



Московский государственный университет имени М.В.Ломоносова

Механико-математический факультет

Кафедра Математической теории интеллектуальных систем

**Киназаров Темирбек**

**«Разработка нейросетевых алгоритмов определения  
пространственных характеристик движущихся объектов»**

**Выпускная квалификационная работа**

Научный руководитель:

доцент кафедры Математической Теории Интеллектуальных Систем,

кандидат физико-математических наук

Часовских Анатолий Александрович.

Москва, 2020

# Оглавление

§ 1	Введение . . . . .	3
§ 2	Постановка задачи . . . . .	4
§ 3	Описание нейросетевой модели . . . . .	5
3.1	Нахождение трехмерной описывающей рамки для автомобилей в кадре . . . . .	5
3.2	Определение марки автомобиля и визуализация . . . . .	8
§ 4	Оценка расстояния до автомобиля . . . . .	9
4.1	Основные обозначения . . . . .	9
4.2	Модель оценки расстояния до автомобиля . . . . .	10
§ 5	Модель оценки центра нижней грани трехмерной рамки и оценка точности . . . . .	12
5.1	Оценка точности . . . . .	13
5.2	Оценка погрешности дискретизации . . . . .	15
§ 6	Заключение . . . . .	16
<b>Литература</b>		<b>17</b>
§ Ж	Пример вывода нейросетевой модели, строящую описанную трехмер- ную рамку . . . . .	19
§ З	Пример вывода модели нахождения марки автомобиля и расстояния до объекта . . . . .	20

## § 1 Введение

Обнаружение трехмерных объектов является одной из важнейших задач для систем восприятия автономных транспортных средств. На текущий момент существует несколько различных систем, позволяющих определять расположение трехмерных объектов на плоских изображениях, представленных набором пикселей [1] "TrackNet" [2] "3D Bounding Box Estimation" [3]. Во многих из них результат работы представлен в виде двумерной рамки, определяющей положение объекта на кадре. Подобное представление решения может влечь за собой потерю точности при использовании упомянутых систем в задаче отслеживания перемещения объекта, при заданном наборе кадров видеоряда. Для увеличения точности отслеживания траектории движения существуют модели построения проекции трехмерной рамки вокруг объекта [4] "Deep MANTA" [5] "Monocular 3D Object Detection". В рамках настоящей работы рассматривается задача определения местоположений проекций центров тяжести автомобилей на дорожное покрытие. Решение указанной задачи позволяет упростить задачу отслеживания перемещения транспортного средства по данным, полученным из видеоряда с некоторой камеры.

В представленной работе предложено решение подзадачи для указанной задачи, состоящей в определении трёхмерной рамки [6], охватывающей транспортные средства на изображении, с последующим определением марки автомобиля. При верном определении трехмерной описывающей рамки для автомобиля на заданном кадре, совместно с маркой автомобиля, можно оценить расположение проекции центра автомобиля с некоторой погрешностью при заданных ограничениях.

Кроме того, в настоящей работе описана математическая модель, позволяющая оценить расстояние до автомобиля, запечатленного на изображении, используя информацию о марке автомобиля и трёхмерной рамке, полученных при помощи нейросетевой модели.

Примеры вывода модели продемонстрированы в приложении.

Данная задача является актуальной для обнаружения нарушения ПДД автотранспортом путем визуальной идентификации электронно-числительной техникой.



Рис. 1: Пример построения траектории автотранспорта

## § 2 Постановка задачи

При съемке камерой происходит проективное преобразование трехмерной сцены на двумерное изображение. Проекция трехмерной точки на изображение получается по формуле преобразования координаты исходной точки сцены в координаты пикселя на изображении:

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \left( K[R \ T] \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \right), \quad (2.1)$$

где  $z_c$  некоторый масштабный коэффициент. Матрица внутренней калибровки  $K$  содержит 5 значимых параметров:

$$K = \begin{pmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

где  $\alpha_x$  и  $\alpha_y$  соответствуют фокусному расстоянию, измеренному в ширине и высоте пикселя,  $u_0$  и  $v_0$  — координаты  $x$  и  $y$  оптического центра в плоскости изображения, а  $\gamma$  — тангенс угла наклона оси  $y$  к оси  $x$ , обычно равен 1.  $R, T$  (где  $R$  — матрица  $3 \times 3$  поворота,  $T$  — вектор  $3 \times 1$  преобразования) — параметры внешней калибровки, определяющие преобразование координат, переводящее координаты точек сцены из трехмерной системы координат в систему координат, связанную с камерой.

В настоящей работе предполагается, что задана матрица  $K[R|T]$  параметров внутренней калибровки камеры, высота, на которой находится камера и угол наклона объектива камеры к перпендикуляру поверхности земли. Подобные ограничения наложены на камеры, при помощи которых ведется съемка автотранспорта на дорогах, для автоматического обнаружения нарушений ПДД. Необходимо построить точку, аппроксимирующую центр нижней грани автотранспорта путем нахождения координат трехмерной рамки, которая охватывает объект. В задачу также входит нахождение направления движения объекта по одному изображению и вывод предполагаемой марки и модели автотранспорта. Также получена оценка расстояния до автомобиля, запечатленного в кадре.

## § 3 Описание нейросетевой модели

Модель состоит из четырех блоков, решающих следующие подзадачи:

1. Нахождение трехмерной описывающей рамки для автомобилей в кадре [6]. В этом блоке используется сверточная нейронная сеть, использующая регрессию на выходе последнего слоя.
2. Определение марки автомобилей [7].
3. Определение проекции центра нижней грани трехмерной рамки в кадре при помощи данных о расположении камеры, а так же используя матрицу камеры.
4. Оценка расстояния до автомобиля. Стоит отметить, что в решениях пунктов 3, 4 в отличие от пунктов 1 и 2, не используется нейросетевых алгоритмов.

### § 3.1 Нахождение трехмерной описывающей рамки для автомобилей в кадре

В статье [6] описывается модель, решающая первую подзадачу. Используя информацию о расположении двумерной охватывающей рамки модель находит координаты углов трёхмерной охватывающей рамки в кадре. Авторы статьи выбрали в качестве параметров для описания трехмерной рамки такие параметры, как центр предполагаемой трехмерной рамки в виде координатной тройки  $T = [t_x, t_y, t_z]$ , длины его трех сторон  $D = [d_x, d_y, d_z]$  и ориентацию  $R(\theta, \phi, \alpha)$  выбранных углов Эйлера. Всего выходит 9 характеристик.

На первом шаге входные изображения обрабатываются обученной моделью для нахождения двумерной охватывающей рамки для запечатлённых в кадре автомобилей. Обозначим  $(x_{min}, y_{min})$  - координаты левой верхней вершины и  $(x_{max}, y_{max})$  - координаты правой нижней вершины двумерной рамки. В качестве системы координат на изображении выбирается прямоугольная декартова система с началом отсчета в левом верхнем углу кадра и направлениями осей абсцисс и ординат вправо и вниз соответственно.

Для каждого автомобиля в пространстве съемки вводится своя система координат, как описано далее. В трехмерной системе координат с началом координат расположенным в центре трёхмерной рамки, можно определить координаты вершин трехмерной рамки следующим образом:  $X_1 = [\frac{d_x}{2}, \frac{d_y}{2}, \frac{d_z}{2}]$ ,  $X_2 = [-\frac{d_x}{2}, \frac{d_y}{2}, \frac{d_z}{2}]$ , ...,  $X_8 = [-\frac{d_x}{2}, -\frac{d_y}{2}, -\frac{d_z}{2}]$ . В рамках описываемой модели принимается допущение, что трёхмерная рамка плотно вписывается в двумерную охватывающую рамку, определенную ранее. Таким образом, несколько проекций вершин описывающей трёхмерной рамки будут находиться на границе двумерной рамки.

Например, рассмотрим проекцию одной из вершин описывающей трехмерной рамки, с исходными координатами  $X_0 = [-\frac{d_x}{2}, \frac{d_y}{2}, \frac{d_z}{2}]$ . Данная проекция будет касаться левой стороны двумерной охватывающей рамки с координатой  $x_{min}$ . Подобное ограничение приводит к следующему уравнению:

$$x_{min} = \left( K[R T] \begin{bmatrix} \frac{d_x}{2} \\ -\frac{d_y}{2} \\ \frac{d_z}{2} \\ 1 \end{bmatrix} \right)_x, \quad (3.1)$$

где  $(.)_x$  означает проекцию по первой координате. Аналогичные уравнения могут быть получены для остальных параметров двумерной рамки, а именно  $x_{max}, y_{min}, y_{max}$ .

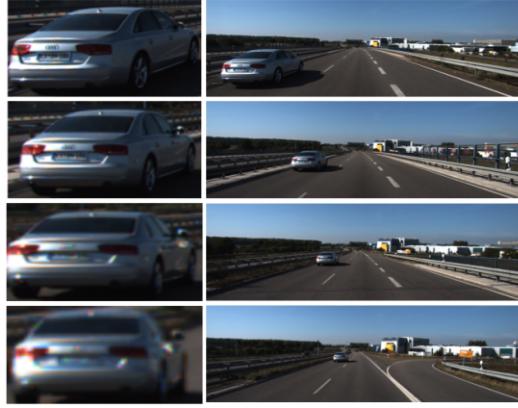
Указанные выше ограничения приводят к четырём уравнениям, которых недостаточно для вычисления девяти свободных характеристик, описанных ранее. Существует несколько геометрических свойств, которые могут быть использованы для введения дополнительных ограничений.

Авторы модели предлагают проводить оценку углов, описывающих направление охватывающей трехмерной рамки в кадре. Помимо этого, предлагается оценивать величину описывающей размеры рамки D, а не его центр T, потому что дисперсия оценки размеров, как правило, ниже (например, автомобили, как правило, имеют примерно одинаковый размер).

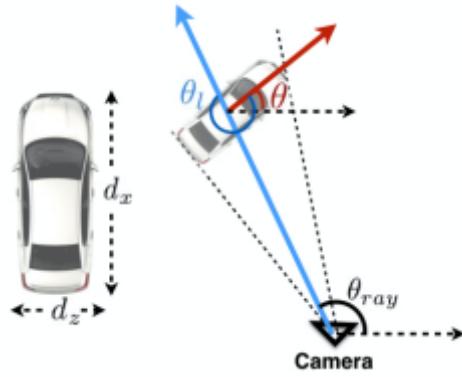
Для этого в модели используется регрессия, параметрами которой выступают параметры поворота трехмерной рамки. От выбора этих параметров в проекции трехмерной рамки будет изменять наложенность на двумерную рамку и соответствующее значение функции ошибки, которая будет описана позже.

Для определения проекции трехмерной охватывающей рамки автомобиля, запечатленного в кадре, необходимо указать восемь проекций вершин этой рамки. В случае полного перебора необходимо проверить не менее  $8^4$  вариантов соответствий, поскольку каждая вершина найденной двумерной рамки может являться одной из проекций трёхмерной охватывающей рамки. Однако, поскольку верхняя и нижняя грани проекции трехмерной рамки в кадре будут расположены в определенном порядке, а так же используя информацию об относительном крене запечатленного объекта, число соответствий может быть сокращено до 64 [6].

Самой важной частью в описываемой статье является определение ориентации трехмерной охватывающей рамки. При перемещении одного и того же реального объекта в кадре некоторые модели могут ошибаться при вычислении ориентации трёхмерной охватывающей рамки, несмотря на то, что реальное направление объекта может не изменяться. На рисунке 2 демонстрируется упомянутый эффект. Подобные ошибки возникают в силу изменений, которые вносит проективное преобразование камеры во время съемки, поэтому необходимо минимизировать погрешность при определении направле-



**Рис. 2:** Слева: обрезанное изображение проезжающей машины. Справа: изображение всей сцены. Как видно, автомобиль на обрезанных изображениях вращается, в то время как направление автомобиля является постоянным среди всех различных рядов.



**Рис. 3:** Иллюстрация локальной ориентации  $\theta_l$  и глобальной ориентации автомобиля  $\theta$ . Локальная ориентация вычисляется относительно луча, проходящего через центр двумерной рамки. Центральный луч обозначен синей стрелкой. Обратите внимание, что центр рамки может не быть фактическим центром объекта. Ориентация автомобиля  $\theta$  равна  $\theta_{ray} + \theta_l$ . Сеть обучена оценивать локальную ориентацию  $\theta_l$ .

ния трёхмерной рамки. В связи с этим рассматривается лишь угол поворота азимута для матрицы  $R$ , то есть сумму двух углов  $\theta_{ray} + \theta_l$ , что подробнее изображено на Рис.3.

В представленной модели решение проблемы нахождения этого угла  $\theta_l$ , которое может принимать значения в промежутке от 0 до 360 градусов, заключается в делении этого промежутка на условные  $n$  пересекающихся пронумерованных подотрезков  $\Delta\theta_i$ . Для каждого подотрезка сверточная нейронная сеть оценивает вероятность  $c_i$  того, что выходной угол находится внутри  $i$ -го подотрезка и поправку на остаточное вращение, которую необходимо применить к ориентации луча на этом подотрезке, чтобы получить выходной угол  $\theta_l$ . Остаточное вращение представлено двумя числами, для синуса и косинуса угла. Таким образом на выходе сети получаем три числовых параметра для каждого подотрезка  $i$ :  $(c_i, \cos(\Delta\theta_i), \sin(\Delta\theta_i))$ . Оценка этих параметров происходит путем применения функции потерь следующего вида:

$$L_\theta = L_{conf} + \omega * L_{loc},$$

где

$$L_{conf} = \sum_i^N L_{conf_i} = \frac{e^{a_i}}{\sum_{k=1}^N e^{a_k}} \quad (3.2)$$

есть softmax, и на выводе получаем вероятности нахождения на каждом подотрезке, то есть параметров  $c_i$ , где  $\omega$  - числовой гиперпараметр.  $L_{loc}$  высчитывается следующим образом:

$$L_{loc} = -\frac{1}{n_{\theta^*}} \sum \cos(\theta^* - c_i - \Delta\theta_i),$$

где  $n_{\theta^*}$  является количеством подотрезков, которые покрывают истинный угол  $\theta^*$ , а  $\Delta\theta_i$  - это изменение, которое необходимо применить к центру подотрезка  $i$ .

В качестве результата работы нейронной сети выбирается ячейка с максимальной вероятностью, и окончательный результат вычисляется путем применения оценочного значения  $\Delta\theta$  этого подотрезка к центру этого подотрезка. В результате необходимо оценить  $3n$  параметров для  $n$  подотрезков. После подсчета всех параметров остается лишь посчитать решение системы, которая была описана выше, состоящая из четырех уравнений для  $x_{min}, x_{max}, y_{min}, y_{max}$ .

### § 3.2 Определение марки автомобиля и визуализация

Модель работает на слоях предобученной модели ResNet 152 [9], состоящей из 152 слоев нейронной сети, являющейся комбинацией сверточных нейронных сетей с различной шириной каждого слоя и операцией субдискретизации (англ. pooling), подробнее изображено на схеме(Рис. 4). Модель дообучается на последнем слое, корректируя веса на размеченных по моделям автомобилей данных, предложенных стэнфордским университетом [11]. Для того чтобы данные соответствовали входному набору сети, выполняется предобработка картинок способом обрезки с отступом от краев для получения изображений размеров 244x244 пикселя.

После расчета вершин трехмерной рамки по представленным вершинам двумерной рамки вырезается картинка из исходного изображения. Так как двумерная рамка слишком плотно прилегает к объекту, делается небольшой отступ в сторону увеличения относительно центра. Объясняется это тем, что на обучении для второй сети в тренировочной базе данных использовались картинки хорошего качества, где объект изображен с четкими границами. Модель работает на сверточной нейронной сети и высчитывает классификацию среди 196 марок автомобилей, находящихся в базе данных, предложенных стэнфордским университетом [11]. Вывод данной части программы добавляется в общий вывод модели.

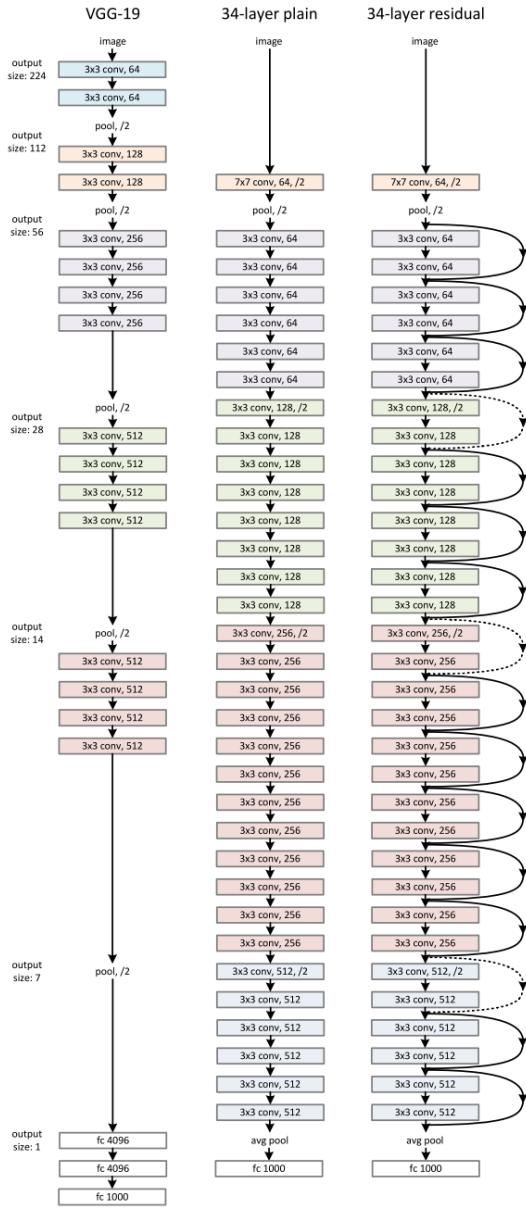


Рис. 4: Схема сети ResNet 152.

## § 4 Оценка расстояния до автомобиля

В настоящем разделе представлена геометрическая модель, позволяющая оценить расстояние до автомобиля, запечатленного в кадре, используя информацию, полученную ранее при помощи нейросетевых моделей.

### § 4.1 Основные обозначения

Заранее известны некоторые параметры камеры, а именно: высота, на которой находится объектив, вертикальный и горизонтальный углы обзора. Также известно из вывода модели, описанной ранее, каковы возможный производитель и марка автомобиля, запечатленного в кадре. В рамках настоящей работы была составлена база данных о реальных физических размерах автомобиля, то есть о высоте, ширине и длине. Так-

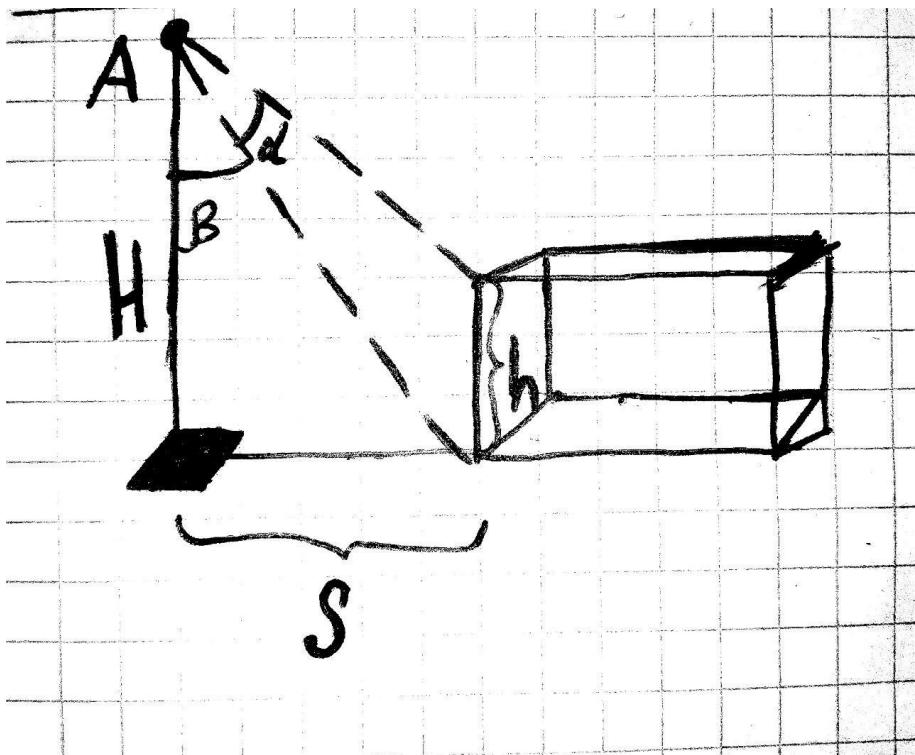


Рис. 5: Рисунок 1

же, как описано ранее, получена информация о проекциях трехмерных охватывающих рамок для автомобилей в кадре

#### § 4.2 Модель оценки расстояния до автомобиля

Обозначим через  $H$  высоту, на которой расположена камера относительно поверхности земли, высоту автомобиля  $h$ , точку, в которой находится объектив камеры  $A$ , а точку перпендикулярной проекции точки  $A$  на плоскость земли обозначим через  $D$ . Выберем одну из вертикальных граней трехмерной рамки и обозначим его концы как  $B$  и  $C$ , где  $C$  лежит на плоскости земли, угол  $\angle BAC = \alpha$ , а угол  $\angle CAD = \beta$ . Расстояние от точки  $D$  до  $C$  обозначим  $s$  - что требуется найти.

Оценим величину угла  $\alpha$  следующим образом: высчитаем количество пикселей на один радиан угла обзора камеры, полученное соотношением вертикального размера изображения на вертикальный угол обзора камеры. Далее умножим это значение на длину проекции ребра  $BC$  на изображение в пикселях.

В таком случае получаем:

1.  $\operatorname{tg}(\beta) = \frac{s}{H}$
2.  $s = (H - h)\operatorname{tg}(\alpha + \beta)$

Тогда

$$s = (H - h) * \frac{\operatorname{tg}(\alpha) + \operatorname{tg}(\beta)}{1 - \operatorname{tg}(\alpha)\operatorname{tg}(\beta)} = (H - h) * \frac{\operatorname{tg}(\alpha) + \frac{s}{H}}{1 - \operatorname{tg}(\alpha)\frac{s}{H}}$$

$$(1 - \operatorname{tg}(\alpha)\frac{s}{H})s = (H - h)(\operatorname{tg}(\alpha) + \frac{s}{H})$$

$$-s^2 \frac{tg(\alpha)}{H} + s = (H - h)tg(\alpha) + \frac{s}{H}(H - h)$$

В итоге получаем квадратное уравнение:

$$\frac{tg(\alpha)}{H} * s^2 - \frac{h}{H} * s + (H - h)tg(\alpha) = 0$$

$$s_{1,2} = \left( \frac{h}{H} \pm \sqrt{\frac{h^2}{H^2} - 4tg^2(\alpha)\frac{H-h}{H}} \right) * \frac{H}{2tg(\alpha)}$$

Из-за ограничений на дискриминант появляется ограничение на значение  $\alpha$ , т.е.:

$$\frac{h^2}{H^2} \geqslant 4tg^2(\alpha)\frac{(H-h)}{H}$$

$$h^2 \geqslant 4tg^2(\alpha)(H-h)H$$

$$tg^2(\alpha) \leqslant \frac{h^2}{4H(H-h)}$$

Так как  $H > h$ , то:

$$|tg(\alpha)| \leqslant \frac{h}{\sqrt{4H(H-h)}}$$

$$arctg\left(-\frac{h}{\sqrt{4H(H-h)}}\right) \leqslant \alpha \leqslant arctg\left(\frac{h}{\sqrt{4H(H-h)}}\right)$$

При  $H > h$  получаются допустимы оба значения  $s_{1,2}$ , т.к.:

$$0 > -4tg^2(\alpha)\frac{H-h}{H}$$

$s$  будет являться строго положительным числом, поскольку  $H - h > 0$

При  $H < h$  можно заметить, что описанные уравнения 1), 2) останутся такими же.

Дискриминант в этом случае является строго положительным, что исключает ограничение угла  $\alpha$ . А положительность значения  $s$  исключает возникновение второго корня, так как рассуждения схожи для случая  $H > h$ .

### Геометрический смысл двух значений оценки расстояния $s$ при $H > h$

Проведем через три пространственные точки А, В и С окружность. Она задается однозначно. Далее проведем через точку А прямую, параллельную поверхности земли так, чтобы она пересекла нашу окружность. Обозначим ее точкой А'. Эта прямая неизбежно является касательной к нашей окружности, что следует из построения. Рассмотрим случай, когда такая точка А' существует. Из свойств окружности и хорд следует, что угол  $\alpha = \angle CAB = \angle CA'B$ , расстояние от поверхности земли до точки А' = Н. Расстояние от точки проекции точки А' на плоскость земли до грани ВС равно  $s' \neq s$ . Если точки А' не существует, значит дискриминант из предыдущих рассуждений

равен нулю, значит решение единствено.

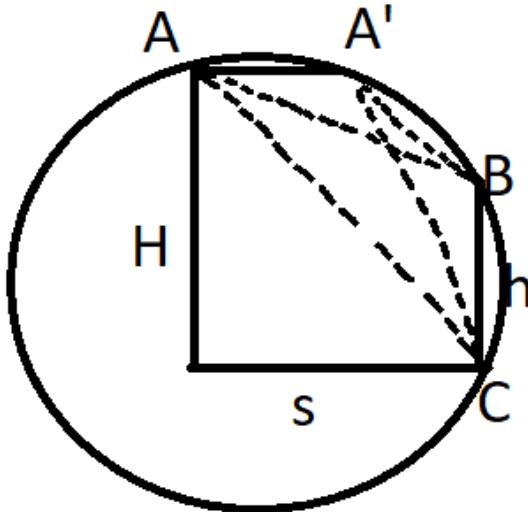


Рис. 6: Рисунок 2

Так как мы владеем информацией о реальных размерах автомобиля, то одним из подходов можно понять, какое из полученных двух решений подходит больше всего:

1. Найдем таким же способом расстояние  $s'_{1,2}$  до другой параллельной грани трехмерной рамки.
2. Найдем 4 разности между потенциальными расстояниями между двумя параллельными гранями.
3. Сравними с реальным значением расстояния между этими гранями, пользуясь знаниями о характеристиках автомобиля.
4. Возьмем из 4 разностей самое близко подходящее к этому значению.

Таким образом мы подберем расстояние до первой грани и задача решена.

Если объект находится далеко, то будут получаться два числовых значения, модуль разности которых будет заведомо больше или меньше определенного заранее заданного порога. Если модуль разности будет больше порога, то меньший из этих корней можно исключить.

## § 5 Модель оценки центра нижней грани трехмерной рамки и оценка точности

Обозначим  $X_1, X_2, X_3, X_4$  координатные пары точек вершин нижней грани трехмерной рамки. Существует несколько способов вычисления приближенного положения центра нижней грани трехмерной рамки. В качестве примера можно рассмотреть подход, в котором вычисляется точка пересечения диагоналей четырехугольника, однако такой подход может давать смещенные результаты, в случае если проекция нижней грани не окажется параллелограммом. В настоящей работе координата центра нижней грани трехмерной рамки приближенно оценивается как среднее арифметическое

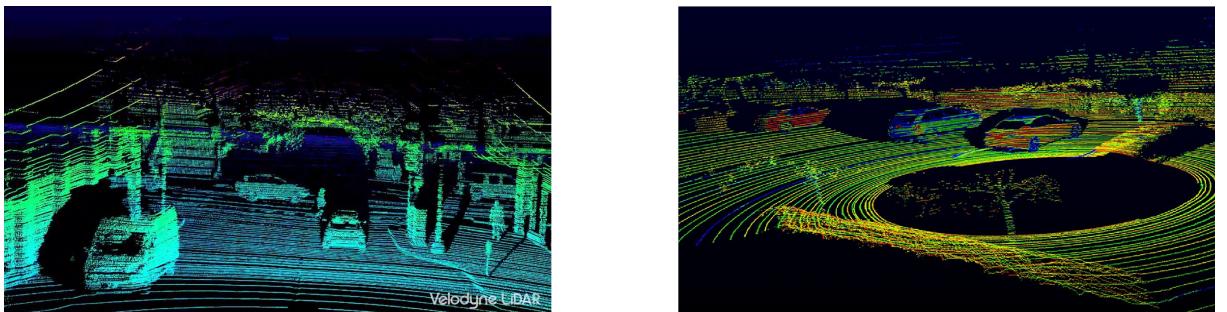
координатных пар вершин нижней грани, то есть

$$X_{center} = \frac{1}{4} \sum_{i=1}^4 X_i$$

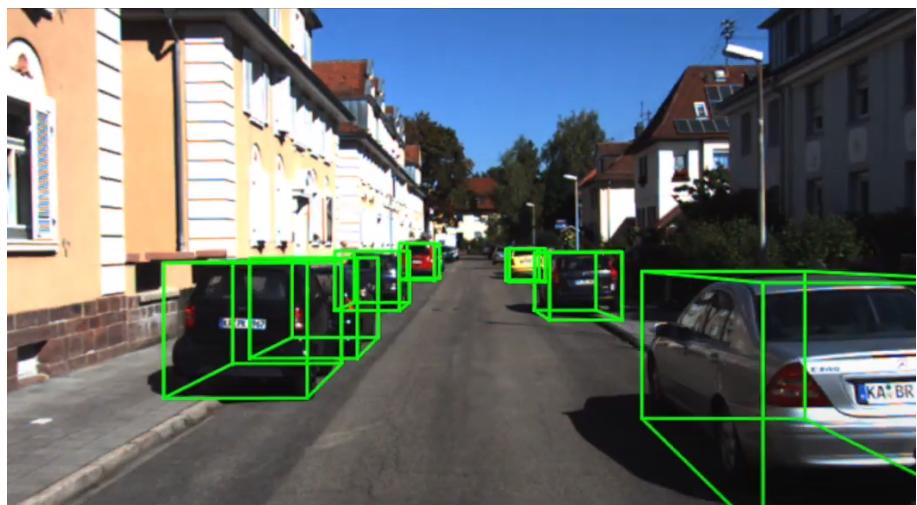
### § 5.1 Оценка точности

Для определения погрешности модели был использован вывод модели построения трехмерной рамки (проверочная модель), использующей данные об облаках точек, полученных с устройства, называемого LiDAR [11], которое с помощью лазеров и сенсоров определяет координаты точек на препятствиях вокруг. Таким образом получается полное представление о трехмерной картине, поскольку использует дополнительную информацию о расположении автомобиля. Такая модель очень точно описывает трехмерную рамку, соответственно и центр ее нижней грани.

В базе данных, которые были использованы для обучения модели, описанной в данной работе, имелись данные о LiDAR, но не были использованы, так как модель опиралась лишь на графическую составляющую.



**Рис. 7:** Визуализация облаков точек, используемых в модели



**Рис. 8:** Пример вывода проверочной модели

Обозначим  $X'$  координатную пару на изображении точки центра нижней грани трехмерной рамки из вывод проверочной модели,  $X$  координатную пару на изображении

точки центра нижней грани трехмерной рамки из вывода модели в данной работе,  $X_i$  - как координатные пары вершин нижней грани трехмерной рамки. Введем обозначение

$$K = \left(1 - \frac{\|X' - X\|}{q}\right) * 100\%$$

где  $\| * \|$  - норма вектора,  $q = \max_i \|X_i - X_{i-1}\|$ .

Оценка качества определения точки была определена, как

$$L = \begin{cases} K & K \geq 0 \\ 0 & K < 0 \end{cases}$$

Значением 0 отсеиваются все те результаты, которые полностью классифицированы неверно.

Для каждого тестового изображения был составлен столбец со значениями оценки качества расчета центра нижней грани трехмерной рамки для каждого автомобиля, который модель смогла распознать. Всего для расчетов точности были использованы данные о центрах нижних граней трехмерных рамок 46 автомобилей из вывода проверочной модели. В представленной таблице в каждой строке каждого столбца представлены значения чисел  $L$  для каждого автомобиля из соответствующего изображения.

№объекта	Рис 1	Рис 2	Рис 3	Рис 4	Рис 5	Рис 6	Рис 7	Рис 8	Рис 9
1	84%	82%	89%	59%	69%	87%	99%	70%	71%
2	72%	0%	71%		70%	80%	63%	58%	44%
3	93%	72%	78%			75%	74%		71%
4	53%	62%	60%						74%
5	93%	89%	84%						

№объекта	Рис 10	Рис 11	Рис 12
1	77%	90%	85%
2	55%	52%	78%
3		0%	79%
4		83%	72%
5		81%	92%
6		72%	
7		94%	
8		76%	
9		0%	

И высчитаны средние по всем тестам без учёта нулевых значений и с учетом нулевых значений оценки точности соответственно.

Итоговое среднее без учета 0 %	Итоговое среднее с учетом 0%
75.3%	70.1%

## § 5.2 Оценка погрешности дискретизации

Далее будем рассматривать целочисленные координатные пары, как пары случайных величин  $\xi_i = (\xi_1^i, \xi_2^i)$  подчиненных некоторому вероятностному распределению с математическим ожиданием  $M$  с дисперсией  $\mathcal{D}\xi_i = D$ . Корень из дисперсии - среднеквадратичное отклонение - можно рассматривать как оценку погрешности дискретизации. Дисперсию центра нижней грани, определенной как пары случайных величин  $\eta = (\eta_1, \eta_2)$ , равна

$$\mathcal{D}\eta = \mathcal{D}\left(\frac{\xi_1 + \xi_2 + \xi_3 + \xi_4}{4}\right) = \frac{\mathcal{D}\xi_1}{4} = \frac{D}{4}$$

Таким образом среднеквадратичное отклонение центра нижней грани будет равно  $\frac{\sqrt{D}}{2}$  - получаем снижение среднеквадратичного отклонения от реальных координат точек на изображении в два раза при вычислении центра нижней грани, по сравнению с погрешностью дискретизации вершин рамки.

## § 6 Заключение

В данной работе исследуется возможность применения ряда нейросетевых моделей в качестве практического инструмента для отслеживания автомобилей в видео ряде. Также приведена математическая модель оценки расстояний до запечатленных автомобилей. Исходный код модели расположен. Модель можно запустить по адресу <https://github.com/TimurKinazar/3DBoundingBox>

В ходе выполнения работы был определен ряд проблем, которые требуют дальнейшего исследования. В частности, если снимать покадрово видео и применять метод, описанный в настоящей работе, то он высчитывает трехмерные рамки с видом человеческим взглядом разрывом, то есть при гладком изменении покадрово объекта довольно резко меняются вершины трехмерной рамки, что влияет на оценку нахождения центра нижней грани. Также вывод второй нейронной сети, вычисляющей марку автомобиля, сильно отличается от того, что в реальности изображено в плане вычисления марки автомобиля. Проблема заключается в низком качестве вырезаемого изображения, на котором находится только изучаемый объект, и в малом количестве марок автомобилей в обучающем датасете, примененных для модели распознавания марки автомобиля.

В перспективе можно улучшить сеть, если заменить часть оценки проекции трехмерной рамки на сеть, способную произвести расчет размеров и ориентации объекта с помощью восстановления трехмерной модели из класса набора данных Pascal3D+.

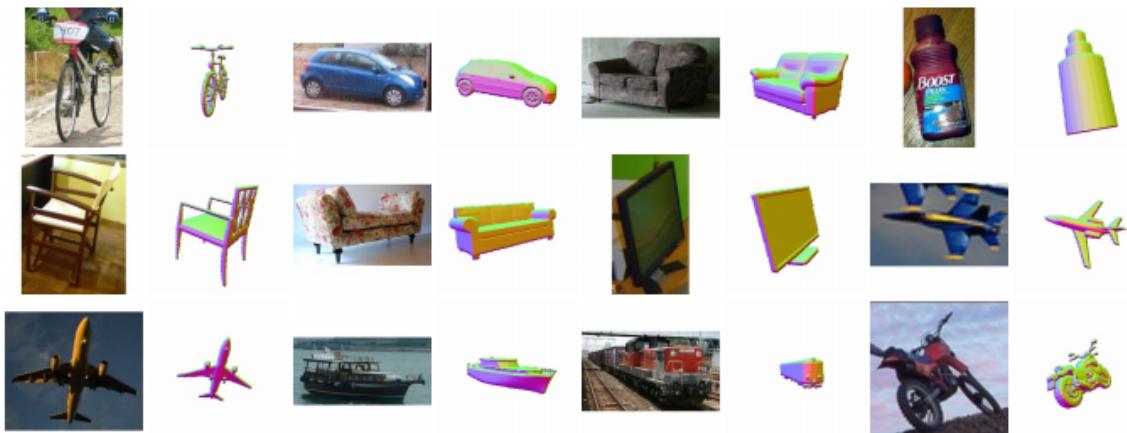


Рис. 9: Визуализация моделей в наборе данных Pascal3D+.

Отдельную благодарность за помощь в написании данной работы хотелось бы выразить аспиранту механико-математического факультета кафедры математической теории интеллектуальных систем Ронжину Дмитрию Владимировичу, а также научному руководителю доценту кафедры Математической теории интеллектуальных систем, кандидат физико-математических наук Часовских Анатолию Александровичу.

# Литература

1. Chenge Li, Gregory Dobler, Xin Feng, Yao Wang *TrackNet: Simultaneous Object Detection and Tracking and Its Application in Traffic Video Analysis* — In Computer Vision and Pattern Recognition 2019
2. Xiaoliang Wang, Peng Cheng, Xinchuan Liu, Benedict Uzochukwu *Focal Loss Dense Detector for Vehicle Surveillance* — Computer Vision and Pattern Recognition 2018
3. Ронжин Д.В. *Распознавание динамики особых точек в видеоряде* — Журнал: Интеллектуальные системы. Теория и приложения Том: 18 Номер 2 2014г. 267-276стр
4. Florian Chabot, Mohamed Chaouch, Jaonary Rabarisoa *Deep MANTA: A Coarse-to-fine Many-Task Network for joint 2D and 3D vehicle analysis from monocular image* — In CVPR 2017
5. Jason Ku, Alex D. Pon, Steven L. Waslander *Monocular 3D Object Detection Leveraging Accurate Proposals and Shape Reconstruction* — In CVPR 2019
6. Arsalan Mousavian, Dragomir Anguelov, John Flynn, Jana Kosecka *3D Bounding Box Estimation Using Deep Learning and Geometry.* — In IEEE CVPR 2017.
7. Nithiroj T. *Car Recognition* — <https://github.com/nithiroj/car-recognition>
8. P. L. A. Geiger and R. Urtasun. *Are we ready for autonomous driving? the KITTI vision benchmark suite.* In CVPR, 2012.
9. Kaiming He, Xianyu Zhang, Shaoqing Ren, Jian Sun *Deep Residual Learning for Image Recognition* — In IEEE CVPR 2015.
10. Soroush. *PyTorch implementation for this paper.* — <https://github.com/skhadem/3D-BoundingBox>
11. A Benchmark for 3D Object Detection in the Wild — <http://cvgl.stanford.edu/projects/pascal3d.html>
12. Yu Xiang, Roozbeh Mottaghi, Silvio Savarese *Beyond PASCAL: A Benchmark for 3D Object Detection in the Wild* — In WACV, 2014

13. Zhaowei Cai, Quanfu Fan, Rogerio S. Feris, Nuno Vasconcelos *A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection* — In ECCV, 2016
14. Karen Simonyan, Andrew Zisserman *Very Deep Convolutional Networks for Large-Scale Image Recognition* — CoRR, abs/1409.1556, 2014
15. Stanford Dataset for Car Recognition — <https://ai.stanford.edu/~jkrause/cars/>
16. Vladimir V. Kniaz Peter V. Moshkantsev Vladimir A. Mizginov *Deep Learning a Single Photo Voxel Model Prediction from Real and Synthetic Images* In XXI International Conference on Neuroinformatics, October 7-11, 2019, Dolgoprudny, Moscow region, Russia
17. M.Zeeshan Zia, Michael Stark, Konrad Schindler *Towards Scene Understanding with Detailed 3D Object Representations* — International Journal of Computer Vision 2014

## Приложение

Пример вывода нейросетевой модели, строящую  
описанную трехмерную рамку



Рис. 10: Пример вывода 1.

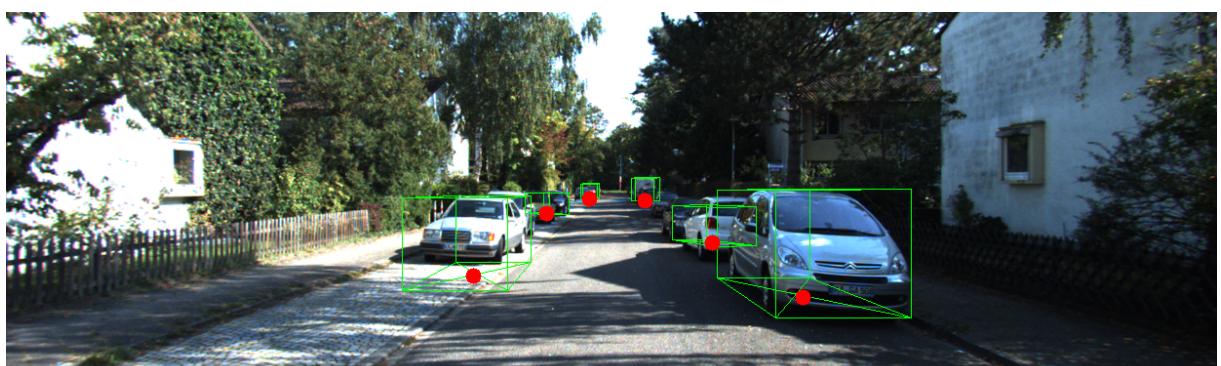


Рис. 11: Пример вывода 2.

## Приложение

### Пример вывода модели нахождения марки автомобиля и расстояния до объекта

```
Estimated pose: 929 214 1243 372
Coordinate of the center of mass ( 1051 , 427 )
Possible car model: Volkswagen Golf Hatchback 1991
Distance to car: 1103.2422585479553 mm
```

```
Estimated pose: 345 183 618 346
Coordinate of the center of mass ( 500 , 326 )
Possible car model: Nissan 240SX Coupe 1998
Distance to car: 3090.3816522604325 mm
```

```
Estimated pose: 601 177 720 258
Coordinate of the center of mass ( 666 , 266 )
Possible car model: Audi 100 Wagon 1994
Distance to car: 3969.383076881309 mm
```

**Рис. 12:** Пример вывода текстовой составляющей.