнях

освещенности, сложном и быстро меняющемся фоне, то классическими алгоритмами обнаружения и сегментации (на основе выделения контуров или однородных областей, морфологических операций, кластеризации, методов оценки оптического потока и др.) решить эти задачи с высокой точностью обнаружения и скоростью работы практически невозможно [4–6].

Наиболее перспективным подходом для решения данных задач являются алгоритмы на основе сверточных нейронных сетей (СНС). Особенность данного подхода заключается в том, что признаки изображений в СНС формируются не самим исследователем, а за счет операции двумерной свертки, при этом используемые фильтры, являющиеся основой СНС, формируются в процессе обучения. Сгенерированные таким образом дескрипторы, как правило, позволяют достичь высокой точности обнаружения и классификации целевых объектов [1–4, 7, 8].

Целью работы является разработка алгоритмов обнаружения и семантической сегментации объектов железнодорожной инфраструктуры на видеоизображениях на основе СНС и проведение исследований, направленных на повышение точности работы алгоритмов в режиме реального времени.

Статья построена следующим образом. В параграфе 1 представлен обзор работ по данной тематике. В параграфе 2 приводится описание набора данных, используемого для обучения сети, а также метрик оценки качества работы алгоритмов. В параграфе 3 представлены экспериментальные исследования алгоритма обнаружения потенциально опасных объектов по пути следования локомотива и элементов железнодорожной инфраструктуры, определяющих движение поездов. В параграфе 4 приведены экспериментальные исследования алгоритма семантической сегментации главных и вспомогательных железнодорожных путей, а также поездов в пределах видимости следования локомотива. Параграф 5 содержит выводы.

1. Обзор публикаций по тематике работы

Решением задач, связанных с применением СТЗ для повышения безопасности движения железнодорожного транспорта, занимаются огромное количество российских и зарубежных ученых и инженеров, о чем свидетельствует значительное количество работ.

Например, в работе [9] рассмотрена модификация сети VGG16 для контроля и диагностики состояния рельсовых стыков и выявления их дефектов. Экспериментальные исследования сети показали достижение точности классификации стыков по метрике «Assugas» не менее 96 %.

В работе [10] предложен способ автоматического контроля объектов железнодорожной инфраструктуры за счёт оснащения локомотива технологией машинного зрения, а именно возможность визуального

контроля показаний сигналов путевых светофоров по ходу его следования. Алгоритм распознавания путевых сигналов на фиксируемом изображении базируется на преобразовании фиксируемого кадра из RGB в пространство HSV и применении модели СНС. По результатам обучения разработанной СНС получена оценка точности отклика для тестовой выборки – 87,15 %.

В ряде работ [3, 4, 11] предлагаются методы обнаружения посторонних объектов, представляющих опасность на железнодорожных путях. В работе [11] для обнаружения препятствий предложено использовать модель «MobileNetV2». Модель реализована на специальном наборе данных, состоящем только из изображений яркого солнечного дня и содержащем три класса объектов (железнодорожные пути, люди, животные). Для представленного набора данных точность обнаружения препятствий по метрике «Assugas» – 97 %. В работе [4] для обнаружения препятствий предложен детектор объектов на основе введения новой параллельной структуры с повышающей дискретизацией и модуля извлечения контекста в архитектуру R-CNN. При размере кадров 1330×800 пкс скорость работы алгоритма достигает 11 кадров в секунду, а средняя точность обнаружения объектов для девяти классов – 90,6 %. В работе [3] рассматривается метод выделения участка железнодорожного пути и обнаружение на нем посторонних объектов с помощью модели YOLOv5. В отличие от других работ рассматриваются оценки рисков поведения вторжения посторонних объектов. Для событий с высоким риском, которые могут привести к аварии, своевременно выдаются сигналы предупреждения. Средняя точность обнаружения объектов трех классов (человек, грузовик, корова) по метрике mAP составляет 0,882, время обработки изображения – 26 мс.

В работах [1,12] рассмотрены методы семантической сегментации объектов железнодорожной инфраструктуры. В рамках работ АО «НИИАС» [1] по семантической сегментации искусственный интеллект определяет железнодорожный путь и вагоны. Точность сегментации двух классов объектов по метрике mIoU составляет около 92–94 %. Высокая точность достигается за счет использования значительно большего датасета. В работе [12] предложен алгоритм сегментации участков железнодорожного пути для обнаружения вторжений в режиме реального времени и на основе легкой нейронной сети (с параметрами 0,5 млн) и декодера кросс-слияния. Точность на сформированном авторами наборе данных по метрике mIoU достигает 92,4 %.

Данных, полученных только с одного типа датчика, часто недостаточно, чтобы обеспечить высокую точность детектирования объектов, особенно в условиях плохой видимости. Для достижения лучших характеристик обнаружения объектов железнодорожной инфраструктуры и людей в работе [13] предлага-

ется использовать совместную обработку данных от датчиков с разной физической сущностью. Методы объединения данных могут быть реализованы на основе архитектур Auto Encoder (AE), Convolution Neural Network (CNN) и Generative Adversarial Network (GAN).

Алгоритм распознавания может быть построен путем последовательных операций слияния и распознавания, или алгоритм слияния может быть встроен в процесс распознавания.

Различные решения задач, связанные с распознаванием объектов, различаются точностью и скоростью работы алгоритмов. Качество решений задач зависит от объема набора данных для обучения сети, количества классов объектов, их распределения в наборе данных, размеров объектов, выбора архитектуры сети, настроек параметров сети. Поэтому методы сегментации и обнаружения объектов железнодорожной инфраструктуры, а также объектов, представляющих собой опасность на железнодорожных путях, по-прежнему нуждаются в постоянном совершенствовании и оптимизации с точки зрения эффективности модели, реального времени и адаптивности к различным условиям. В связи с этим данная работа направлена на повышение эффективности алгоритмов обнаружения и сегментации объектов за счет методов аугментации данных, дополнительной предобработки, выбора архитектуры и наилучших гиперпараметров сети.

2. Набор данных и метрики

В данной работе для обучения и тестирования алгоритмов обнаружения и сегментации использовался набор данных, предложенный НИИ информатизации, автоматизации и связи на железнодорожном транспорте.

Набор данных включает в себя:

- 8203 изображения в формате png (2688×1512 и 1920×1080 px);
- 8203 файла в формате json с разметкой для задачи обнаружения;
- 8203 маски в формате png для задачи сегментации;
- 1000 неразмеченных изображений для тестирования.

Изображения для обнаружения содержат 11 классов объектов:

- автомобиль (car);
- человек (human);
- вагон (wagon);
- стрелочные переводы по ходу движения влево (facingSwitchL), вправо (facingSwitchR) и вне видимости (facingSwitchNV);
- стрелочные перевод против хода движения влево (trailingSwitchL), вправо (trailingSwitchR) и вне видимости (trailingSwitchNV);

- разрешающий (signalE) и запрещающий (signalF) сигналы светофора.

Для задачи сегментации размечаемыми объектами на изображениях являлись:

- главный железнодорожный полигон (MainRailPolygon);
- вспомогательный железнодорожный полигон (AlternativeRailPolygon);
- поезд (Train).

Обучение алгоритмов обнаружения и сегментации производилось в облачном сервисе Google Colab на графическом ускорителе NVidia Tesla P100-PCIe-16 GB. Объем тренировочной выборки состоял из 5742 изображений, валидационной выборки – из 1230 изображений.

Для оценки точности работы алгоритмов обнаружения и сегментации использовались общепринятые метрики: полнота (recall) и точность (precision). В алгоритме обнаружения данные метрики использовались для вычисления основных метрик: средней точности и усредненной по всем категориям величины средней точности. А в алгоритме сегментации – для оценки величины, определяемой как отношение площадей пересечения к объединению рамок, ограничивающих объекты (IoU, Intersection over Union).

Метрики полноты (R) и точности (P) вычислялись по формулам:

$$R = \frac{TP}{TP + FN}, \quad (1)$$

$$P = \frac{TP}{TP + FP}, \quad (2)$$

где TP соответствует числу пикселей верно определенных объектов, FN – числу пикселей, относящихся к объектам интереса, но алгоритм их к таковым не относит, FP – числу пикселей, которые алгоритм относит к объектам интереса, но они таковыми не являются.

Для оценки метрики средней точности (AP) применялось выражение:

$$AP = \frac{\sum_{r=1}^N P_r}{N}, \quad (3)$$

где P_r – значение точности, соответствующее определенному значению полноты r , а N – множество различных значений полноты. Оно получается путем варьирования порога оценки достоверности в оцениваемом алгоритме обнаружения. Эта метрика используется, если алгоритм применяется для обнаружения объектов только одного класса.

В случае применения алгоритма для обнаружения объектов нескольких классов используется среднее значение AP , усредненное по всем классам объектов:

$$mAP = \frac{\sum_{i=1}^K AP_i}{K}, \tag{4}$$

где K – число классов объектов.

Метрика IoU вычислялась по формуле:

$$IoU = \frac{S(A \cap B)}{S(A \cup B)} = \frac{TP}{TP + FP + FN}, \tag{5}$$

где $S(\cdot)$ – площадь, A и B – рамки, ограничивающие объекты, соответствующие экспертной разметке и предсказанные алгоритмом.

3. Алгоритм обнаружения объектов

Для решения задачи обнаружения могут применяться детекторы на основе сверточных нейронных сетей (СНС) архитектур R-CNN [4,14], YOLO [3], MobileNetV2 [11], SSD [15,16], FPN [17]. Возможные решения различаются точностью обнаружения объектов по классам и средней точностью, скоростью работы и затрачиваемыми вычислительными мощностями на работу алгоритма.

Поскольку в решаемой задаче алгоритм должен анализировать видеопоток в режиме реального времени, то для задачи обнаружения объектов был выбран одноэтапный алгоритм на основе СНС YOLOv5 (You Look Only Once).

Основой архитектуры СНС YOLOv5 является экстрактор признаков (Backbone) и два полносвязных слоя (Neck, Head), которые отвечают за формирование предсказаний объектов, а также вероятностей принадлежности этих объектов к заданным классам [3]. Модель YOLOv5 позволяет добиться хороших соотношений по точности обнаружения и скорости работы (до 230 fps на NVIDIA RTX 4090 и до 190 fps на TESLA V100) по сравнению с сетями других архитектур, например, таких как SSD, R-CNN.

Для обучения модели использовалась библиотека yolov5 и предобученные веса yolov5. В начале работы модели выполнялась загрузка базы изображений и готовой разметки в виде масок для каждого изображе-

ния отдельно. Изначальная разметка представляет собой текстовый файл с названием класса объекта и значениями координат рамки, обрамляющей объект. Исходные данные приводятся к формату COCO, в котором указываются метки классов, координаты верхнего левого угла, ширина и высота рамки. В таком виде данные поступают на вход СНС для обучения.

После обучения модели выполнялась оценка потерь и принималось решение о повторном обучении. Если функция потерь удовлетворяла заданным пользователем условиям, то результаты сохранялись и обучение на этом заканчивалось.

Обучение длилось в течение 30 эпох. Время обучения модели составило 17 ч.

3.1. Экспериментальные исследования алгоритма обнаружения объектов

Для повышения точности обучения сети выполнялся ряд процедур: расширение базы данных, предобработка, экспериментальный выбор наилучшей топологии алгоритма и гиперпараметров сети.

Расширение базы изображений выполнялось за счет методов аугментации данных, таких как мозаика, масштабирование и поворот на несколько градусов. В процессе обучения на вход СНС YOLOv5 поступает батч размером 1920×1920 пкс с девятью изображениями различного размера. Благодаря аугментации в каждый батч попадает большое разнообразие классов объектов, что позволяет повысить скорость обучения без увеличения вычислительных мощностей.

Дополнительная аугментация данных и дообучение модели проводились для объектов небольших размеров и редких классов, таких как стрелочные переводы.

В работе проводилось сравнение нескольких моделей семейства YOLOv5 (YOLOv5m, YOLOv5l и YOLOv5x), которые различаются количеством внутренних слоёв и параметров. Экспериментальные графики точности (precision), полноты (recall) и потерь (loss) от числа эпох (epoch) для трех данных моделей представлены на рис. 1.

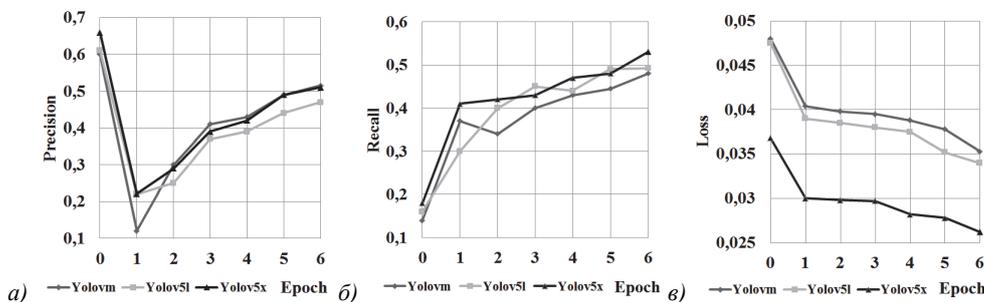


Рис. 1. Графики точности (а), полноты (б) и потерь (в) для трех моделей YOLOv5

Анализ графиков показывает, что модель YOLOv5l работает хуже, чем YOLOv5m, несмотря на повышенное количество слоев и параметров. Наилучшие результаты по точности и полноте, самые

низкие потери, а также высокую скорость, несмотря на увеличение веса модели показывает YOLOv5x. На тестовой выборке модель YOLOv5x позволила получить результаты на 5% лучше, чем YOLOv5m.

В ходе тестирования модели были трудности с определением сигналов светофора. Причина – небольшое количество примеров, искажение цветов из-за разного времени суток, туманов, дождей и т.п. Для улучшения различия цветов и повышения точности обнаружения сигналов светофоров выполнялось преобразование изображений из цветового пространства RGB в HSV с подбором параметров оттенка (H), насыщенности (S) и яркости (V). В результате были выбраны наиболее подходящие параметры HSV, нормализованные в диапазоне от 0 до 1: $H=0$, $S=0,85$, $V=0,6$.

Для повышения производительности и стабилизации работы СНС применялся метод пакетной нормализации (Batch normalization) [18]. Активным указывался параметр SyncBatchNorm (Synchronized Batch Normalization), который в отличие от обычного параметра BatchNorm рассчитывает среднее и дисперсию пакета нормализации не только внутри каждого GPU, а синхронно по всем графическим процессорам. Это позволяет бороться с переобучением, а также достичь более высокой точности и сходимости нейросети.

Следующим шагом, направленным на повышение скорости и точности работы модели, был сравнительный анализ рекомендуемых к применению оптимизаторов: стохастического градиентного спуска (SGD) с накоплением импульса и адаптивной оценки момента (Adam). Метод SGD основан на итерационной оптимизации целевой функции с подходящими свойствами гладкости. При использовании SGD с накоплением импульса учитываются значения предыдущих градиентов, сохраняется общее направление, и траектория остаётся правильной. Метод Adam основан на расчете экспоненциальной средней градиента и экспоненциального среднего квадратов градиента. Каждая экспоненциальная средняя имеет свой гиперпараметр, определяющий период усреднения. Данный подход позволяет снизить скорость обучения часто обновляемых параметров и в большей степени учитывать последние значения градиента, которые характеризуют текущее состояние модели.

На рис. 2 показаны графики оценки функции потерь для тренировочной (а) и валидационной (б) выборок для моделей с оптимизаторами SGD и Adam.

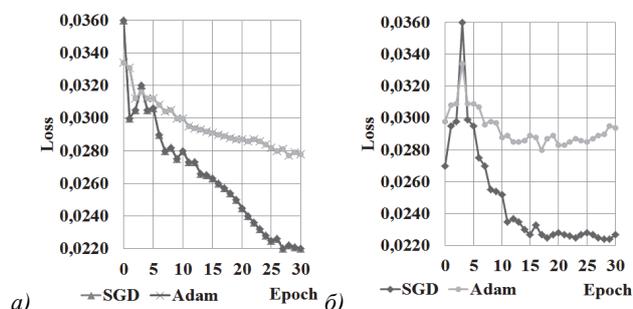


Рис. 2. Графики потерь от числа эпох для тренировочной (а) и валидационной (б) выборок

Лучший результат по оценке точности потерь показал оптимизатор SGD, и его последующее использование позволило избежать переобучения сети.

3.2. Результаты экспериментов

Для оценки эффективности детектирования объектов железнодорожной инфраструктуры использовались метрики полноты (R) и точности (P), с их помощью вычислялась основная метрика оценки качества mAP. Тестирование проводилось на изображениях, полученных при разных погодных условиях, в том числе с недостаточной видимостью. Оценка результатов обучения на тестовой выборке показана в табл. 1. В колонке «Исх.» приведены результаты для исходной модели YOLOv5x, а в колонке «Мод.» – результаты для улучшенной (модифицированной) модели за счет аугментации данных, цветокоррекции и выбора наилучших гиперпараметров сети.

Табл. 1. Сравнение результатов тестирования исходной и улучшенной модели

Класс	P		R		mAP	
	Исх.	Мод.	Исх.	Мод.	Исх.	Мод.
car	0,67	0,88	0,75	0,84	0,77	0,86
human	0,67	0,77	0,66	0,83	0,68	0,75
wagon	0,77	0,90	0,76	0,85	0,79	0,90
facingSwitchL	0,41	0,73	0,58	0,66	0,43	0,60
facingSwitchR	0,46	0,61	0,43	0,71	0,36	0,64
facingSwitchNV	0,39	0,43	0,30	0,51	0,24	0,33
trailingSwitchL	0,51	0,63	0,69	0,70	0,60	0,62
trailingSwitchR	0,49	0,67	0,54	0,57	0,50	0,55
trailingSwitchNV	0,35	0,36	0,29	0,40	0,23	0,39
signalE	0,56	0,75	0,48	0,66	0,44	0,68
signalF	0,48	0,77	0,54	0,68	0,43	0,72

Из результатов, приведенных в табл. 1, видно, что проведенные процедуры позволили значительно повысить точность обнаружения объектов по классам. Наибольший прирост по метрике mAP наблюдается для объектов «сигналы светофоров» (signalE, signalF) – до 29%. Средняя точность улучшенного алгоритма обнаружения крупных объектов: автомобилей (car), людей (human), вагонов (wagon) по метрике mAP составила 85%.

Наихудшие результаты по точности обнаружения имеют стрелочные переводы вне видимости из-за их небольшого размера и возможного их принятия за стрелочные переводы влево и вправо.

На рис. 3 показан пример детектирования объектов железнодорожной инфраструктуры на тестовом наборе: а – исходное изображение, б – экспертная разметка, в – результаты работы модели с указанием значений уверенности, с которой СНС относит объект к определенному классу.

На рис. 4 представлены примеры работы алгоритма на неразмеченном тестовом наборе в светлое (а) и темное (б) время суток.

Визуальная субъективная оценка показывает, что алгоритм успешно обнаруживает объекты на тестовых изображениях, так как при обучении удалось извлечь все необходимые для детектирования признаки.



Рис. 3. Пример детектирования объектов железнодорожной инфраструктуры: (а) исходное изображение, (б) экспертная разметка, (в) результаты работы модели

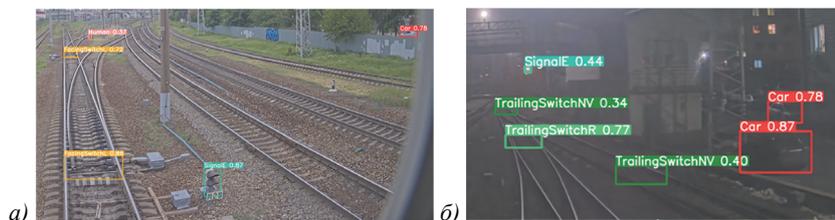


Рис. 4. Пример детектирования объектов на тестовых изображениях: (а) в светлое время суток; (б) в темное время суток

Алгоритм работает в режиме реального времени, время обработки: 1231 изображения размером 1920×1088 пкс составляет 50 секунд (24,6 кадр/с).

4. Алгоритм семантической сегментации

Семантическая сегментация применяется для детектирования железнодорожного пути и является основой для последующих алгоритмов обнаружения и классификации объектов [1]. Результатом семантической сегментации в отличие от детектирования объектов являются не параметры ограничивающего прямоугольника и метки, а маркировка пикселей изображения в соответствии с определенным классом. В связи с этим для реализации алгоритма сегментации требуется более сложная структура сети. Для решения задачи сегментации могут применяться архитектуры U-Net [7,8,19], Mask R-CNN [20], SegNet [21], PSPNet [22], DeepLabv3 [23] и др. Наиболее широкое применение в задачах семантической сегментации получила CHC U-Net. Базовая архитектура U-net состоит из путей сжатия и расширения и выполняет следующую последовательность действий: пошагово уменьшает размерность изображения с некоторыми преобразованиями и затем восстанавливает предсказанные маски из сжатого изображения. Основными достоинствами модели U-Net являются: возможность работы с небольшим количеством примеров, высокие показатели точности и скорости работы.

4.1. Экспериментальные исследования алгоритма семантической сегментации

Задачей алгоритма сегментации являлась разметка трех классов объектов: поездов, главных и вспомогательных железнодорожных путей. Для обучения модели использовалась библиотека unet, предобученные веса и набор данных.

При обучении на первых этапах возникали трудности при выделении главного полигона из вспомогательных путей. Повышение точности и производи-

тельности в процессе обучения выполнялось за счет выбора наилучших гиперпараметров: оптимального числа эпох, размера батча, размера изображения, оптимизатора и функции активации.

Оптимизатором был выбран AdamW. Он эффективен в вычислительном отношении, имеет небольшие требования к памяти, инвариантен к диагональному масштабированию градиентов и хорошо подходит для больших задач с точки зрения данных.

Также в работе были проведены сравнения нескольких функций активаций для заданного набора данных. Сравнение выполнялось по функции потерь, вычисляемой по формуле:

$$Dice\ loss = 1 - \frac{2TP}{2TP + FP + FN} \tag{6}$$

Результаты оценки потерь для разных функций активаций представлены в табл. 2.

Табл. 2. Оценки потерь для разных функций активаций

Функция активации	Потери
Softmax	0,0845
Sigmoid	0,0858
Tanh	0,0750
PReLU	0,0635
LeakyReLU	0,0774

В качестве функции активации выбрана функция LeakyReLU, которая обеспечивает меньшие потери на заданном наборе данных, чем Softmax и Sigmoid, менее требовательна к вычислительным ресурсам, чем функция Tanh, и не требует тонкой настройки архитектуры модели по сравнению с PReLU.

4.2. Результаты экспериментов

На рис. 5 приведены зависимости точности (а) и потерь (б) от числа эпох (epoch) на валидационной выборке, которые свидетельствуют о корректности работы алгоритма на тестовых изображениях. Сред-

няя точность по всем классам объектов по метрике IoU составила 92 % (рис. 5а).

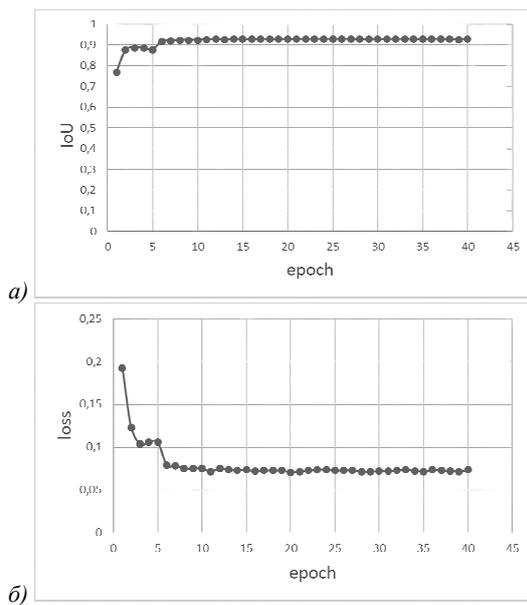


Рис. 5. Зависимости точности (а) и потерь (б) от числа эпох

Оценка результатов тестирования по классам приведена в табл. 3. В колонке «Исх.» представлены оценки метрики IoU для исходной модели U-Net, а «Мод.» – для улучшенной модели.

Табл. 3. Результаты оценки точности сегментации

Класс	IoU	
	Исх.	Мод.
"MainRailPolygon" (главный ЖД-полигон)	0,663	0,811
"AlternativeRailPolygon" (вспомогательный ЖД-полигон)	0,864	0,929
"Train" (поезд)	0,736	0,798

Улучшенный алгоритм существенно повысил точность сегментации объектов. Наибольший прирост по метрике IoU наблюдается для объекта «главный ЖД-полигон» – 14,8 %.

Пример сегментации изображений для светлого и темного времени суток показан на рис. 6. Здесь представлены исходные изображения, разметки экспертом и результаты сегментации.

Из приведенного примера видно, что алгоритм сегментации успешно выделяет объекты интереса на тестовых изображениях: главный и вспомогательные пути, а также поезда.

Время работы алгоритма сегментации последовательности из 1000 изображений размером 1920×1088 пкс занимает 3 минуты (5,6 кадр/с).

Заключение

Разработаны два алгоритма распознавания объектов железнодорожной инфраструктуры на изображениях: детектирование объектов, создающих препятствия, и объектов железнодорожной инфраструкту-

ры, а также семантическая сегментация главных и вспомогательных железнодорожных путей и поездов по ходу движения локомотива. Алгоритмы реализованы на основе СНС-архитектур YOLOv5 и U-Net. Алгоритмы работают в режиме реального времени. Средняя точность по 11 классам алгоритма детектирования по метрике mAP составляет 64 %, а алгоритма сегментации для трех классов по метрике IoU – 92 %.

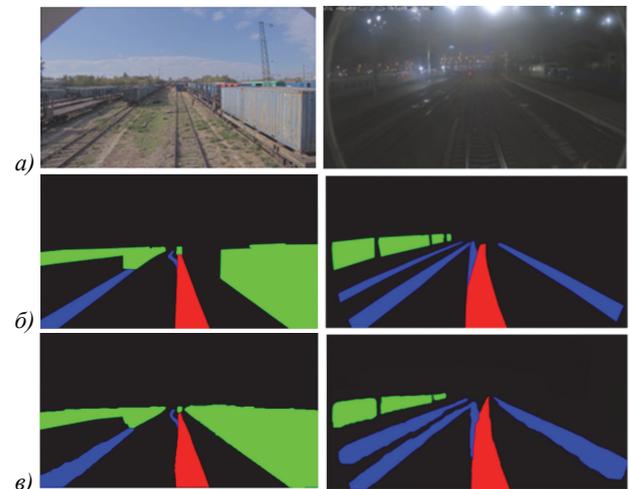


Рис. 6. Пример сегментации: а) исходные изображения, б) эталонная сегментация (маски), в) результат сегментации: светло-серый цвет – главный железнодорожный полигон, темно-серый – вспомогательные, белый – поезда

В качестве путей совершенствования можно рассмотреть последние версии модели YOLO v8, v9, решающие не только задачу детектирования объектов, но и задачу сегментации, а также позволяющих достичь более высоких соотношений по точности обнаружения и скорости работы алгоритма. Более высоких показателей по точности детектирования объектов также можно достичь за счет увеличения набора данных для обучения сети и комплексирования данных от разных сенсоров комплекса систем технического зрения.

References

- [1] Popov PA. Application of artificial intelligence technologies for railway transport [In Russian]. Bulletin of the Institute for Problems of Natural Monopolies: Railway Engineering 2024; 1(65): 38-41.
- [2] Ozerov AV, Okhotnikov AL. Computer vision as part of advanced train protection system [In Russian]. Intelligent transport systems. Materials of the II International Scientific and Practical Conference 2023: 620-625. DOI 10.30932/9785002182794-2023-620-625.
- [3] Chen P, Zhang Z, Huang Y, Dai L, Xu F, Hu H. Railway obstacle intrusion warning mechanism integrating YOLO-based detection and risk assessment. J Ind Inf Integr 2024; 38: 100571. DOI: 10.1016/j.jii.2024.100571.
- [4] He D, Ren R, Li K, Zou Z, Ma R, Qin Y, Yang W. Urban rail transit obstacle detection based on Improved R-CNN. Measurement 2022; 196: 111277. DOI: 10.1016/j.measurement.2022.111277.

- [5] Medvedeva EV, Kurbatova EE. Image segmentation based on two-dimensional markov chains. In Book: Favorskaya MN, Jain LC, eds. Computer vision in Control Systems-2. Innovations in practice. Cham, Switzerland: Springer International Publishing Switzerland; 2015: 277-295. DOI: 10.1007/978-3-319-11430-9_11.
- [6] Silar Z, Dobrovolny M. Utilization of directional properties of optical flow for railway crossing occupancy monitoring. 2013 Int Conf on IT Convergence and Security (ICITCS), 2013, pp. 1-4. DOI: 10.1109/ICITCS.2013.6717896.
- [7] Bagaev S, Medvedeva E. Experimental assessment of accuracy of multiclass segmentation of objects from satellite images based on a modified convolutional neural network U-net [In Russian]. Current Problems in Remote Sensing of the Earth from Space 2021; 18(6): 35-45. DOI: 10.21046/2070-7401-2021-18-6-35-45.
- [8] Bagaev S, Medvedeva E. Segmentation of satellite images of the earth's surface using neural network technologies. 28th Conf of Open Innovations Association (FRUCT) 2021: 15-21. DOI: 10.23919/FRUCT50888.2021.9347653.
- [9] Orlov SP, Efimushkin NA, Efimushkina NV. Deep neural network for diagnosing railway track elements [In Russian]. Vestnik of Samara State Technical University. Technical Sciences Series 2022; 30(1): 63-72. DOI: 10.14498/tech.2022.1.4.
- [10] Minakov VA, Fomenko VK. Machine vision technology on locomotives for identifying track signals [In Russian]. World of Transport 2019; 17(6): 62-72. DOI: 10.30932/1992-3252-2019-17-62-72.
- [11] Rahman F, Ahmed Md, Hasan Md, Jahan N. Real-time obstacle detection over railway track using deep neural networks. Procedia Comput Sci 2022; 215: 289-298. DOI: 10.1016/j.procs.2022.12.031.
- [12] Chen Z, Yang J, Chen L, Feng Z, Jia L. Efficient railway track region segmentation algorithm based on lightweight neural network and cross-fusion decoder. Autom Constr 2023; 155: 105069. DOI: 10.1016/j.autcon.2023.105069.
- [13] Zeng Y, Wang X, Zhao H, Jin Y, Giannopoulos GA, Lia Y. Image fusion methods in high-speed railway scenes: A survey. High-Speed Railway 2023; 1(2): 87-91. DOI: 10.1016/j.hspr.2023.01.001.
- [14] Mahendrakar T, Ekblad A, Fischer N, White R, Wilde M, Kish B, Silver I. Performance study of YOLOv5 and Faster R-CNN for autonomous navigation around non-cooperative targets. 2022 IEEE Aerospace Conf (AERO) 2022: 1-12. DOI: 10.1109/AERO53065.2022.9843537.
- [15] Rahmani W, Hernawan A. Real-time human detection using deep learning on embedded platforms: A review. J Robot Control 2021; 2(6): 462-468. DOI: 10.18196/jrc.26123.
- [16] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg A-C. SSD: Single shot multibox detector. In Book: Leibe B, Matas J, Sebe N, Welling M, eds. Computer Vision – ECCV 2016. 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I. Cham, Switzerland: Springer International Publishing AG; 2016: 21-37. DOI: 10.1007/978-3-319-46448-0_2.
- [17] Yang G, Wang Z, Zhuang S. PFF-FPN: A parallel feature fusion module based on FPN in Pedestrian detection. 2021 Int Conf on Computer Engineering and Artificial Intelligence (ICCEAI) 2021: 377-381. DOI: 10.1109/ICCEAI52939.2021.00075.
- [18] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ICML'15: Proc 32nd Int Conf on International Conference on Machine Learning 2015; 37: 448-456.
- [19] Zhang H, Jiang Z, Zheng G, Yao X. Semantic segmentation of UAV remote sensing images based on improved U-Net. 2023 8th Int Conf on Intelligent Computing and Signal Processing (ICSP) 2023: 1735-1740. DOI: 10.1109/ICSP58490.2023.10248526.
- [20] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV) 2017: 2961-2969. DOI: 10.1109/ICCV.2017.322.
- [21] Khruvakhray S, Srinonchat J. Railway track detection based on SegNet deep learning. TENCON 2023 – 2023 IEEE Region 10 Conf (TENCON) 2023: 409-413. DOI: 10.1109/TENCON58879.2023.10322378.
- [22] Seeli D, Thanammal KK. Brain disease detection using PSPNet for multiclass segmentation. 2nd Int Conf on Applied Artificial Intelligence and Computing (ICAAIC) 2023: 1030-1034. DOI: 10.1109/ICAAIC56838.2023.10140366.
- [23] Zhang Y, Zhang Y, Zhang Q. Semantic segmentation of traffic scene based on DeepLabv3+ and attention mechanism. 3rd Int Conf on Neural Networks, Information and Communication Engineering (NNICE) 2023: 542-547. DOI: 10.1109/NNICE58320.2023.10105805.

Сведения об авторах

Медведева Елена Викторовна, окончила Кировский политехнический институт (1991 г.) по специальности «Инженер-конструктор-технолог ЭВА». Доктор технических наук (2012 г.), профессор кафедры радиоэлектронных средств Вятского государственного университета. Область научных интересов: цифровая обработка сигналов и изображений. E-mail: emedv@mail.ru ORCID: <https://orcid.org/0000-0002-0677-1418>

Перевощикова Анастасия Александровна, окончила Вятский государственный университет (2023 г.) по направлению 10.05.02 «Информационная безопасность телекоммуникационных систем». Область научных интересов: техническое зрение, методы машинного обучения. E-mail: frlqwq@gmail.com

ГРНТИ: 28.23.37

Поступила в редакцию 12 мая 2024 г. Окончательный вариант – 24 июля 2024 г.

Research of neural network algorithms for recognizing railway infrastructure objects in video images

E.V. Medvedeva¹, A.A. Perevoshchikova¹

¹ Vyatka State University, Moskovskaya Str. 36, Kirov, 610000, Russia

Abstract

The article describes the development of two neural network algorithms for recognizing objects of the railway infrastructure in video images. Both algorithms are aimed at improving railway traffic safety. One algorithm detects foreign objects on railway tracks and objects relating to the railway infrastructure. The other algorithm implements the semantic segmentation of main and auxiliary railway tracks, as well as trains within the visible range of the locomotive. The algorithms are implemented based on convolutional neural networks (CNN) YOLO and U-Net. The CNN is trained and tested using the image database of the Research Institute of Information, Automation and Communications in Railway Transport. The experimental studies conducted are aimed at increasing the efficiency of algorithms for object detection and segmentation through the use of data augmentation methods and additional preprocessing, as well as selecting an architecture and optimal network hyperparameters. The detection algorithm works in real time, achieving an average accuracy of 64% for 11 object classes according to the mAP metric. The operating speed of the semantic segmentation algorithm is 5 frames/s, the average accuracy for three classes of objects according to the IoU metric is 92%.

Keywords: object detection, semantic segmentation, railway infrastructure objects, railway traffic safety, machine vision systems, neural network algorithms.

Citation: Medvedeva EV, Perevoshchikova AA. Research of neural network algorithms for recognizing railway infrastructure objects in video images. *Computer Optics* 2025; 49(3): 443-450. DOI: 10.18287/2412-6179-CO-1563.

Author's information

Elena Viktorovna Medvedeva, graduated from the Kirov Polytechnic Institute (1991) with a degree in “Engineer design-technologist of electronic computer equipment.” Doctor of Technical Sciences (2012), Professor of the Department of Radio Electronic Systems at Vyatka State University. Area of scientific interests: digital signal and image processing. E-mail: emedv@mail.ru ORCID: <https://orcid.org/0000-0002-0677-1418>

Anastasia Aleksandrovna Perevoshchikova, graduated from Vyatka State University (2023) with a degree in 10.05.02 “Information security of telecommunication systems.” Area of scientific interests: technical vision, machine learning methods. E-mail: f1lqwq@gmail.com

Received May 12, 2024. The final version – July 24, 2024.
