

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Пермский государственный национальный исследовательский университет»

**Физико-математический институт**

**ОТЧЕТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ  
на тему  
СВЕТОФОР, ОПРЕДЕЛЯЮЩИЙ ЗАГРУЖЕННОСТЬ ПОТОКА  
АВТОМОБИЛЕЙ И ПЕШЕХОДОВ С ПОМОЩЬЮ  
КОМПЬЮТЕРНОГО ЗРЕНИЯ, ДЛЯ АВТОМАТИЧЕСКОГО  
ПЕРЕКЛЮЧЕНИЯ СИГНАЛОВ**

Направление: 01.04.02 Прикладная математика и информатика

**Выполнил:**

магистр 2 курса

\_\_\_\_\_ **Ю. В. Санников**

**Научный руководитель:**

Директор института, доктор физико-  
математических наук

\_\_\_\_\_ **М.А. Барулина**

Пермь 2025

## РЕФЕРАТ

Научно исследовательская работа 13 с., 27 рис., 26 источн., 1 прил.

**НЕЙРОННЫЕ СЕТИ, СЕГМЕНТАЦИЯ ОБЪЕКТА НА ВИДЕО, ДЕТЕКЦИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ, УМНЫЙ СВЕТОФОР.**

Данная работа посвящена исследованию способов сегментации объекта на видео, с дальнейшей разработки модели используя нейронные сети.

Во введении представлено обоснование актуальности темы регулировки Дорожного движения за счет умного светофора с использованием моделей сегментации видео.

Первая глава содержит обзор сегментационных моделей и анализ предметной области.

Во второй главе содержится построение моделей и разработка класса для сегментации объектов в видеопотоке.

Во второй главе содержится анализ существующих решений с использованием компьютерного зрения в рассматриваемой области.

В третьей главе разработка класса для сегментации объектов в видеопотоке и проводятся сравнения сегментации объекта на рассмотренной модели.

В заключении подведен итог проделанной работы и приведены перспективы развития.

Приложение А содержит ссылку на репозиторий с исходным кодом программы.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	4
1. Анализ предметной области.....	7
1.1. Сегментирование с использованием нейронных сетей, обзор архитектур .....	7
1.2. Ultralytics YOLO.....	13
1.3. Обзор библиотеки cv2.....	16
1.3.1. Методы для работы с видео .....	17
2. Существующие решения проблемы трафика .....	18
2.1. Адаптивный светофор .....	18
2.2. Использование ML для детекции автомобилей в России .....	18
2.2.1. Видеодетектор автотранспортных средств ИНФОПРО и Видеодетектор транспорта РТК .....	19
2.2.2. НПО ИТС: Паук Трафик .....	20
2.3. Использование светофоров, управляемых ИИ, в странах Европы.....	21
3. Обучение и тестирование модели .....	22
3.1. Результаты поисков обучающего датасета.....	22
3.2. Результаты сегментации видео на предобученной модели .....	23
3.3. Направления для развития.....	26
ЗАКЛЮЧЕНИЕ .....	27
ГЛОССАРИЙ .....	28
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	29
Приложение А ссылка на репозиторий с кодом программы и результатами сегментации .....	32

## ВВЕДЕНИЕ

Данная работа посвящена исследованию возможностей умных светофоров, основанных на использовании искусственного интеллекта, с целью повышения пропускной способности дорожного движения.

В условиях современного города, пробки, особенно в час-пик, приводят к огромным экономическим, временным, психологическим и иным задержкам. По статистике из-за этой проблемы в крупных мегаполисах теряется около 40% от времени необходимого на поездку [1]. Для стран Европы и США ущерб, производимый скоплением автомобилей, может составлять миллиарды долларов. Например, в США на 2023 ущерб экономике от автомобильных пробок составил \$70,4 млрд [2]. В России, как и в остальных странах мира проблема пробок так же актуальна. За последние десять лет число легковых автомобилей в России выросло на 8.7 млн. Появление новых развязок и полос движения решает проблему в краткосрочном периоде, но из-за долгих сроков на строительство, стоимости этих работ и постоянно растущего трафика данное решение нельзя назвать оптимальным. Более дешевым решением является модернизация светофоров – простых регуляторов движения, не требующих масштабных перестроек. Современные светофоры работают по трехступенчатой схеме: часы пик, дневное время, ночь, но в случае непредсказуемости потока, данная система становится малоэффективной.

Исходя из исследований ученых пермского политеха, сокращение фаз светофора способствует сокращению заторов на 15-20%, а значит своевременное переключение фаз светофоров будет способствовать более высокой пропускной способности [3]. Существующие адаптивные светофоры частично решают проблему трехступенчатой системы, но они часто работают на устаревших технологиях и сильно зависят от множества физических датчиков, чувствительных к атмосферным условиям, что снижает их эффективность.

Представленное в текущей научно исследовательской работе решение заключается в разработке нейросетевых светофоров, которые будут более устойчивы к внешним влияниям и способны эффективно распознавать транспортные средства и пешеходов. Системы, основанные на искусственном интеллекте, могут самостоятельно анализировать дорожную ситуацию и оперативно переключать сигналы светофора в зависимости от загруженности. Это позволит минимизировать время ожидания на перекрестках и улучшить общую пропускную способность транспортных потоков.

В рамках текущей работы необходимо исследовать существующие решения в выбранной области, изучить существующие модели для сегментации видео, обработать полученный с камер наблюдения видеопоток передвижений транспорта и пешеходов.

Процесс исследования можно разбить на несколько этапов:

Этап сбора данных — получение необработанных данных о дорожном трафике, а именно видео с камер наблюдения.

Подготовительный этап — изучение существующих решений и моделей для сегментации транспорта на видео.

Заключительная фаза — оценка эффективности существующей модели на тестовых данных.

Задача сегментации автомобилей и пешеходов осложняется множеством факторов, например, таких как неоднородность изображений, изменяемость фона, ложные детекции и т.д.

**Объектом исследования** являются архитектуры сегментации объектов на изображении.

**Предметом исследования** являются методы автоматизации и оптимизации препроцессинга видео.

**Целью** Научно исследовательской работы является исследование архитектур нейросетей, методов и существующих решений в области регулировки движения с использованием компьютерного зрения и разработка программы для сегментации автомобилей и пешеходов на видео.

Для достижения поставленной цели необходимо было решить следующие задачи:

1. изучить необходимый материал для построения архитектуры модели,
2. изучить существующие решения и аналоги по сегментации транспортного трафика с использованием камер,
3. найти датасет для обучения модели,
4. разработать программу, для определения каждого пикселя на принадлежность объекту автомобиля или пешехода на видео (выделение нескольких экземпляров классов разными цветами, умение отделять объекты от фона),
5. протестировать и выбрать лучшую модель для сегментации объекта в видеопотоке.

Основная задача – надежное выделение контура объектов.

Вспомогательная задача – оптимизация решения.

Инструменты, использованные при разработке программы: Python, CUDA, Google Colab, YOLO.



## 1. Анализ предметной области

Далее будут рассмотрены:

- сегментирование с использованием нейронных сетей,
- Ultralytics YOLO,
- обзор библиотеки cv2 для работы с видео и изображениями,

### 1.1. Сегментирование с использованием нейронных сетей, обзор архитектур

С развитием нейросетей за последнее десятилетие, методы анализа изображений переживают ренессанс. Нейросетевые подходы позволили повысить уровень точности и эффективности в задачах компьютерного зрения. Новые методы превосходят традиционные в сложных задачах сегментации изображения и видео. Задача сегментации объекта на картинке, представляет собой классификацию по пикселям, определение принадлежности пикселя объекту или фону.

Современные методы сегментации изображений обычно адаптированы с архитектур свёрточных нейронных сетей (CNN – Convolution Neuro Network) или трансформеров [4]. Существует множество архитектур, разработанных для решения данной задачи. К наиболее распространенным архитектурам относятся:

- DeepLab (2015) [5] архитектуры, используют ASPP (Atrous Spatial Pyramid Pooling) свёрточные сети, улавливающие контекст на разных масштабах, последние версии DeepLabV2 и DeepLabv3+ (2021) позволяют более точно восстанавливать пространственную информацию, выводя ее в число ведущих отраслей в этой сфере. Общая схема метода DeepLabV3 и DeepLabv3+ представлена на рисунке 1.

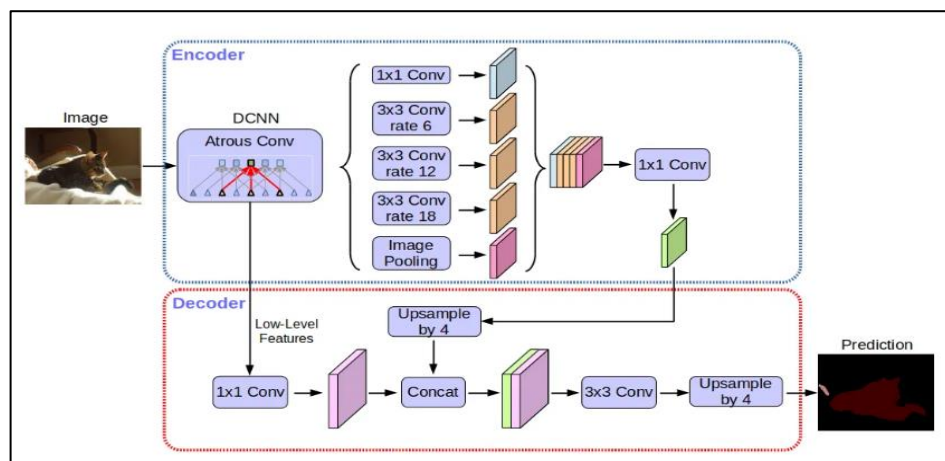
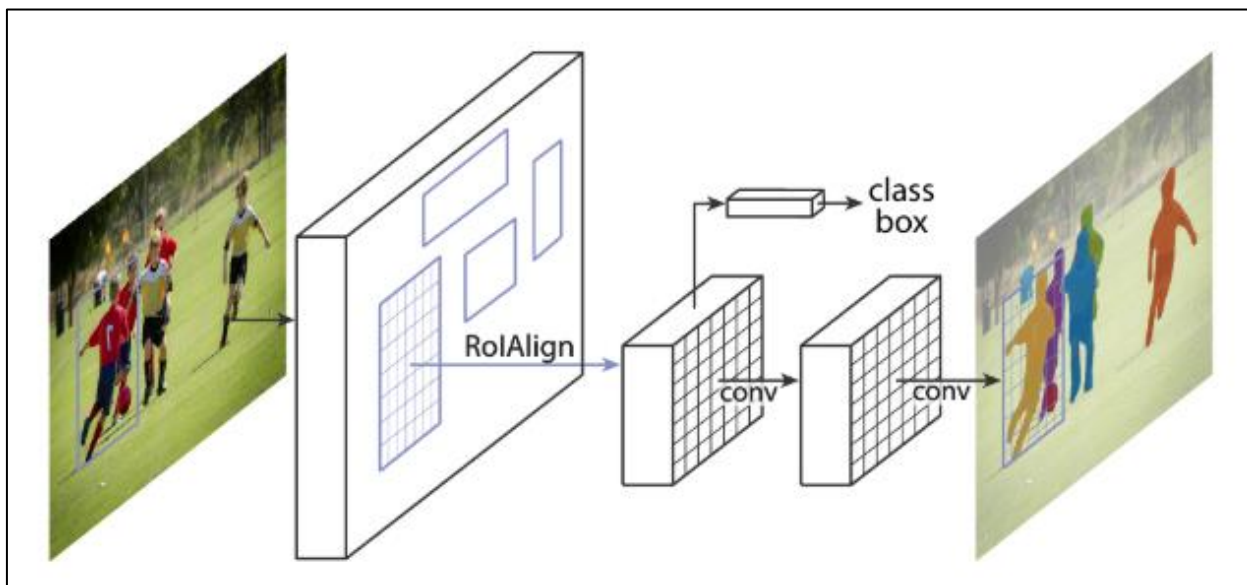


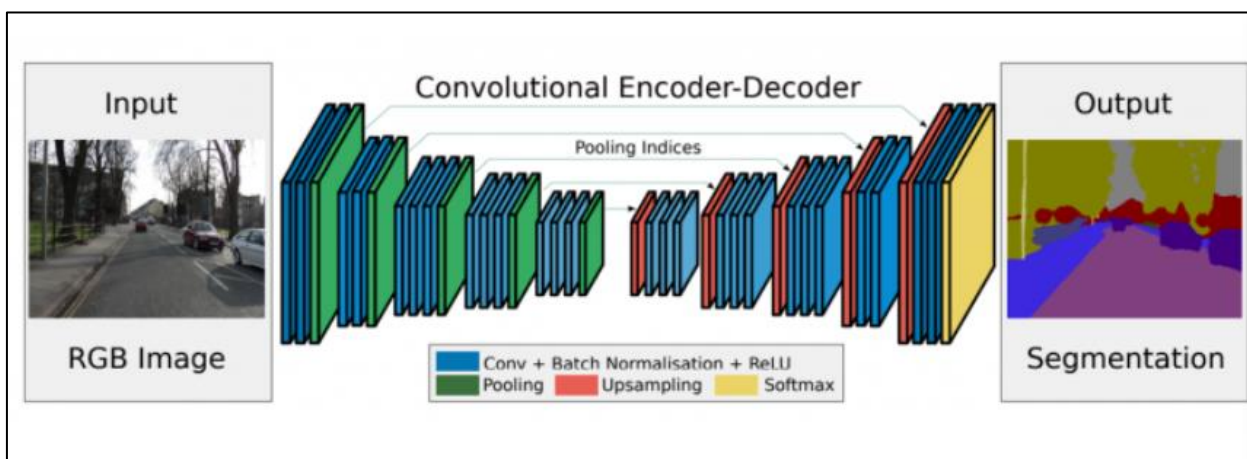
Рисунок 1. Общая схема метода DeepLabV3 и DeepLabv3+.

– Mask R-CNN (2017) [6], представляет собой улучшенную версию R-CNN (Region-based Convolutional Neural Network) [7]. Главное новшество Mask R-CNN заключается в внедрении механизма сегментации на уровне масок, который позволяет точно определить границы объектов, что делает возможным не только выделение классов объектов, но и их формы. Общая схема метода Mask R-CNN представлена на рисунке 2.



**Рисунок 2.** Общая схема метода Mask R-CNN.

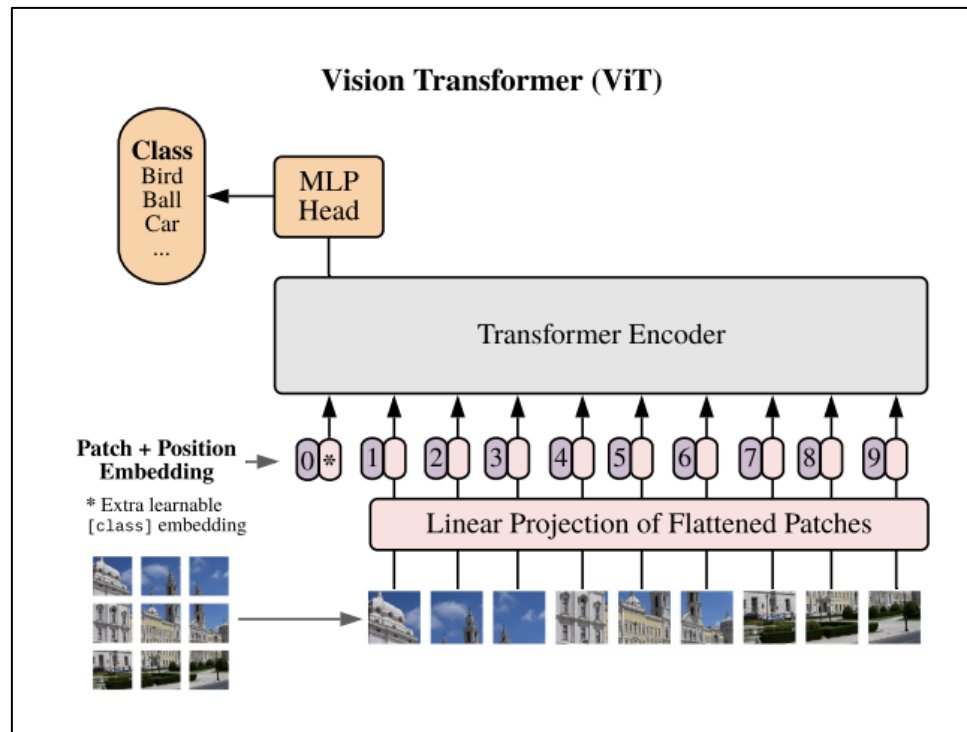
– SegNet (2015) [8]. Была разработана для того, чтобы эффективно решать задачи сегментации с относительно низкими вычислительными затратами, особенно для приложений, требующих высоких разрешений изображений. Не смотря на свою эффективность и оптимизированность, главным недостатком такой модели является ее точность. Схема архитектуры SegNet представлена на рисунке 3.



**Рисунок 3.** Общая схема архитектуры SegNet.



– Transformer-базированные архитектуры ViT (VisionTransformer) (2021) [9] представляют изображения в виде последовательных патчей 16x16 пикселей, которые равномерно преобразуются в векторные представления и подаются на вход трансформеру. В отличие от CNN, ViT не использует свертки, что позволяет захватывать более обширные контексты. К недостаткам метода можно отнести необходимость большого количества данных для обучения. Схема архитектуры Visual Transformer проиллюстрирована на рисунке 4.



**Рисунок 4.** Схема архитектуры Visual Transformer.

Основными недостатками описанных выше методов, являются чрезмерные вычислительные затраты из-за сложных архитектур.

– U-Net является одной из стандартных архитектур CNN для задач сегментации изображений, использующаяся, для сегментации области изображения по классам. Для U-Net характерно получение хороших результатов в различных реальных задачах, при

использовании небольшого количества данных. Архитектура U-Net, изначально разработанная Олафом Роннебергером и др. в 2015 году [10], использовалась для медицинских целей в сегментации биомедицинских изображений. Свое название она получила, из-за U-образной формы схемы модели, представленной на рисунке 5.

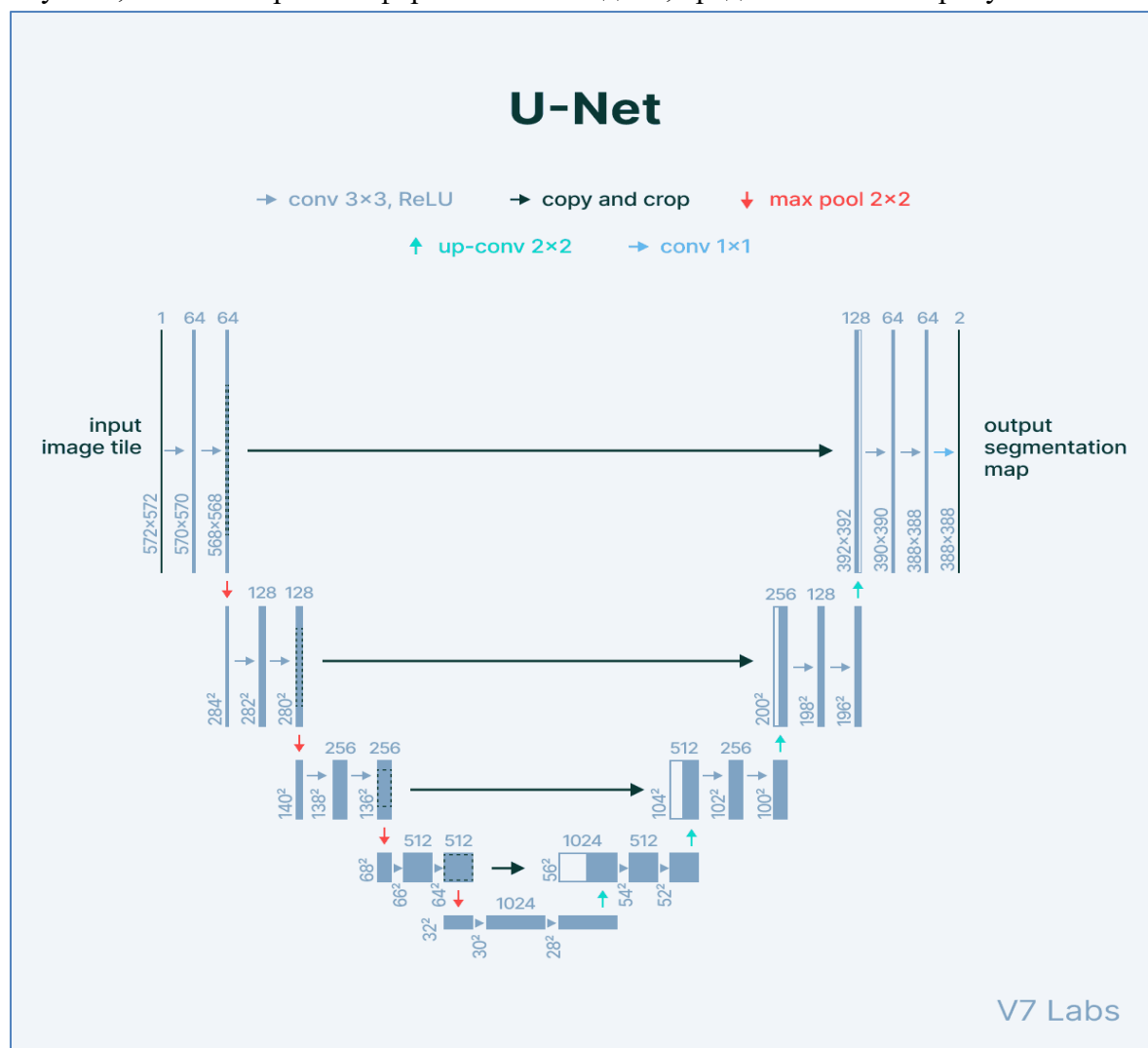
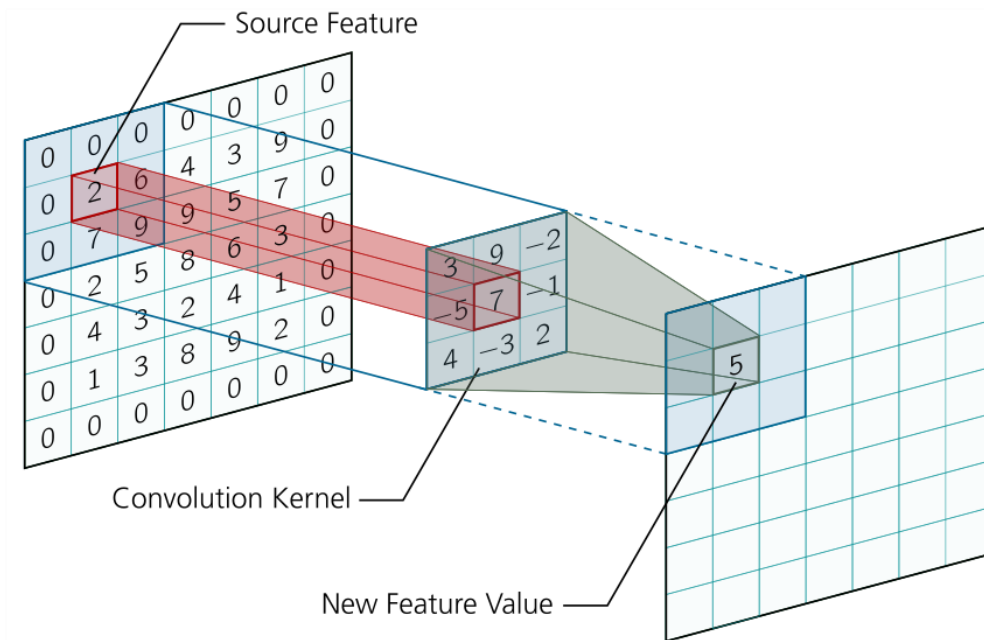


Рисунок 5. Схема архитектуры U-Net.

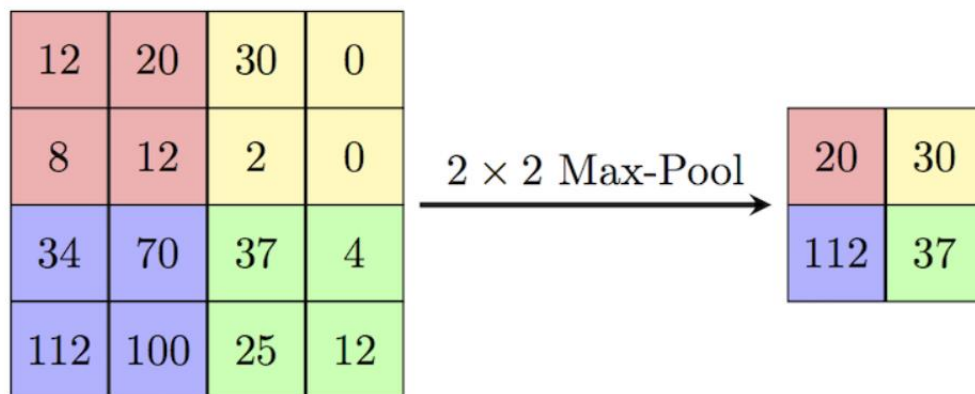
На рисунке 5, в левой части схемы располагается свёрточная сеть – Encoder, в правой стороне Decoder – расширяющийся путь [11]. Ключевой идеей схемы, заключается в том, что каждый промежуточный результат encoder-а конкатенируется с промежуточным результатом правой части архитектуры. Такое решение (skip-connection) было представлено для решения проблемы потери информации во время операции max-pooling.

Encoder работает следующим образом, если на вход подается изображение 572x572x1 и с помощью операции свертки (convolution 3x3) с функцией активации ReLU, и нулевым отступом, для предотвращения потери информации на границах. Слой, создаваемый сверткой, позволяет изучать более локальные признаки, например, обнаружение глаза на изображении. Пример свертки 3x3 представлен на рисунке 6.



**Рисунок 6.** Пример свертки 3x3.

После данных манипуляций получаются тензоры  $572 \times 572 \times 64$ . Следующим шагом применяется maxPooling  $2 \times 2$ , уменьшающий размер тензора по ширине и высоте вдвое. Помимо того, что данный шаг помогает уменьшить количество параметров модели, тем самым предотвратив переобучение, он также повышает надежность модели, предотвращая неточности при небольших изменениях во входных данных. Суть метода maxPooling  $2 \times 2$  представлена на рисунке 7.

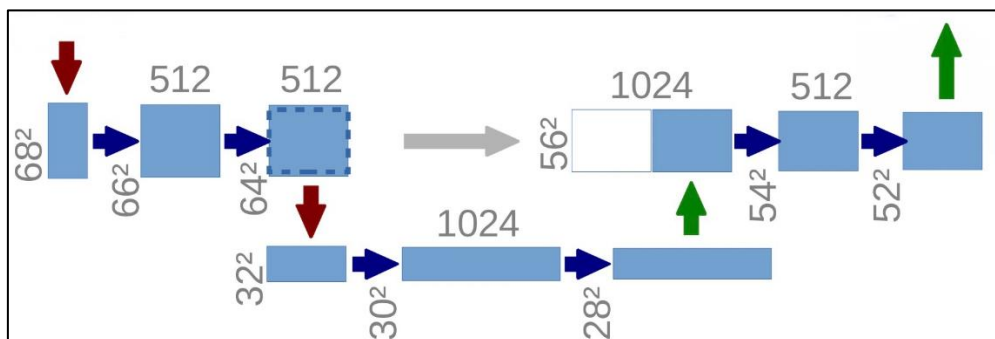


**Рисунок 7.** Пример maxPooling  $2 \times 2$ .

Эти два шага повторяются до момента, когда тензор не станет размером  $28 \times 28 \times 1024$ .

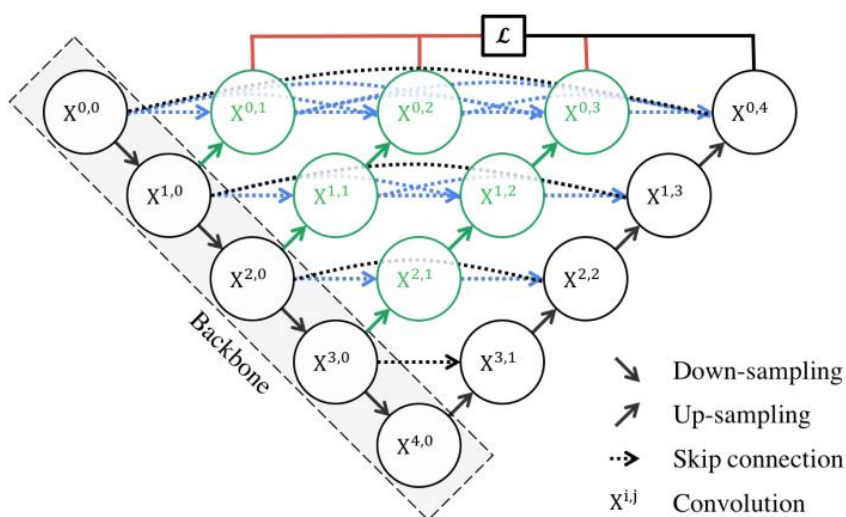
Основной целью декодера, является реконструкция карты сегментации высокого разрешения из изученных признаков, тем самым восстановив пространственную детализацию. Так как при max-pooling терялась важная информация, при выполнении восстановления изображения с помощью up-sampling результат получался не точным, из-за

потери контуров объектов, поэтому для приведения тензора к целевому размеру используется up-convolution 2x2 или conv2dTranspose 2x2, после чего берется тензор 64x64 с предыдущего этапа и конкатенируется с полученным результатом up-conv, чтобы размеры тензоров совпадали, от предыдущий тензор обрезают до 56x56. Данная операция позволяет восстановить признаки, утраченные во время сжатия до тензора 28x28. На рисунке 8 представлена часть схемы с переходом от свёрточной части к восстанавливающей.



**Рисунок 8.** Переход от *encoder* части к *decoder*.

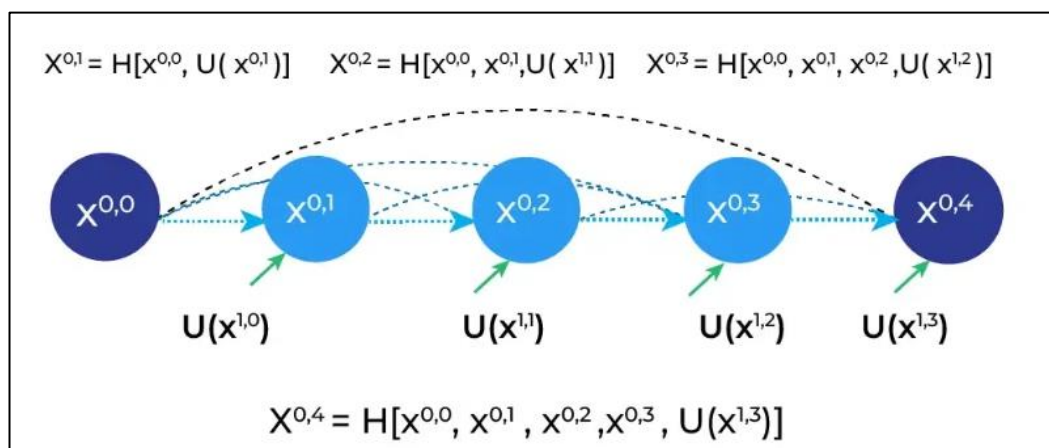
– U-Net++, представленная в 2018 году является модификацией существующей модели U-Net [12]. Для улучшения показателей традиционной U-Net было предложено использование плотных пропускных соединений (skip-connections). На рисунке 9 представлена схема U-Net++.



**Рисунок 9.** Архитектура U-Net++.

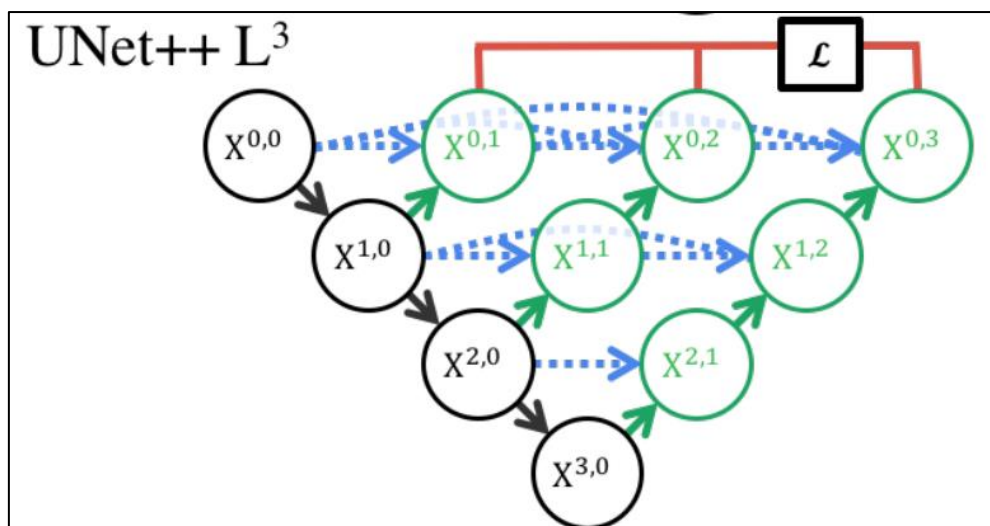
Обновленный дизайн архитектуры предлагает проводить up-convolution с предыдущими параметрами на каждом шаге свертки, полученный skip-connection передается далее, пока не упирается в декодер. Каждый такой мост строится за счет соединения (concatenate) предыдущих узлов up-sampling с свертыванием на пропускном

соединении (рисунок 10). Такой подход позволяет более точно сократить семантический разрыв при восстановлении изображения.



**Рисунок 10.** Оптимизация skip-connection с использованием комбинирования узлов.

Количество шагов вниз и вверх в этих моделях обозначаются буквой  $L$ . Экспериментальным путём было доказано, что архитектура U-Net++  $L3$  при только 3-х слоях обучается на 32,2% быстрее чем U-Net++ $L4$ , при минимальном отклонении в итоговом результате [13]. Схема U-Net++ $L3$  представлена на рисунке 11.



**Рисунок 11.** Архитектура U-Net++ $L3$ .

## 1.2. Ultralytics YOLO

Задача сегментации – одна из ключевых задач в области компьютерного зрения, делится на три группы [14]:

- семантическая сегментация, тип сегментации, классифицирующий каждый пиксель изображения, без возможности определить разные экземпляры одного и того-же объекта (рисунок 12);



**Рисунок 12.** Пример семантической сегментации.

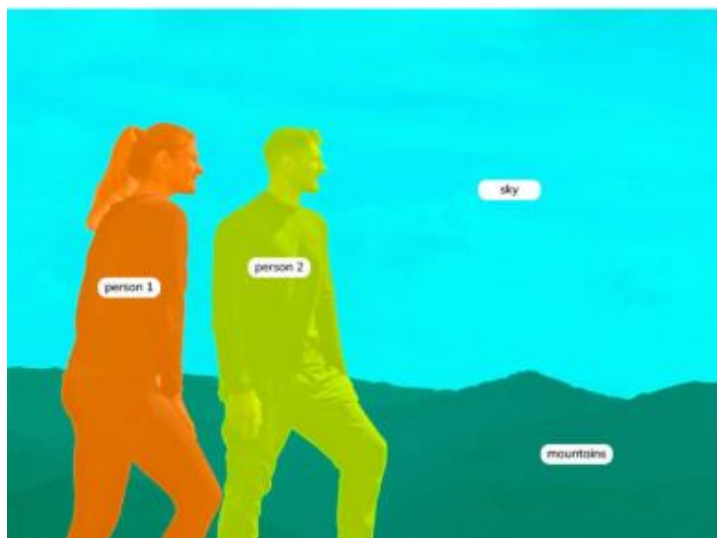
– instance segmentation, тип сегментации, способный различать не только класс объекта, но и различные экземпляры (рисунок 13);



**Рисунок 13.** Пример instance сегментации.

– паноптическая сегментация, распознает не только отдельные экземпляры объектов, но и фоновые (небо/земля), что позволяет более комплексно анализировать сцены (рисунок 14).



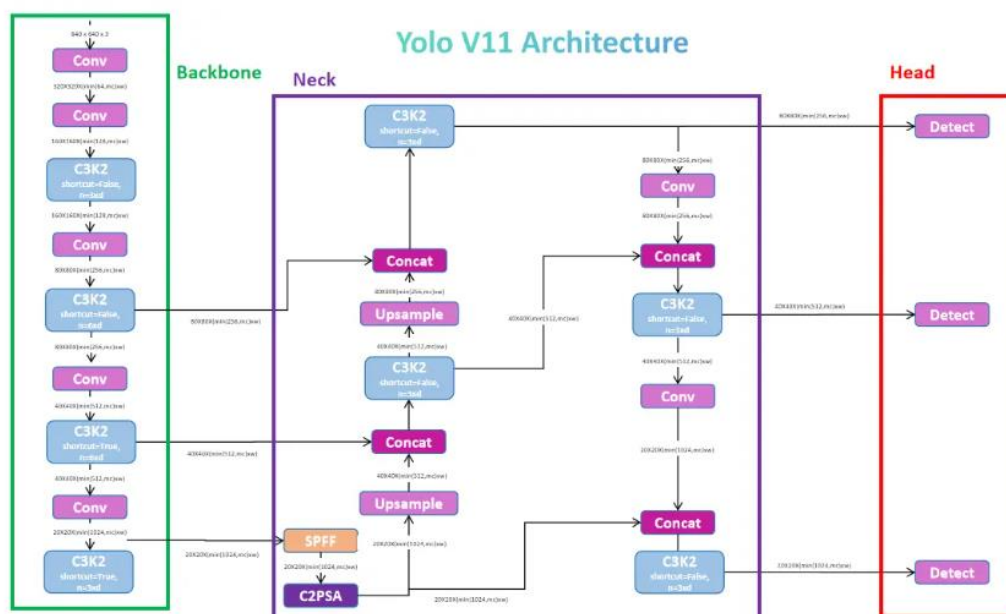


**Рисунок 14.** Пример паноптической сегментации.

В текущей работе, для определения объектов и их наличия будет использоваться метод Instance Segmentation.

Платформа ultralytics YOLO (You Only Look Once) специализируется на быстром и точном обнаружении объектов, а также поддерживает задачи сегментации. Помимо быстрого и точного определения объектов, система указывает количество экземпляров, обнаруженных на картинке, что необходимо для решения текущей задачи.

В рамках НИР использовалась YOLO11 с трехсекционной архитектурой Backbone, Neck и Head [15], представленной на рисунке 15.



**Рисунок 15.** Архитектура YOLO v11.

Backbone часть извлекает из изображения полезные признаки на основе эффективного “bottleneck-based” алгоритма.

Neck – обрабатывает выходные данные Backbone и передает полученные признаки головной части

Head (головная) часть, работает исходя из поставленной задачи:

- обнаружение объектов на изображении,
- семантическая сегментация объектов,
- определение ключевых точек для позы (например, на человеке голова, локти и т.д.),
- классификация изображений и т.д.

Bottleneck-based слой представляет из себя два сверточных блока, соединенных последовательно с функцией конкатенации, если параметр `shortcut = true`, входные параметры объединяются с выходными параметрами второго сверточного слоя. Такая структура повышает эффективность и процесс обучения. На рисунке 16 представлен Bottleneck слой.

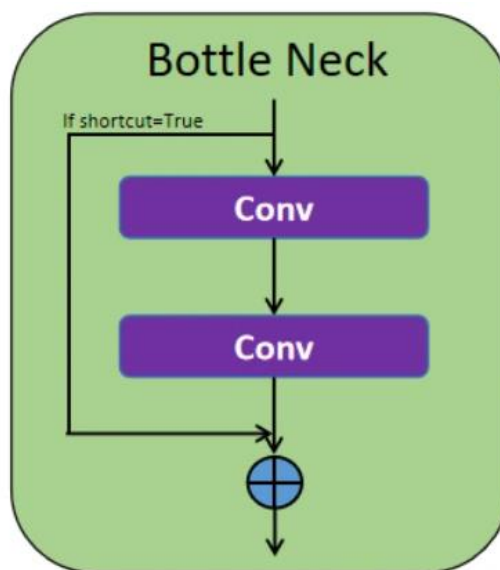


Рисунок 16. *Bottleneck* слой.

### 1.3. Обзор библиотеки cv2

Cv2 это модуль OpenCV в Python - открытой библиотеки для работы с алгоритмами компьютерного зрения [16]. В рамках работы используется для преобразований изображений и видео.

В этом пункте будут рассмотрены:

- методы для работы с видео.



### 1.3.1. Методы для работы с видео

В подпункте будут рассмотрены методы для работы с видеорядом, использовавшиеся во время разработки:

- `cv2.VideoCapture(fileName)` – функция для захвата видео записывается в значение `cap`, на вход подается путь к файлу, далее в цикле `while True`: с помощью `cap.read()` считывается кадр видео, данная функция возвращает 2 значения (логическое значение корректности считывания кадра и сам кадр);

- `cap.get(cv2.CAP_PROP_FPS)` – используется для получения кадров в секунду у обрабатываемого видео, `fps` (frames per second) понадобятся для записи полученных изображений в видео;

- `cv2.waitKey(1)` - функция, прерывающая выполнение бесконечного цикла, ждет нажатия клавиши 1 миллисекунду, чтобы проверить нужно ли освободить захват видео с помощью `cv2.destroyAllWindows()` и выходить из цикла;

- `cv2.VideoWriter(pathToVideo, cv2.VideoWriter_fourcc(*'DIVX'), fps, (width, height))` - функция для записи видео в файл, на вход подается путь до файла для сохранения, переменная `fourcc` - 4-байтовый код, который используется для указания видеокodeка, количество кадров в секунду и размер видео в пикселях;

- `video.write(cv2.imread(os.path.join(pathToVideoImages, image)))` – используется для конвертации папки изображений в видеоряд, в рамках работы программы, во время сегментации, обработанные изображения сохраняются в папку, чтобы в случае вылета программы полученный результат можно было склеить в один видеоряд.

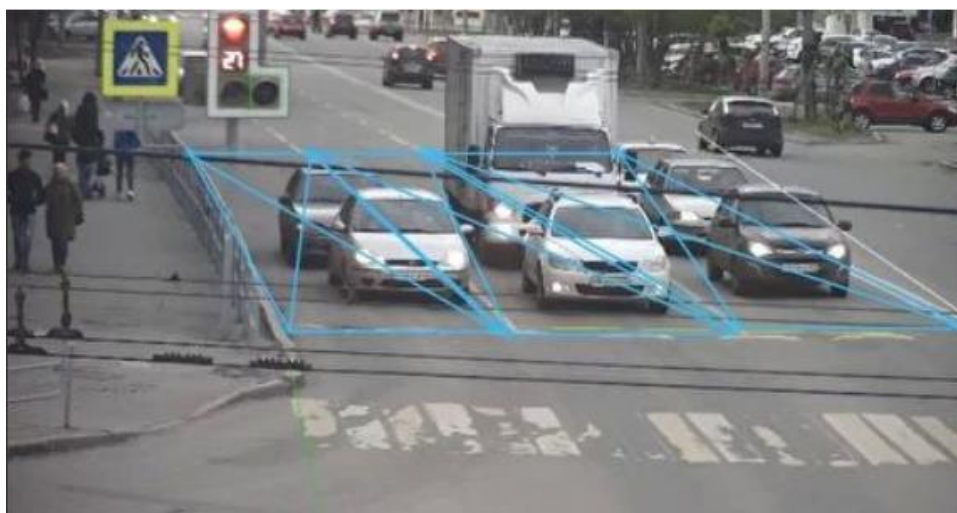
## 2. Существующие решения проблемы трафика

Далее будут рассмотрены:

- адаптивный светофор,
- использование ML для детекции автомобилей в России,
- использование светофоров, управляемых ии, в странах Европы.

### 2.1. Адаптивный светофор

Адаптивные светофоры используют устаревшую технологию подсчета автомобилей на перекрестках, а именно физические датчики или видеодетектор фона в определенных зонах. Ограниченность этой технологии заключается в том, что обнаружение транспорта с помощью датчиков ограничено несколькими метрами перед переходом, а значит видеодетектор фона показывает заполненность только этого участка. К тому-же, если камера не сможет различить эту область на видео из-за расстояния, смещения или погодных условий, то светофор перестанет работать оптимально. Пример адаптивного светофора представлен на рисунке 17.



**Рисунок 17.** Видеодетектор адаптивного светофора.

Помимо устаревших технологий, установка физических датчиков для зоны детекции требует реконструкции дорожного полотна и местами стоимость установки может исчисляться миллионами [17 - 18].

Таким образом, адаптивные светофоры работают только в ситуациях, где правила переключения абсолютно очевидны или на мало загруженных перекрестках.

### 2.2. Использование ML для детекции автомобилей в России

Преимущества нейросетевых светофоров над адаптивными:

- не нуждаются в установке физических датчиков,

- получают более информативную информацию, например, тип транспорта, его длина,
- помимо транспортных средств, могут учитывать скопления пешеходов,
- количество объектов для детекции не ограничивается только типом транспорт и пешеход.

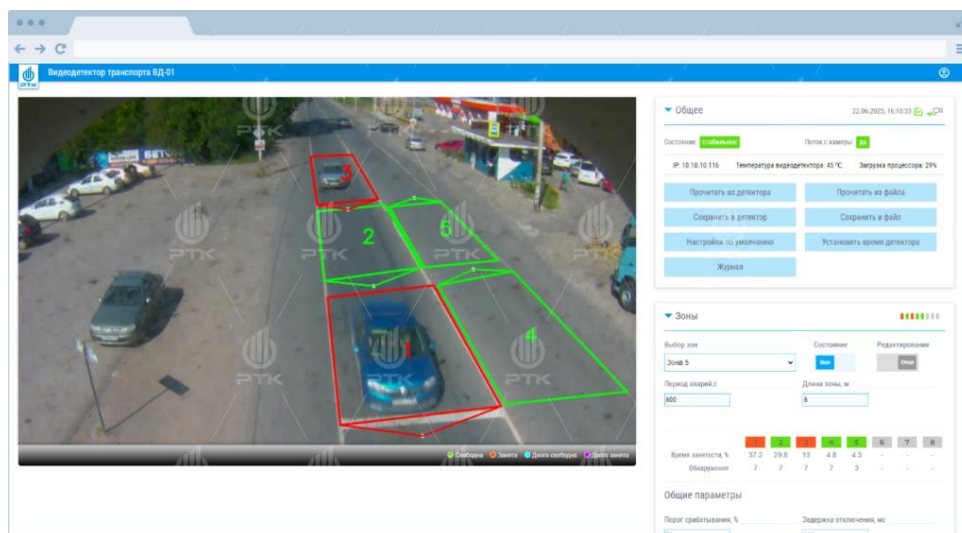
Рассмотренный в предыдущем пункте адаптивный светофор, хоть и умеет переключать сигналы светофора за счет датчиков, но делает это без использования компьютерного зрения. Детекция автомобилей в России используется в основном для фиксации нарушений на дорогах.

Далее будут рассмотрены проекты из России использующие компьютерное зрение для контроля движения на дорогах:

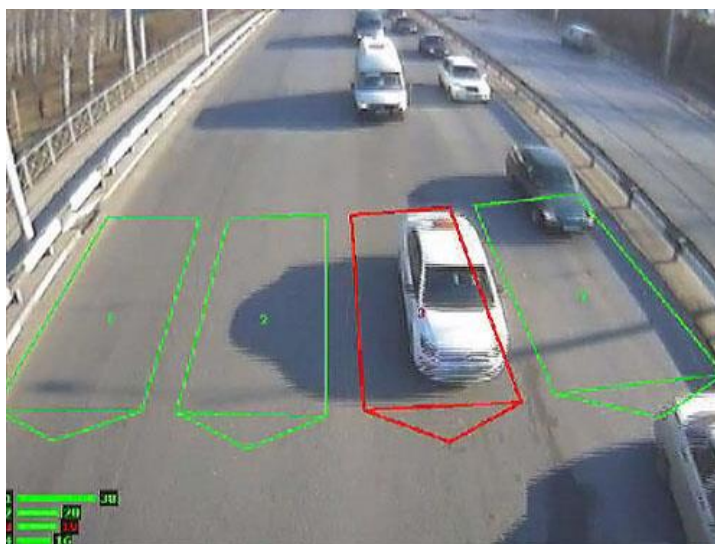
- видеодетектор автотранспортных средств ИНФОПРО и видеодетектор транспорта РТК,
- НПО ИТС: Паук Трафик,

### **2.2.1. Видеодетектор автотранспортных средств ИНФОПРО и Видеодетектор транспорта РТК**

Оба проекта работают по принципу адаптивных светофоров, но используют нейросетевые алгоритмы для вычисления геометрического центра транспортного средства. Особенности этих проектов являются получение и анализ статистики по следующим параметрам: объем потока, средняя скорость, занятость дороги, интервал следования автотранспортных средств [19]. Система ИНФОПРО, может переключаться в режим адаптивного светофора [20], когда Видеодетектор РТК может похвастаться определением класса транспортного средства. На рисунке 18 представлен интерфейс Видеодетектора РТК, а на рисунке 19 видеодетектор ИНФОПРО.



**Рисунок 18.** *Видеодетектор РТК.*



**Рисунок 19.** *Видеодетектор ИНФОПРО.*

### **2.2.2. НПО ИТС: Паук Трафик**

Проект НПО ИТС: Паук Трафик тоже используется для аналитики дорожного движения, но в отличие от рассмотренных ранее проектов активно использует технологии компьютерного зрения. Видеодетектор способен определять количество автомобилей, плотность трафика, среднюю скорость на дороге и индивидуальную скорость транспортного средства, а также его тип. Разработанный инженерами «НПО «ИТС» детектор устойчив к погодным условиям и может контролировать движение даже на многополосных магистралях с погрешностью детектирования не выше 5%. Помимо анализа и сбора данных, комплекс видеофиксации «ПАУК» выявляет факты нарушения скоростного режима, проезда на запрещающий сигнал светофора, правил пользования ремнями безопасности и телефоном во время движения, пересечение стоп-линии и другие

нарушения ПДД, а также умеет распознавать госномера нарушителей [21]. На рисунке 20 представлен видеодетектор Паук Трафик.

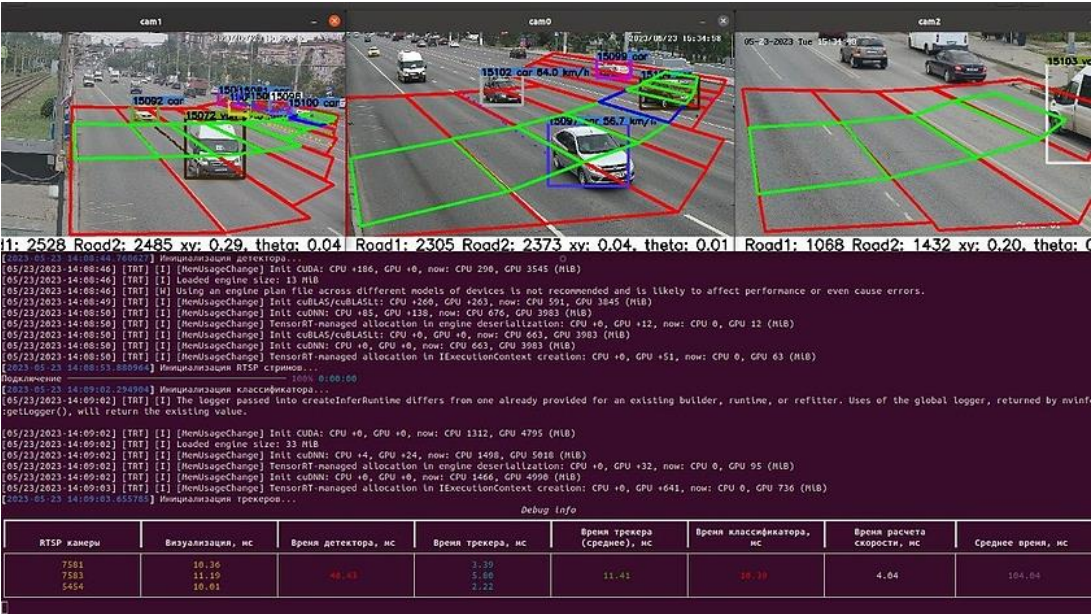


Рисунок 20. Видеодетектор Паук Трафик.

### 2.3. Использование светофоров, управляемых ИИ, в странах Европы

В отличие от, Российских проектов с использованием Компьютерного зрения, для анализа дорожного движения, в Германии уже проводятся испытания светофоров с искусственным интеллектом.

Так в 2023 году в немецком городе Лемго проект KI4LSA установил светофоры с камерами высокого разрешения и радарными датчиками, для фиксирования реальной дорожной ситуации, включая среднюю скорость автомобилей и время ожидания, для интеллектуального прогнозирующего переключения светофоров [22 - 23].

В конце наблюдений, использование искусственного интеллекта улучшило транспортный поток на 10-15% [24]. На рисунке 21 представлен видеодетектор KI4LSA.





**Рисунок 21.** Видеодетектор KI4LSA.

### **3. Обучение и тестирование модели**

Будут рассмотрены:

- результаты поисков обучающего датасета,
- результаты сегментации видео на предобученной модели,
- направления для развития.

В качестве результатов будут представлены результаты метрик и их сравнение. По результатам сегментации будут выявлены недочеты использования предобученной модели.

#### **3.1. Результаты поисков обучающего датасета**

Для обучения модели необходимо иметь датасет с размеченными данными. После тщательного поиска на сайтах / платформах, связанных с машинным обучением, подходящих размеченных данных с камер наблюдения обнаружено не было.

Вторым вариантом было искать прямые трансляции улиц городов в России, для ручной разметки видеоматериала. Есть множество сайтов с прямыми трансляциями частных камер установленных по городу. Некоторые камеры установлены на перекрестках и рядом со светофорами. Такие камеры можно использовать, как и для обучения, так и для тестирования распознавания. Также можно запросить доступ к видеозаписи с муниципальных камер, интегрированных в единый центр хранения и обработки данных, обратившись в необходимые правоохранительные органы [25].

В предыдущем году обучения для разметки изображений при создании датасета использовался сервис Roboflow, но помимо него существует множество широко используемых инструментов, в том числе сервисов позволяющих автоматизировать процесс, из самых популярных можно выделить:

- LabelMe – бесплатный инструмент с открытым исходным кодом для разметки изображений с возможностью добавления лейблов и выделения объектов на изображении;
- VGG Image Annotator (VIA) - бесплатное программное решение, для создания аннотаций на уровне пикселей, с поддержкой множества форматов;
- SuperAnnotate – платформа с возможностью автоматической разметки с ручной корректировкой;
- CVAT – позволяет автоматизировать процесс разметки на основе предобученных моделей.

### 3.2. Результаты сегментации видео на предобученной модели

В рамках текущей НИР был написан класс *VideoConverterClass* на языке python для сегментации объектов в видеопотоке с использованием предобученной модели YOLO11l-seg. Модели Yolo для сегментации делятся на n, s, m, l, x, различия между моделями заключаются в скорости и точности детекции. Если необходима скорость детекции, но не ее точность выбирается модель n, иначе, если необходима точность, но не скорость выбирается модель x [26].

На рисунке 22 представлен график точнос, по результатам обучения модели на 100 эпохах и с использованием данных из датасета COCO, взятого из открытого доступа.



**Рисунок 22.** График точности обучения модели.

Класс *VideoConverterClass* состоит из двух методов:

- maskVideoImg – с использованием библиотеки cv2, для захвата видеопотока, алгоритм обрабатывает каждый видеокادر в классе predictSingleImg и записывает в отдельную папку, метод представлен на рисунке 23;

```

def predictSingleImg(self, img, model, pathToVideoImages, curIm):
    '''
    Predict single img with the model
    img - Изображение для сегментации
    model - обученная модель для сегментации
    '''
    results = model(img) # predict on an image

    # Access the results
    for result in results:
        print(type(result.masks))
        if result.masks:
            xy = result.masks.xy # mask in polygon format
            xyn = result.masks.xyn # normalized
            masks = result.masks.data # mask in matrix format (num_objects
            x H x W)
            name = pathToVideoImages + "/" + str(curIm).zfill(32) + '.jpg'
            result.save(filename=name)

    return result

```

**Рисунок 23.** Метод для сегментации единственного изображения с помощью модели Yolo11l-seg.

– convertToVid - метод, конвертирует все обработанные кадры из папки в одно видео с использованием cv2

Далее будут представлены результаты сегментации на 3-х видеороликах, записанных на онлайн-трансляциях видеокамер с сайта “Сети Тагила”.

На рисунке 24 можно заметить, что предобученная модель смогла определить большинство объектов правильно.



**Рисунок 24.** Корректное определение объектов на видеоряде 1.



Но уже через пару секунд, сеть стала путать пешеходов с водяными гидрантами (рисунок 25) и перестала обнаруживать большинство автомобилей (рисунок 26).

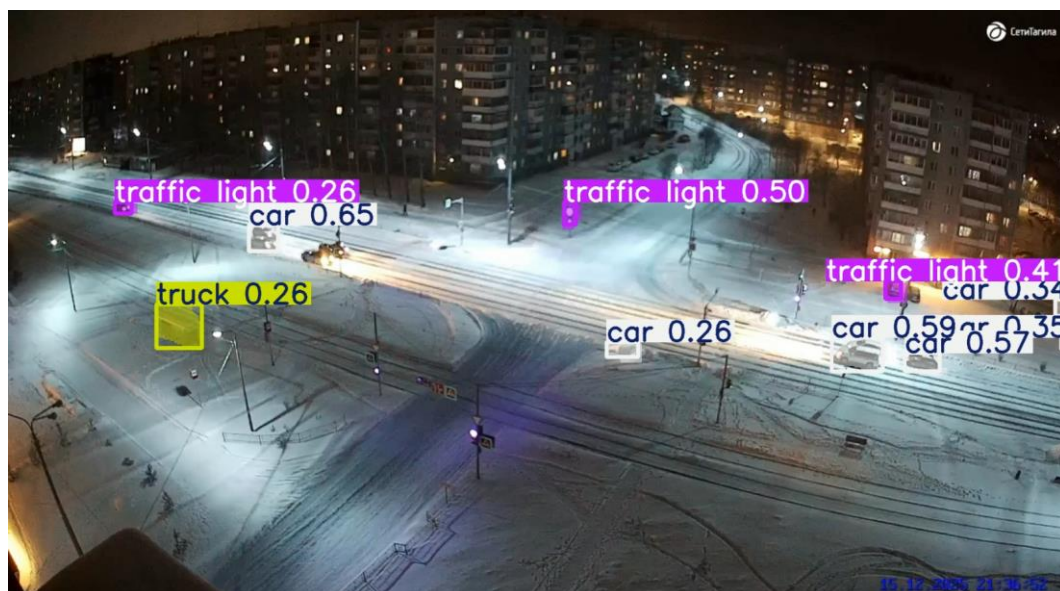


**Рисунок 25.** Некорректное определение объектов на видеоряде 3.



**Рисунок 26.** Некорректное определение объектов на видеоряде 1.

В случае с камерой, расположенной более удаленно, некоторые машины модель воспринимала за светофоры, как на рисунке 27.



**Рисунок 27.** Некорректное определение объектов на видеоряде 2.

По результатам сегментации видеоряда, можно сделать вывод, что сегментация модели на данных из открытого доступа работает, но точность определения объектов очень нестабильна.

### **3.3. Направления для развития**

Для улучшения сегментации необходимо дообучить модель на корректных данных, заранее аннотированных в сторонних сервисах.

Также можно попробовать написать собственную архитектуру, для возможности более гибкой модификации модели.

## ЗАКЛЮЧЕНИЕ

В результате выполнения научно исследовательской работы был изучен материал, необходимый для построения архитектуры модели, а также рассмотрены проекты решающие похожие задачи. Также был разработан класс, использующий модель YOLO11-seg, для сегментирования видео с целью определения каждого пикселя на принадлежность к транспортным средствам или людям. В заключении разработанный класс был протестирован на подготовленных заранее видеороликах и по полученным результатам, определены дальнейшие направления для развития.

Были реализованы задачи, а именно:

- подготовлена система к выполнению задания,
- написана программа для обработки данных,
- проведены циклы обучения на данных из открытого доступа и тестирования модели,
- получены формальные и субъективные оценки модели.

Таким образом, цель и поставленные задачи были достигнуты.

Для дальнейшего развития необходимо:

- разметить данные для создания подходящего датасета,
- реализовать модель по типу YOLO, для множественного обнаружения объектов на изображении,
- разработать алгоритм для управления фазами.

## ГЛОССАРИЙ

Workflow - это процесс, который включает последовательность действий.

Препроцессинг - это процесс предварительной обработки данных.

CNN – (Convolutional Neural Network) это тип алгоритма глубокого обучения, предназначенный для обработки визуальных данных.

Тензор - это представление данных, которое используется для решения задач сегментации изображений или временных рядов. Например, в задаче семантической сегментации изображений, элементы тензора отвечают о принадлежности каждого пикселя к определённому классу.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 А.Б. Рахманов. Четыре колеса апокалипсиса: причины автомобильных пробок в крупных странах мира [Научная статья] // cyberlinka.ru [Сайт]. URL: <https://cyberleninka.ru/article/n/chetyre-kolesa-apokalipsisa-prichiny-avtomobilnyh-probok-v-krupnyh-gorodah-mira/viewer> (Дата последнего обращения 22.12.2025)
- 2 В США подсчитали, во сколько обходятся стране автомобильные пробки [Электронный ресурс] // officelife.media [Сайт]. URL: <https://officelife.media/news/54320-v-ssha-podschitali-vo-skolko-obkhodyatsya-strane-avtomobilnye-probki-summa-vpechatlyayet/> (Дата последнего обращения 22.12.2025)
- 3 Ученые Пермского Политеха разработали алгоритм оптимизации перекрестков на основе компьютерного моделирования [Электронный ресурс] // pstu.ru [Сайт]. URL: <https://pstu.ru/news/2025/10/17/17823/> (Дата последнего обращения 22.12.2025)
- 4 В. А. Офицеров, А. С. Конушин. Нейросетевые методы сегментации изображений высокого разрешения [Научная статья] // cyberleninka.ru [Сайт]. URL: <https://cyberleninka.ru/article/n/neyrosetevye-metody-segmentatsii-izobrazheniy-vysokogo-razresheniya/viewer> (дата последнего обращения 22.12.2025).
- 5 Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam. Rethinking Atrous Convolution for Semantic Image Segmentation [Научная статья] // arxiv.org [Сайт]. URL: <https://arxiv.org/pdf/1706.05587> (дата последнего обращения 22.12.2025).
- 6 Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. Mask R-CNN, 2018 [Научная статья] // arxiv.org [Сайт]. URL: <https://arxiv.org/pdf/1703.06870> (дата последнего обращения 22.12.2025).
- 7 Abirami Vina. What is R-CNN? A quick overview [Электронный ресурс] // www.ultralitics.com [Сайт]. URL: <https://www.ultralitics.com/ru/blog/what-is-r-cnn-a-quick-overview> (дата последнего обращения 22.12.2025).
- 8 Vijay Badrinarayanan, Alex Kendall, SegNet: A Deep Convolutional EncoderDecoder Architecture for Image Segmentation, 2016 [Научная статья] // arxiv.org [Сайт]. URL: <https://arxiv.org/pdf/1511.00561> (дата последнего обращения 22.12.2025).
- 9 Визуальные трансформеры (ViT) [Электронный ресурс] // habr.com [Сайт]. URL: <https://habr.com/ru/companies/otus/articles/849756/> (дата последнего обращения 22.12.2025).
- 10 Olaf Ronneberger, Philipp Fischer, U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015. [Научная статья] // arxiv.org [Сайт]. URL: <https://arxiv.org/pdf/1505.04597> (дата последнего обращения 22.12.2025).

- 11 U-Net: нейросеть для сегментации изображений [Электронный ресурс] // neurohive.io [Сайт]. URL: <https://neurohive.io/ru/vidy-nejrosetej/u-net-image-segmentation/> (дата последнего обращения 22.12.2025).
- 12 Unet++ Architecture Explained [Электронный ресурс] // www.geeksforgeeks.org [Сайт]. URL: <https://www.geeksforgeeks.org/machine-learning/unet-architecture-explained/> (дата последнего обращения 22.12.2025).
- 13 Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. [Научная статья] // arxiv.org [Сайт]. URL: <https://arxiv.org/pdf/1807.10165v1> (дата последнего обращения 22.12.2025).
- 14 Семантическая сегментация: самый полный гайд 2024. [Электронный ресурс] // habr.com [Сайт]. URL: [https://habr.com/ru/companies/data\\_light/articles/855336/](https://habr.com/ru/companies/data_light/articles/855336/) (дата последнего обращения 22.12.2025).
- 15 Nikhileswara Rao Sulake Guide on YOLOv11 Model Building from Scratch using PyTorch. [документация] // www.analyticsvidhya.com [Сайт]. URL: <https://www.analyticsvidhya.com/blog/2025/01/yolov11-model-building/> (дата последнего обращения 22.12.2025).
- 16 Open Source Computer Vision [Электронный ресурс] // docs.opencv.org [Сайт]. URL: <https://docs.opencv.org/3.4/index.html> (дата последнего обращения 22.12.2025) или <https://opencv.org/> (дата последнего обращения 22.12.2025).
- 17 Как сделать светофор умнее, а проезжать и загруженные перекрестки быстрее? [Электронный ресурс] // habr.com [Сайт]. URL: <https://habr.com/ru/articles/553162/> (дата последнего обращения 22.12.2025).
- 18 На «умные» светофоры в Уфе в 2023 году потратили более 270 млн рублей [Электронный ресурс] // www.kommersant.ru [Сайт]. URL: <https://www.kommersant.ru/doc/6509735> (дата последнего обращения 22.12.2025).
- 19 Видеодетектор автотранспортных средств ИНФОПРО [Электронный ресурс] // infoprocess.ru [Сайт]. URL: <https://www.infoprocess.ru/products.html> (дата последнего обращения 22.12.2025).
- 20 Видеодетектор автотранспортных средств ИНФОПРО сценарии использования [документация] <https://www.infoprocess.ru/downloads/scenario.pdf> (дата последнего обращения 22.12.2025).
- 21 НПО ИТС: Паук Трафик [электронный ресурс] // www.tadviser.ru [Сайт] URL: <https://clck.ru/3QxVQh> (дата последнего обращения 22.12.2025).

- 22 В Германии тестируют умные светофоры. Ими управляют самообучающиеся алгоритмы [электронный ресурс] // digitalocean.ru [Сайт] URL: <https://digitalocean.ru/n/v-germanii-testiruyut-umnye-svetofory> (дата последнего обращения 22.12.2025).
- 23 Светофоры с искусственным интеллектом начали испытывать в Германии [электронный ресурс] // kolesa.kz [Сайт] URL: <https://kolesa.kz/content/news/svetofory-s-iskusstvennym-intellektom-nachali-ispytyvat-v-germanii/> (дата последнего обращения 22.12.2025).
- 24 KI4LSA Künstliche Intelligenz für Lichtsignalanlagen. (KI4LSA Искусственный интеллект для Системы световой сигнализации) [научная статья] URL: [https://www.iosb-ina.fraunhofer.de/content/dam/iosb/iosb-ina/documents/WhitePaper\\_KI4LSA.pdf](https://www.iosb-ina.fraunhofer.de/content/dam/iosb/iosb-ina/documents/WhitePaper_KI4LSA.pdf) (дата последнего обращения 22.12.2025).
- 25 Как получить запись с городских камер видеонаблюдения [Электронный ресурс] // samara-ohrana.ru [Сайт] URL: <https://clck.ru/3QxVyt> (дата последнего обращения 22.12.2025).
- 26 Instance Segmentation [документация] // docs.ultralytics.com [Сайт] URL: <https://docs.ultralytics.com/tasks/segment/> (дата последнего обращения 22.12.2025).

## **Приложение А**

**ссылка на репозиторий с кодом программы и результатами сегментации**

Репозиторий проекта можно найти, перейдя по ссылке:

<https://github.com/Uranus28/CVTrafficLight> (github.com)