

Аннотация

Документ представляет собой техническое задание к выпускной работе бакалавра на тему «Портирование сверточной нейросети на ARM архитектуру с ограниченными вычислительными ресурсами и ресурсами памяти», выполненную студентом группы ИВТ-461, Мельниковым Тимофеем Алексеевичем.

Составлено и оформлено согласно ГОСТ 19.201-78.

Объём технического задания составил 21 страниц и включает 2 рисунка и 2 таблицы.

Министерство образования и науки Российской Федерации  
ФГБ ОУ ВПО «Волгоградский государственный технический  
университет»

Кафедра «Системы автоматизированного проектирования и ПК»

Утверждаю

и.о. зав. кафедрой САПРиПК,

д.т.н., проф.

_____	М. В. Щербаков
(подпись)	(инициалы, фамилия)
«_____»	_____ 2017

Портирование сверточной нейросети на ARM архитектуру с  
ограниченными вычислительными ресурсами и ресурсами памяти

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

ВРБ-40 461 806-10.27-10-17.91-91

Листов 21

Научный руководитель

к.т.н., доцент каф. САПРиПК

\_\_\_\_\_ А. В. Катаев

«\_\_\_\_\_» \_\_\_\_\_ 2017

Нормоконтролер

к.т.н., доцент каф. САПРиПК

\_\_\_\_\_ О. А. Шабалина

«\_\_\_\_\_» \_\_\_\_\_ 2017

Исполнитель

студент группы ИВТ-461

\_\_\_\_\_ Т. А. Мельников

«\_\_\_\_\_» \_\_\_\_\_ 2017

Волгоград, 2017

Утверждаю

и.о. зав. кафедрой САПРиПК,

д.т.н., проф.

_____	М. В. Щербаков
(подпись)	(инициалы, фамилия)
«_____»	_____ 2017

Портирование сверточной нейросети на ARM архитектуру с  
ограниченными вычислительными ресурсами и ресурсами памяти

ЛИСТ УТВЕРЖДЕНИЯ

А.В.00001-01 91 01-1-ЛУ

Листов 1

Научный руководитель

к.т.н., доцент каф. САПРиПК

\_\_\_\_\_ А. В. Катаев

«\_\_\_\_\_» \_\_\_\_\_ 2017

Исполнитель

студент группы ИВТ-461

\_\_\_\_\_ Т. А. Мельников

«\_\_\_\_\_» \_\_\_\_\_ 2017

Волгоград, 2017

## СОДЕРЖАНИЕ

1 Введение	5
1.1 Наименование программы	5
1.2 Область применения	5
2 Основание для разработки	5
2.1 Документ, на основании которых ведется проектирование	5
2.2 Организация, утвердившая этот документ, и дата его утверждения	6
2.3 Наименование и условное обозначение темы разработки	6
3 Назначение разработки	6
4 Требование к программе	7
4.1 Требования к функциональным характеристикам	7
4.1.1 Требования к составу выполняемых функций	7
4.2 Организация входных и выходных данных	9
4.2.1 Входные данные	9
4.2.2 Выходные данные	10
4.3 Требования к надежности	11
4.3.1 Требования к надежному функционированию	11
4.3.2 Время восстановления после отказа	11
4.3.3 Требования к составу и параметрам технических средств	12
4.4 Требования к информационной и программной совместимости	13
4.4.1 Требования к методам решения	13
4.4.2 Требования к языкам программирования	13
5 Требования к программной документации	13
6 Стадии и этапы разработки	14
6.1 Стадии разработки	14
6.2 Этапы разработки	14
7 Порядок контроля и приемки	15
7.1 Виды испытаний	15
7.2 Общие требования к приемке	16
Приложение 1 — Формат конфигурационных файлов нейронной сети	17

Приложение 2 — Диаграмма последовательности	18
Приложение 3 — Визуализация локализации объектов	20
Приложение 4 — Макеты экранных форм	21

## 1 Введение

### 1.1 Наименование программы

Разработке подлежит программный продукт, представляющий собой клиент-серверное приложение для детектирования объектов на изображении, используя сверточную нейронную сеть, на устройствах с ARM-архитектурой. Полное наименование приложения — "Приложение для прямого прохода сверточной нейронной сети на ARM-устройстве". Краткое наименование — "С.Н.I.P. Vision". Далее используется краткое название — Приложение.

### 1.2 Область применения

Приложение предназначено для использования компьютерного зрения, основанного на машинном обучении, в встраиваемых системах с архитектурой ARM.

## 2 Основание для разработки

### 2.1 Документ, на основании которых ведется проектирование

Разработка ведется на основании задания на выполнение выпускной квалификационной работы бакалавра по направлению «Информатика и вычислительная техника».

2.2 Организация, утвердившая этот документ, и дата его утверждения

Задание на выполнение выпускной квалификационной работы бакалавра выдано к.т.н., доцентом кафедры «Системы автоматизированного проектирования и поискового конструирования» Катаевым А. В.

Задание выдано «\_\_» \_\_\_\_\_ 2016 г.

Срок окончания работ «\_\_» \_\_\_\_\_ 2017 г.

2.3 Наименование и условное обозначение темы разработки

Наименование темы разработки — Приложение осуществляющее прямой проход сверточной нейронной сети на устройстве с ARM-архитектурой.

### 3 Назначение разработки

Приложение предназначено для исследования вычислений сверточных нейронных сетей на ARM-архитектуре, выполняющих детектирование объектов на изображении, и визуализации выполненных вычислений на клиентском компьютере.

## 4 Требование к программе

### 4.1 Требования к функциональным характеристикам

Приложение разделено на клиентскую и серверную части. Серверная часть запускается на ARM-устройстве и выполняет детектирование объектов на изображении, используя сверточные нейронные сети, и функции для работы с клиентом. Клиентская часть запускается на персональном компьютере пользователя и предоставляет возможность подключаться к серверной программе, выбирать изображение для детектирования на нем объектов и просматривать результаты работы детектирования объектов.

#### 4.1.1 Требования к составу выполняемых функций

Серверная программа должна выполнять следующие группы функций:

- Функции работы с клиентом;
- Функции парсинга и сериализации конфигураций нейронной сети;
- Функции детектирования объектов на изображении.

Для работы с клиентом должны быть реализованы следующие функции:

- Создание сокета для входных подключений;
- Связка сокета с сетевым оборудованием;
- Создание клиентского сокета, после подключения клиента;
- Ожидание и выполнение команд клиента.



Для реализации парсинга и сериализации данных должны быть реализованы следующие функции:

- Парсинг и сериализация меток детектируемых объектов. Формат файла меток представлен в приложении 1, рисунок 1.2;
- Парсинг и сериализация файла конфигурации сверточной нейронной сети YOLO. Файл конфигурации представлен в приложении 1, рисунок 1.2;
- Парсинг и сериализация весов сверточной нейронной сети Tiny YOLO. Веса хранятся последовательно в двоичном формате.

Для детектирования объектов на изображении необходимо реализовать следующие функции:

- Сериализация изображения для детектирования объектов на нем;
- Прямой проход следующих слоев:
  - а) Сверточный слой;
  - б) Слой объединения;
  - в) Слой контуров.
- Активация следующих нелинейных функций:
  - а) Жесткая пороговая функция;
  - б) Линейная функция;
  - в) Leaky функция.
- Локализация найденных объектов;
- Отрисовка меток и границ объектов.

Клиентская программа должна выполнять следующие группы функций:

- Функции работы с сервером;
- Функции взаимодействия с пользователем.

Для работы с сервером необходимо реализовать следующие функции:

- Получение хоста сервера;
- Соединение с сервером.

Для взаимодействия с пользователем необходимо реализовать следующие функции:

- Соединение с сервером;
- Выбор изображения для детектирования объектов на нем;
- Отображение процесса работы прямого прохода Tiny YOLO;
- Отображение результатов прямого прохода Tiny YOLO. А именно изображение с метками и отрисованными границами объектов.

Для осуществления передачи данных между клиентом и сервером необходимо реализовать интерфейс работы с сокетами. Данный модуль является общим для клиента и сервера и должен выполнять следующие функции:

- Отправка сообщения;
- Получение сообщения;
- Отправка изображения;
- Получение изображения.

## 4.2 Организация входных и выходных данных

### 4.2.1 Входные данные

Входными данными для клиентской части приложения является изображение в формате png или jpg. Любой файл иного формата, но с расширением png или jpg, открываться не должен. В процессе вычислений серверная программа должна отправлять клиенту информацию о пройденных этапах вычислений. В приложении 2 описан процесс обмена данными между клиентской и серверной частями.

Входными данными для серверной части приложения является сообщение от клиента. В таблице 1 приведены входные данные запрашиваемые сервером в зависимости от полученного сообщения.

Таблица 1 — Входные данные серверной части приложения

Сообщение	Входные данные
"yolo"	изображение в формате png или jpeg в байтовом представлении
"exit"	—
другое сообщение	сообщение с корректной командой ("yolo"или "exit")

#### 4.2.2 Выходные данные

Выходными данными приложения является изображение с метками найденных объектов. Каждый найденный объект должен быть замкнут прямоугольником. Подробная информация об визуализации выходного изображения представлена в приложении 3.

Клиентская часть программы должна иметь текстовый браузер для вывода информации о основных этапах работы серверной части приложения. В приложении 2 на диаграмме последовательности показано, какая информация о этапах серверной части должна быть выведена.

В таблице 2 приведены выходные данные серверной части, в зависимости от сообщения.

Серверная часть приложения должна выводить в стандартный поток ввода/вывода основную информацию о этапах работы. В приложении 2 на диаграмме последовательности показаны информационные блоки, которые должны выводиться в стандартный поток ввода/вывода.

Таблица 2 — Выходные данные серверной части приложения

Сообщение	Входные данные
"yolo"	изображение в формате png

Таблица 2 — Выходные данные серверной части приложения

Сообщение	Входные данные
	с отрисованными метками и границами объектов в байтовом представлении
"exit"	—
другое сообщение	— —

### 4.3 Требования к надежности

#### 4.3.1 Требования к надежному функционированию

Надежное функционирование приложения должно быть обеспечено выполнением Заказчиком совокупности организационно-технических мероприятий, перечень которых приведен ниже:

- организацией бесперебойного питания технических средств
- использованием лицензионного программного обеспечения

#### 4.3.2 Время восстановления после отказа

Время восстановления после отказа, вызванного неисправностью технических средств, фатальным сбоем (крахом) операционной системы, не должно превышать времени, требуемого на устранение неисправностей технических средств и переустановки программных средств.

#### 4.3.3 Требования к составу и параметрам технических средств

Эксплуатация клиентской части приложения конечным пользователем подразумевает у него наличие Linux-подобной операционной системы с подключением к интернету. В операционной системе должна быть предустановлена библиотека Qt 5.8.0.

Ниже приведены требования к техническим средствам компьютера:

- процессор мощностью не менее 1 ГГц;
- оперативная память не менее 512 Мб;
- свободное место не менее 100 Мб;
- устройства взаимодействия с пользователем — клавиатура, мышь и монитор.

Запуск серверной части приложения должен осуществляться на встраиваемом решении с процессором ARM. Должны поддерживаться следующие семейства процессоров:

- ARM7
- ARM9
- Cortex A

Устройство должно обладать Linux-подобной операционной системой с подключением к интернету.

Ниже приведены требования к техническим средствам компьютера:

- процессор мощностью не менее 60 МГц;
- оперативная память не менее 512 Мб;
- свободное место не менее 500 Мб;
- устройства взаимодействия с пользователем — клавиатура и монитор.

#### 4.4 Требования к информационной и программной совместимости

##### 4.4.1 Требования к методам решения

Методы решения должны обеспечивать выполнение всех этапов проектирования приложения в соответствии с их порядком и сроками выполнения, указанными в разделе 7 данного документа.

##### 4.4.2 Требования к языкам программирования

Клиентская часть должна быть написана на языке C/C+.

Серверная часть должна быть написана на языке С. Серверная часть не должна содержать внешних зависимостей. Разрешается использовать только системные библиотеки.

#### 5 Требования к программной документации

К приложению прилагается следующая документация:

- Техническое задание согласно ГОСТ 19.201-78;
- Пояснительная записка.

## 6 Стадии и этапы разработки

### 6.1 Стадии разработки

Разработка должна включать следующие стадии:

- Изучение используемого фреймворка нейронных сетей «darknet» (ноябрь-декабрь);
- Изучение архитектур сверточных нейронных сетей: «YOLO», Tiny YOLO»;
- Разработка технического задания (февраль);
- Рабочее проектирование (февраль-март);
- Реализация и тестирование программы (апрель-май)

### 6.2 Этапы разработки

На стадии изучения фреймворка «darknet» должны быть выполнены следующие этапы:

- Изучить функции работы с изображениями и способ представления изображения;
- Изучить функции чтения конфигурации нейронной сети, весовых коэффициентов;
- Изучить построение архитектуры работы слоев и связи между ними.

На стадии разработки технического задания должны быть выполнены следующие этапы:

- Разработка технического задания;
- Согласование и утверждение технического создания.

На стадии рабочего проектирования должны быть выполнены перечисленные ниже этапы:

- Разработка клиентской части приложения;
- Разработка серверной части приложения;
- Реализация интерфейса взаимодействия клиентской и серверной части.

На стадии реализации программы должны быть выполнены перечисленные ниже этапы:

- Реализация вертикального прототипа;
- Доработка прототипа до конечного продукта.

На стадии тестирования программы должны быть выполнены перечисленные ниже этапы:

- Анализ корректности вычислений при прямом проходе сверточной нейронной сети Tiny YOLO;
- Анализ быстродействия программы вычислений на ARM-устройстве.

## 7 Порядок контроля и приемки

### 7.1 Виды испытаний

Испытания программы и верификация документации должны проводиться в организации заказчика.

Приемо-сдаточные испытания приложения должны производиться доцентом кафедры САПриПК Катаевым А. В.

Программа должна соответствовать всем требованиям, изложенным в Техническом задании.



## 7.2 Общие требования к приемке

Приемка программы должна производиться доцентом кафедры САПРиПК Катаевым А. В.

Программа должна считаться годной для приемки, если в процессе тестирования заказчиком она удовлетворяет всем пунктам данного технического задания.

## Формат конфигурационных файлов нейронной сети

```

classes= 20 _____ количество классов
train  = /home/pjreddie/data/voc/train.txt _____ обучающая выборка
valid  = /home/pjreddie/data/voc/2007_test.txt _____ тестовая выборка
names  = data/voc.names _____ метки классов
backup = backup _____ директория промежуточных файлов

```

Рисунок 1.1 — Данные о выборке и метках объектов

```

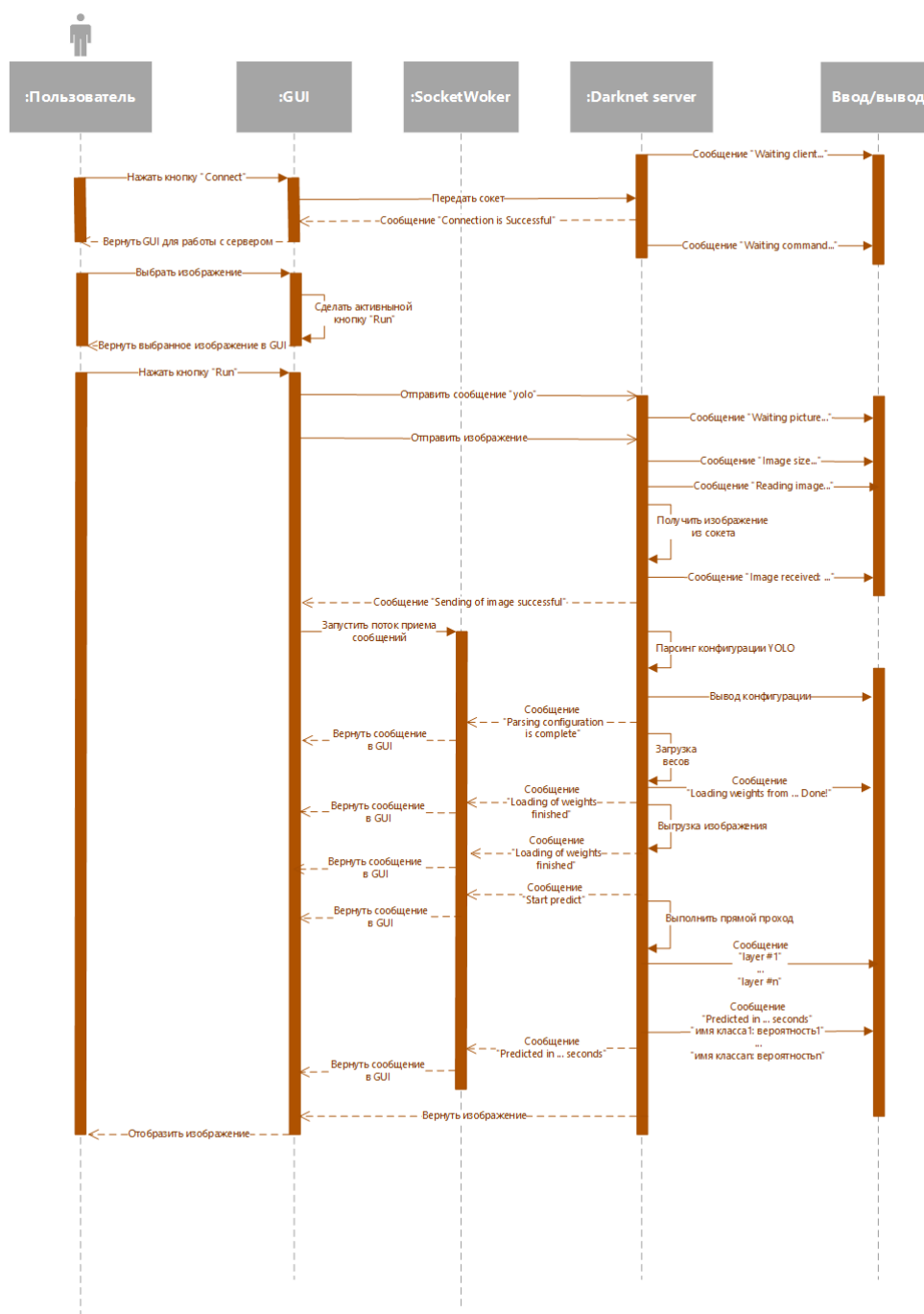
[net] _____ инициализация сети
параметр1=значение1
...
параметрN=значениеN _____ описание внутренних параметров

[слой1]
параметр1=значение1
...
параметрN=значениеN
[слойN]
параметр1=значение1
...
параметрN=значениеN _____ описание слоев, и их параметров

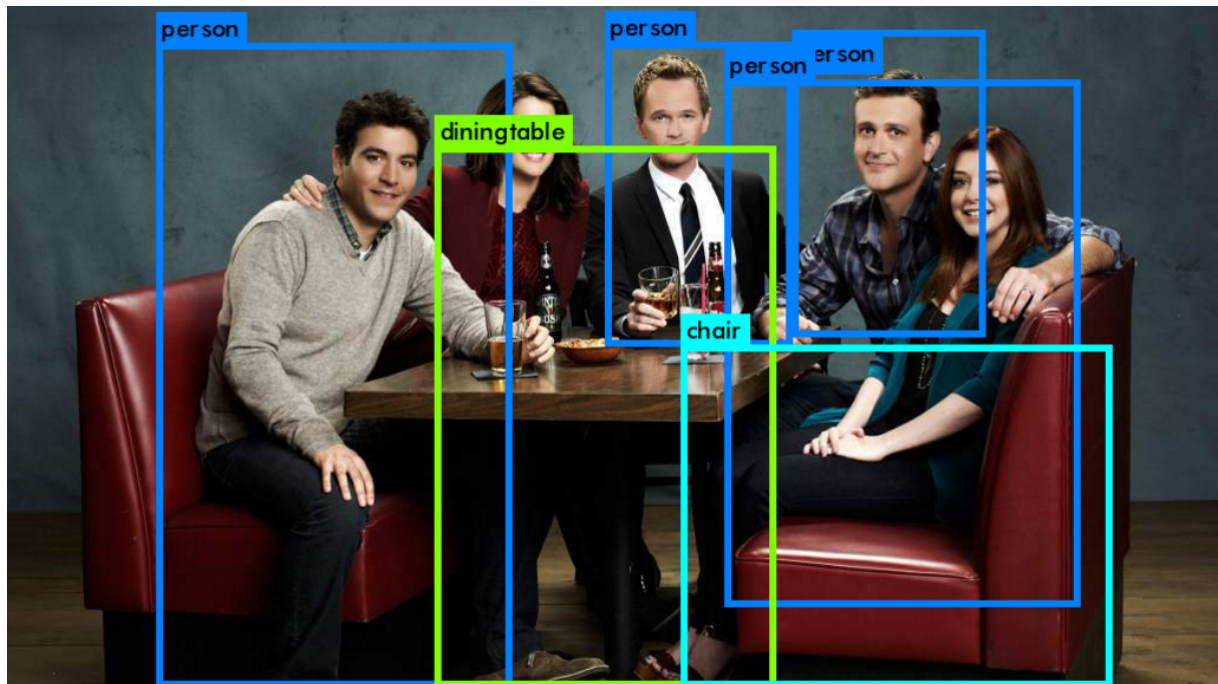
```

Рисунок 1.2 — Конфигурационный файл нейронной сети

# Диаграмма последовательности



## Визуализация локализации объектов



## Макеты экранных форм

