

Министерство образования Республики Беларусь

Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.383

Артемьев Владислав Сергеевич

**Оценка позы человека в реальном времени с помощью алгоритмов  
глубокого обучения**

**АВТОРЕФЕРАТ**

на соискание академической степени  
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель

Хотеев А.Л.

кандидат технических наук, доцент

Минск 2019

## КРАТКОЕ ВВЕДЕНИЕ

Мировые технологии развиваются по экспоненциальной кривой. Примеров этому можно привести массу. Например, закон Мура, согласно которому количество транзисторов, размещаемых на кристалле интегральной схемы, удваивается каждые 24 месяца, выполняется и в наше время. Соответственно, размышляя о разработке алгоритмов, необходимо избавиться от линейного мышления и начать думать на 10 шагов вперед.

Автоматизация неизбежна. Уже сейчас понятно, что такие профессии, как грузчик, водитель автомобиля, и другие в ближайшее время станут неактуальными в связи с появлением роботов, оснащенных искусственным интеллектом и беспилотных автомобилей.

Один из алгоритмов, который уже используется в данный момент, и будет активно использоваться в ближайшие 5-10 лет, является алгоритм оценки позы человека.

Одним из примеров уже существующих решений алгоритма является Kinect от компании Microsoft, использующий алгоритм оценки позы человека (Pose estimation) для отслеживания движений человека и использования их для взаимодействия с персонажем в игре. Для работы данного алгоритма используют данные с несколько камер, а также датчик глубины. Однако существуют и ситуации, когда PE необходим не только для игр.

Например, вы захотели потренироваться в спортзале, и вместо того, чтобы нанимать фитнес тренера, вы просто на свой телефон установили приложение, которое оценивает положение вашего тела в режиме реального времени, считает ваши повторения упражнения, а также говорит, если вы выполнили упражнение неправильно. В таком случае не подходит Kinect, т.к. таскать его всегда с собой не представляется возможным. А телефон в наше время есть у каждого человека.

С другой стороны, телефон не обладает таким железом, как Kinect, поэтому задача усложняется. Соответственно, есть несколько требований к алгоритму оценки позы:

1. Он должен быть точным
2. Должен быть быстрым ( в режиме реального времени )
3. Должен работать с выполнением предыдущих условий на телефоне с одной камерой

Целью данной работы является реализация алгоритма, соответствующего всем трем предыдущим условиям.

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

### **Цель и задачи исследования**

*Целью* диссертационной работы является изучение методов применения нейронных сетей, методов глубокого обучения, для решения задачи оценки позы человека в реальном времени, а также для практического применения данного алгоритма в реальной жизни. Для достижения поставленной цели необходимо решить следующие задачи:

1. Ознакомиться с задачей оценки позы человека.
2. Изучить существующие подходы и разработать новый подход, который будет работать быстро, сохраняя достаточно хорошую точность.
3. Разработать алгоритм, используя глубокие нейронные сети, на основе подхода, разработанного в предыдущем пункте.
4. Собрать необходимые данные, обучить модель.
5. Провести экспериментальные исследования разработанного алгоритма и модели нейронной сети.

*Объектом* исследования являются фитнес-приложения.

*Предметом* исследования является применение нейронных сетей для оценки позы человека по фотографии.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность использования глубоких нейронных сетей для задачи оценки позы человека в реальном времени, а также возможно использовать алгоритм с достаточно хорошей точностью в реальном времени на телефоне.

### **Личный вклад соискателя**

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя А.Л.Хотеев, заключается в формулировке целей и задач исследования.

## Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, пяти глав, заключения, списка использованных источников и приложений. В первой главе представлены общие сведения о нейронных сетях, а также рассмотрены основные архитектуры для задачи классификации объектов на изображении. Вторая глава посвящена обзору популярных задач компьютерного зрения, в данный момент решаемых с помощью алгоритмов машинного обучения. В третьей главе происходит обзор и описание фреймворков, используемых для решения задачи диссертации. В четвертой главе произведён обзор различных техник для параллельного выполнения процедур в задачах машинного обучения. В пятой главе описан метод и реализация нейронной сети, работающей в режиме реального времени на телефоне для задачи оценки позы человека.

Общий объем работы составляет 76 страниц, из которых основного текста 61 страница, 26 рисунков, использованных источников из 23 наименований на 2 страницах и 2 приложения на 14 страницах.

## ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы, кратко затронута тема аналогов.

В **первой главе** проведен обзор существующих достижений в машинном обучении, а также выбрана оптимальная стратегия для решение задачи диссертации в реальном времени с хорошим качеством.

Во **второй главе** происходит небольшой экскурс в задачи компьютерного зрения, а также установлены взаимосвязи между похожими задачами.

В **третьей главе** рассматриваются основные технологии и библиотеки, которые используются в данной работе. В качестве фреймворков для машинного обучения и глубокого обучения использовались Tensorflow и Keras, которые являются наиболее популярными инструментами среди разработчиков и исследователей.

**Четвертая глава** рассказывает про основные подходы для распараллеливания вычислений в задачах машинного обучения. Была выбрана

оптимальная концепция и оптимальный фреймворк для того, чтобы обученная нейронная сеть работала в режиме реального времени на телефоне с iOS.

В пятой главе рассмотрена оптимальная по соотношению скорости / точности нейронная сеть для оценки позы человека в реальном времени, обучение ее на датасете COCO, а также реализация прямого прохода сети на телефоне с операционной системой iOS.

## **ЗАКЛЮЧЕНИЕ**

### **Основные научные результаты диссертации**

В рамках данной работы была предложена архитектура математической модели на базе сверточной нейронной сети типа MobileNet, позволяющая в режиме реального времени находить ключевые точки на теле человека (вместе с постобработкой), при этом сохраняя достаточно хорошую точность, если сравнивать с более тяжеловесными моделями. В результате получили обученную модель на датасете COCO. Также разработан фреймворк для проигрывания слоев, участвующих в нейронной сети MobileNet на iOS, являющий оберткой над Metal Performance Shaders.

### **Рекомендации по практическому использованию результатов**

Полученные результаты формируют теоретическую и практическую базу для разработки ПО для оценки позы человека в реальном времени на телефоне с операционной системой iOS. Данный алгоритм может использоваться как основная технология таких приложений, как виртуальный фитнес-тренер, йога-тренер, тренер по танцам и т.д.

Были подтверждены гипотезы про использование сети MobileNet, которая решает задачу классификации, для более сложных задач, например оценки позы человека, а также возможность использовать сверточную нейронную сеть на телефоне в режиме реального времени.