

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 18 страниц и включает 2 рисунка и 2 таблицы.

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«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

ВКР-40 461 806-10.27-10-17.81

Листов 20

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 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	5
2.3 Наименование и условное обозначение темы разработки	6
3 Назначение разработки	6
4 Требование к программе	6
4.1 Требования к функциональным характеристикам	6
4.1.1 Требования к составу выполняемых функций	6
4.2 Организация входных и выходных данных	8
4.2.1 Входные данные	8
4.2.2 Выходные данные	9
4.3 Требования к надежности	10
4.3.1 Требования к надежному функционированию	10
4.3.2 Время восстановления после отказа	10
4.3.3 Требования к составу и параметрам технических средств	11
4.4 Требования к информационной и программной совместимости	12
4.4.1 Требования к методам решения	12
4.4.2 Требования к языкам программирования	12
5 Требования к программной документации	12
6 Стадии и этапы разработки	12
6.1 Стадии разработки	12
6.2 Этапы разработки	13
7 Порядок контроля и приемки	14
7.1 Виды испытаний	14
7.2 Общие требования к приемке	14
Приложение А Формат конфигурационных файлов нейронной сети	15
Приложение Б Диаграмма последовательности	16
Приложение В Визуализация локализации объектов	17

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:

- обмен сообщениями с клиентом;
- парсинг и сериализация конфигураций нейронной сети;
- детектирование 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 А, рисунок А.2;
- парсинг и сериализация файла конфигурации сверточной нейронной сети YOLO. Файл конфигурации представлен в приложении А, рисунок А.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:

- обмен сообщением с сервером;
- взаимодействие с пользователем.

Для работы с сервером необходимо реализовать следующие функции:

- получение хоста сервера;
- соединение с сервером.

Для взаимодействием с пользователем необходимо реализовать следующие функции:

- соединение с сервером;
- выбор изображения для детектирования объектов на нем;
- отображение процесса работы прямого прохода Tiny YOLO;
- отображение изображения 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 Б 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 прямоугольник. Подробная информация об визуализации выходного изображения представлена в приложении В.

Клиентская часть программы должна иметь текстовый браузер для вывода информации о основных этапах работы серверной части приложения. В приложении Б 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 Б 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 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 C. 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 изображением 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.

Приложение А

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 А.1 — Данные о выборке и метках объектов

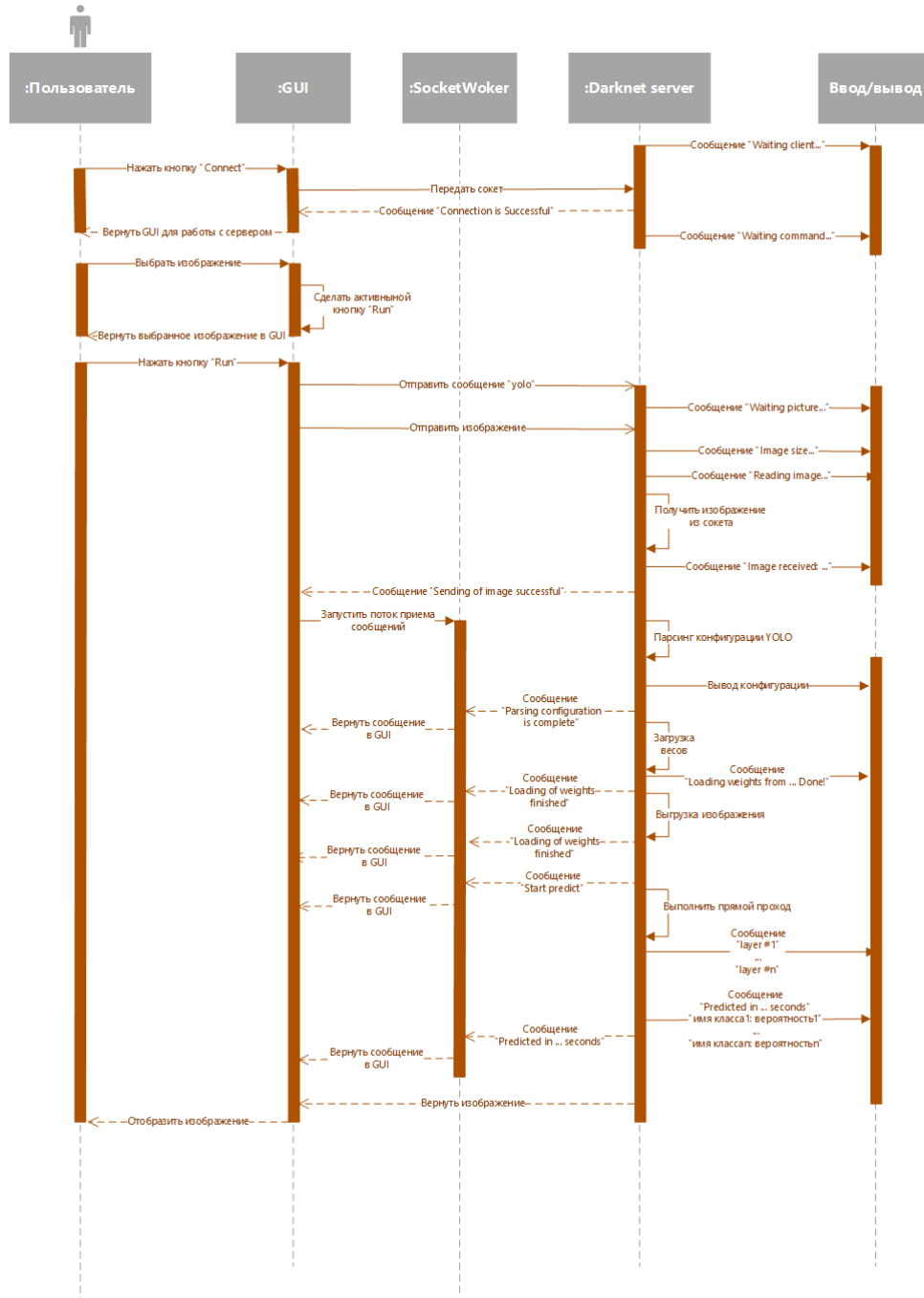
`[net]` — инициализация сети
`параметр1=значение1`
`...`
`параметрN=значениеN` — описание внутренних параметров

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

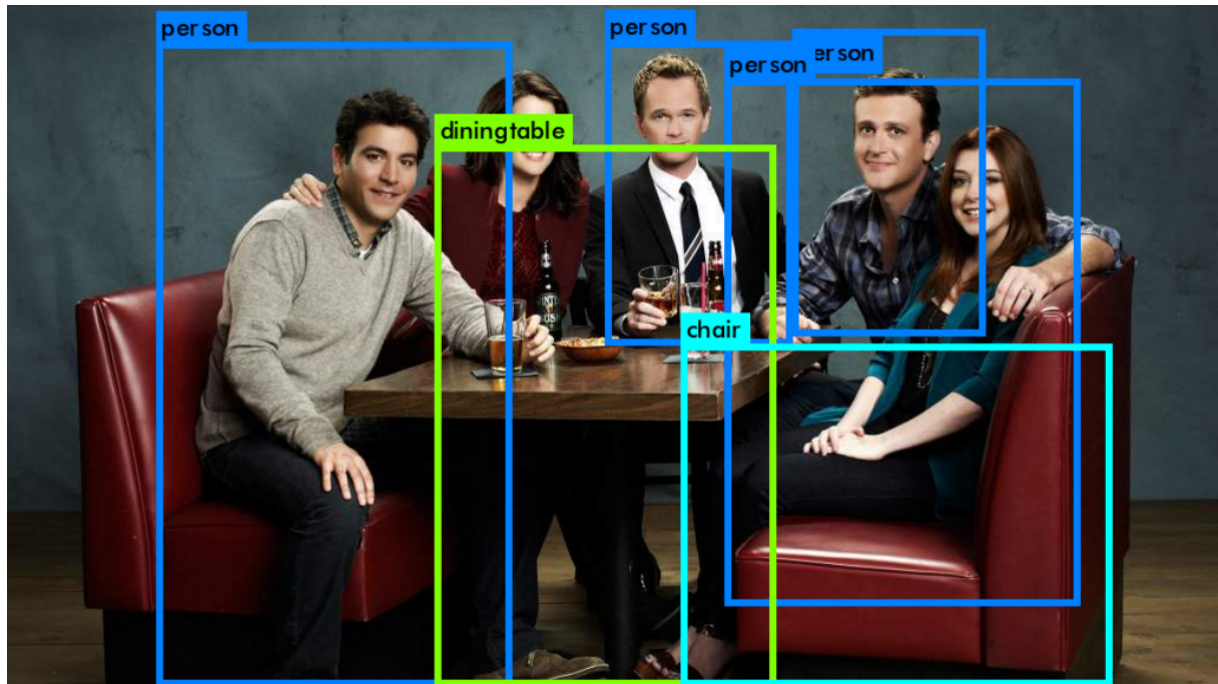
Рисунок А.2 — Конфигурационный файл нейронной сети

Приложение Б

Диаграмма последовательности



Приложение В
Визуализация локализации объектов



Приложение Г

Макеты экранных форм

