

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

УДК 004.428

Стасюк
Александр Васильевич

**ПРОГРАММНОЕ СРЕДСТВО АНАЛИЗА ИЗОБРАЖЕНИЙ НА
ОСНОВЕ НЕЙРОННОЙ СЕТИ**

АВТОРЕФЕРАТ

диссертации на соискание степени
магистра информатики и вычислительной техники

по специальности 1-40 81 01 – Информатика и технологии разработки
программного обеспечения

Минск 2020

Работа выполнена на кафедре информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель: **СКОБЦОВ Вадим Юрьевич**,
кандидат технических наук, доцент кафедры программного обеспечения информационных технологий «Белорусский государственный университет информатики и радиоэлектроники»

Рецензент: **БУЛОВА Александр Дмитриевич**,
кандидат технических наук, доцент кафедры экономической информатики «Белорусский государственный экономический университет»

Защита диссертации состоится «24» июня 2020 г. года в 10⁰⁰ часов на заседании Государственной экзаменационной комиссии по защите магистерских диссертаций в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, Минск, ул. Гикало, 9, копр. 4, ауд. 111, тел. 293-85-91, e-mail: inform@bsuir.by

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

ВВЕДЕНИЕ

Оценка позы человеческого тела по изображениям является сложной задачей, которая вызвала большой интерес со стороны сообщества специалистов, занимающихся компьютерным зрением. Автоматическое определение позы тела имеет много сфер применения, такие как отслеживание человека, захват движения, распознавание активности, наблюдение и анализ рабочего процесса хирургов во время операций.

Компьютерное зрение активно способствует эволюции современного общества, продвигая автоматизацию процессов в различных областях. Хорошо изученным предметом в области компьютерного зрения являются люди. Обнаружение человека, отслеживание и оценка позы по изображениям являются фундаментальными проблемами высокого уровня, которые сообщество неоднократно решало в прошлом. При обнаружении и отслеживании люди обычно локализуются с помощью ограничивающей рамки в пределах одного или нескольких видов обзора. При оценке поз цель состоит в том, чтобы восстановить положение человеческого тела в 2-мерном или 3-мерном пространстве из данных изображения. Положение человеческого тела обычно описывается скелетом, а данные изображения поступают из одного или нескольких видов. Несмотря на то, что были предложены надежные алгоритмы обнаружения и отслеживания человека в реальных условиях, оценка поз все еще остается открытой проблемой. В этой работе будет рассмотрена проблема оценки позы человека в сложных условиях.

Основная работа будет проводиться в области машинного анализа, исследуя достижения глубокого обучения в компьютерном зрении. Глубокое обучение - это новая область машинного обучения, которая моделирует абстракции высокого уровня в модальностях данных с использованием глубоких архитектур. Основная идея заключается в изучении представлений из необработанных данных с использованием сложных и нелинейных преобразований. Наиболее популярными архитектурами для реализации глубокого обучения являются сети глубокого убеждения (DBN) и сверточные глубокие нейронные сети (ConvNets) со многими приложениями в области компьютерного зрения, распознавания речи и обработки естественного языка (NLP). Обе модели являются типом нейронной сети. Хотя они были представлены несколько десятилетий назад, в последние годы они приобрели большую популярность, главным образом, благодаря достижениям аппаратных систем.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Магистерская диссертация относится к актуальной области современных информационных технологий – использованию нейронных сетей для решения задачи распознавания позы человека на изображении. Актуальность работы обусловлена тем, что в настоящее время всё более широкое распространение получают биометрические системы анализа движений человека и именно использование нейронных сетей для решения данной задачи является наиболее перспективным направлением, на что и направлено основное внимание данной работы.

Результаты работы носят как теоретический (комплекс рекомендаций, методик и т.п.), так и практический характер: предложенная композиция из нейронных сетей может применяться в системах анализа движений, системах детекции с идентификацией личности, системах анализа видеоконтента за счет надежности, а также высокой скорости идентификации.

Степень разработанности проблемы

Исследование задачи распознавания позы человека по изображению с использованием нейронных сетей осуществлялось на основе работ российских и белорусских ученых: А. Глазунова, Д. Самаля, Г. Вороновского, А. Скурихина, В. Базаревского, а также зарубежных авторов: Ш. Панканти, Э. Джейна, А. Ранга, П. Пернера, М. Зейлера и др.

Одним из недостатков исследований, представленных в современной технической литературе, является рассмотрение задачи нахождения ключевых точек, а не определения позы.

Предложенная работа направлена на устранение этого недостатка, в работе целостно рассмотрена задача распознавания позы на изображении, а также проведены исследования различных архитектур нейронных сетей для достижения необходимой точности и производительности при реализации системы распознавания позы человека на изображении.

Цель и задачи исследования

Целью диссертационной работы является изучение алгоритмов обучения и архитектур нейронных сетей для распознавания позы человека на изображениях, а также разработка программного обеспечения для решения задач определения позы и дальнейшего распознавания в режиме реального времени.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Исследовать современные архитектуры нейронных сетей;
2. На основании произведенного анализа выбрать необходимые технологии и инструменты для реализации поставленной задачи;

3. Спроектировать и разработать программный продукт для распознавания позы человека на изображениях и видео;

4. Провести экспериментальные исследования разработанного программного продукта.

Область исследования

Содержание диссертации соответствует образовательному стандарту высшего образования второй ступени (магистратуры) ОСВО 1-39 81 01-2012 специальности 1-40 81 01 «Информатика и технологии разработки программного обеспечения».

Теоретическая и методологическая основа исследования

В основу диссертации легли работы белорусских и зарубежных ученых в области использования глубоких нейронных сетей для решения задач компьютерного зрения.

Информационная база исследования сформирована на основе литературы, открытой информации, технических нормативно-правовых актов, сведений из электронных ресурсов, а также материалов научных конференций и семинаров.

Научная новизна

Научная новизна и значимость полученных результатов работы заключается в исследовании архитектур глубоких нейронных сетей для решения задачи распознавания позы человека, в предложенной композиции нейронных сетей для решения данной задачи, а также разработке мобильного приложения, использующего композицию из PoseNet и MobileNet сетей со специальными модификациями.

Теоретическая значимость работы заключается в детальном анализе архитектур и подходов к использованию нейронных сетей для распознавания поз.

Практическая значимость диссертации состоит в проведенных исследованиях различных архитектур, также предложенная композиция из нейронных сетей может применяться в системах биометрической идентификации, системах анализа движений, системах анализа видеоконтента за счет надежности, а также высокой скорости идентификации.

Основные положения, выносимые на защиту

1. Систематизация основных классов решаемых задач в распознавании позы человека на изображении, преимущества нейросетевых методов распознавания позы человека на изображении.

2. Архитектуры сверточных нейронных сетей для решения задачи распознавания позы. Анализ их точности и производительности работы.

3. Обзор CoreML для конвертации моделей сверточных нейронных сетей в формат, который может быть использован на мобильных устройствах, а также разработка экспериментального мобильного приложения, которое использует композицию из PoseNet и MobileNet сетей со специальными модификациями.

Апробация диссертации и информация об использовании ее результатов

Основные положения диссертационной работы использовались в разработке системы анализа правильности исполнения физических упражнений.

Публикации

Результаты исследований, вошедшие в диссертацию, были опубликованы в статье в рецензируемом издании: Дацик А.А., Стасюк А.В. Алгоритмы обнаружения объектов. Образование и наука в России и за рубежом выпуск N15(2019) Vol.63.

Структура и объем работы

Диссертация состоит из введения, общей характеристики работы, четырех глав с краткими выводами по каждой главе, заключения, библиографического списка и приложений.

В первой главе приведен обзор основных классов решаемых задач в распознавании позы человека на изображении, рассмотрены нейросетевые методы распознавания человека на изображении, а также рассмотрены различные типы искусственных нейронных сетей.

Во второй главе рассмотрены архитектуры сверточных нейронных сетей для решения задачи распознавания позы, также проведены эксперименты для сравнения точности и производительности разных архитектур.

В третьей главе описан процесс обучения нейронной сети с разработанной ранее архитектурой сети. Приведены результаты экспериментов и метрик. Приведен обзор набора данных и аргументация выбранных целевых метрик.

В четвертой главе приведен обзор средств для использования сверточных сетей на мобильных устройствах. А также приведены результаты разработки экспериментального мобильного приложения, работающего с помощью PoseNet и MobileNet сетей со специальными модификациями.

В приложении представлены публикации автора и результаты исследований производительности и точности сетей.

Общий объем работы составляет 73 страницы, из которых основного текста – 55 страниц, 18 рисунков на 14 страницах, 5 таблиц на 3 страницах,

список использованных источников из 22 наименований на 2 страницах и 3 приложений на 3 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** описана проблематика сферы. Задача оценки позы человеческого тела по изображениям является сложной задачей, которая вызвала большой интерес со стороны сообщества специалистов, занимающихся компьютерным зрением. Автоматическое определение позы тела имеет много сфер применения, такие как отслеживание человека, захват движения, распознавание активности, наблюдение и анализ рабочего процесса хирургов во время операций, анализ видеопотока в контексте решения задач правопорядка, в спорте. В работе рассмотрена задача оценки позы человека в сложных условиях. Для решения поставленной задачи выбрана одна из основных технологий компьютерного зрения и машинного обучения - технология глубокого обучения и модели сверточных нейронных сетей.

В **общей характеристике работы** показана актуальность проводимых исследований, степень разработанности проблемы, сформулированы цель и задачи диссертации, обозначена область исследований, научная (теоретическая и практическая) значимость исследований, а также апробация работы.

В **первой главе** приведен обзор задач распознавания позы человека на изображении.

Задача оценки позы человеческого тела по изображениям является сложной задачей, которая вызвала большой интерес со стороны сообщества специалистов, занимающихся компьютерным зрением. Автоматическое определение позы тела имеет много сфер применения, такие как отслеживание человека, захват движения, распознавание активности, наблюдение и анализ рабочего процесса хирургов во время операций, анализ видеопотока в контексте решения задач правопорядка, в спорте. В работе рассмотрена задача оценки позы человека в сложных условиях. Для решения поставленной задачи выбрана одна из основных технологий компьютерного зрения и машинного обучения - технология глубокого обучения и модели сверточных нейронных сетей. принимается за объект, относящийся к известным системе классам, и ему разрешается доступ. При распознавании требуется отнести объект к одному из n известных классов или выдать заключение о том, что этот объект не относится к известным классам.

Существует несколько способов задания позы человека на изображении. В первых работах под позой понимали положение множества частей тела человека, таких как голень, предплечье, туловище, голова и др. Положение каждой части тела описывалось прямоугольником, стороны которого могли не быть параллельны осям координат. Решение задачи в такой постановке оказалось сложным поскольку размеры частей тела на

изображении могут сильно изменяться в зависимости от позы человека. Поэтому в одной из работ авторы определяют позу как положение суставов, соединяющих части тела человека, на изображении. В настоящее время эта постановка используется для определения позы.

Во второй главе рассмотрена архитектура нейронной сети, разработанная в ходе исследований.

В настоящее время, как это было установлено и описано в разделе литературного обзора, для решения большинства прикладных задач анализа на изображениях и видеоданных в качестве подходов к формированию архитектуры сверточных нейросетевых классификаторов патологий используются следующие решения:

- перебор случайных конфигураций сети: установка случайных параметров архитектуры и проверка точности работы конфигурации каждой модели на определенном наборе данных. Данный способ является самым неэффективным и реже всего может использоваться для решения прикладных задач;

- экспертный подход: установка параметров сети как для решения похожей успешно решенной задачи. Данным подходом пользуются большинство исследователей, использующих в качестве модели классификации СНС;

- автоматизированный подбор параметров с оптимизацией: подбор значений параметров сети с использованием алгоритмов оптимизации, например, байесовская оптимизация. Данный подход используется достаточно редко и зачастую может являться неэффективным из-за больших затрат времени на его реализацию;

- построение индивидуальной архитектуры СНС, оптимизированной для решения определенного класса задач. В большинстве случаев данный подход используется в исследованиях, касающихся непосредственно машинного обучения и разработки оптимальных моделей СНС.

MobileNetV2. Особенностью данной архитектуры является отсутствие max-pooling - слоёв. Вместо них для снижения пространственной размерности используется свёртка с параметром stride, равным 2.

Двумя гиперпараметрами архитектуры MobileNetV2 являются α (множитель ширины) и ρ (множитель глубины или множитель разрешения). Множитель ширины отвечает за количество каналов в каждом слое. Например, $\alpha = 1$ даёт нам архитектуру, описанную в статье, а $\alpha = 0.25$ — архитектуру с уменьшенным в четыре раза числом каналов на выходе каждого блока.

В третьей главе был описан процесс обучения нейронной сети с разработанной ранее архитектурой сети. Приведены результаты экспериментов и исследований. Проведен сравнительный анализ с архитектурами сетей с существенно большим количеством параметров. Приведен обзор набора данных и аргументация выбранных целевых метрик.

Показатели метрик у нейронной сети с основным блоком в виде архитектуры EfficientNet-B1 являются на порядок выше, чем у MobileNetV2. Но время исполнения (обработки одного кадра) является неприемлемым для использования в мобильных приложениях.

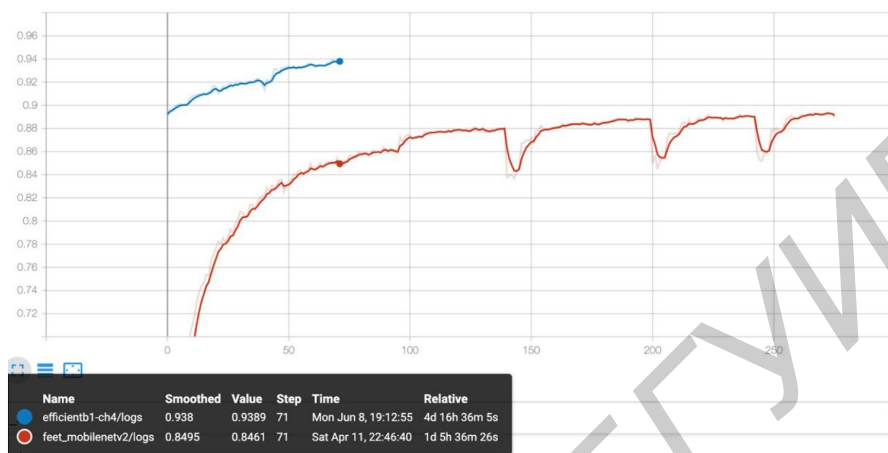


Рисунок 1 – Графики значений метрик нейронных сетей с основными блоками MobileNetV2 и EfficientNet-B1

Выбранная архитектура сети содержит 3.4 миллиона параметров и имеет размер входного изображения 352 пикселя на 352 пикселя. Также, для обеспечения хорошей сходимости нейронной сети, должен быть использован размер мини-пакета не менее 16. Данные параметры довольно жестко определяют требования к вычислительным мощностям.

Параметр	Значение
Архитектура центрального процессора	Intel Haswell Xeon
Количество ядер центрального процессора	32
Объем оперативной памяти (5-е поколение)	64 Гб
Архитектура графического процессора	NVIDIA Tesla P100
Количество ядер графического процессора	3584
Количество памяти графического процессора	16 Гб

В таблице выше приведены основные характеристики вычислительной машины, на которой проводилось обучение нейронной сети. Из-за использования графического процессора с меньшим количеством памяти пришлось бы значительно уменьшать размер пакета при обучении. Уменьшение размера пакета привело бы ухудшению сходимости.

В четвертой главе было разработано экспериментальное мобильное приложения под ОС IOS. Интерес представляет, как эти сети будут работать на не флагманских мобильных устройствах, а также с какой скоростью приложение будет распознавать позы людей.

Также в данной главе был изучен CoreML как инструмент, который позволяет исполнение модели на различных устройствах, в том числе мобильных телефонов и интернета вещей, что дает вам более чем 3- кратное ускорение вывода по сравнению с оригинальным Pytorch-моделями. С помощью него можно получить машинное обучение на своем Raspberry Pi или на телефоне.

Связка из CoreML, графического ускорителя iPhone и разработанной архитектуры для определения позы показала достаточно хорошие результаты в плане производительности, а также точности.

Алгоритм на языке Swift работает в 2,6 раза быстрее, чем такой же алгоритм на Objective-C, и в 8,4 раза быстрее, чем на Python. Использование Swift делает процесс разработки программ более гибким и удобным.

К сожалению в Swift и Objective-C нет возможности найти максимум в матрице. Данный функционал приходится реализовывать вручную, циклическими проходами по элементам матриц.

Name	Type	Flexibility	Description
▼ Inputs			
input_0	Image (Color 208 x 352)	208 x 352 352 x 208 ...	
▼ Outputs			
heatmaps	MultiArray (Float32 14 x 22 x 13)	14 x 22 x 13 14 x 13 x...	MultiArray of shape (1, 1, 14, 22, 13).
short_offsets	MultiArray (Float32 28 x 22 x 13)	28 x 22 x 13 28 x 13 x...	MultiArray of shape (1, 1, 28, 22, 13).
472	MultiArray (Float32)		MultiArray of shape (1, 1, 52, 22, 13).

Рисунок 2 – Формат выхода разработанной нейронной сети в CoreML

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Дан обзор нейросетевых методов распознавания позы человека на изображении. Указаны преимущества и перспективы нейросетевых методов. Проведены исследования различных алгоритмов обучения, предобработки изображений, архитектур нейронных сетей, которые могут использоваться для решения задачи распознавания позы человека на изображении. Проведен сравнительный анализ производительности и точности распознавания нескольких современных архитектур нейронных сетей. Отмечены наиболее перспективные архитектуры для данной задачи.

2. Предложена композиция из нейронных сетей, которая способна работать на различных устройствах, в том числе мобильных телефонах и интернета вещей. Данный способ обработки изображений позволяет получить достаточно большой прирост в производительности даже на относительно слабых устройствах. На флагманских моделях мобильных телефонах процедура распознавания позы человека занимает несколько миллисекунд и отлично работает в режиме реального времени.

3. Разработан экспериментальный программный продукт под ОС IOS с использованием композиции из оптимизированных моделей нейронных сетей.

Рекомендации по практическому использованию результатов

1. Полученные результаты формируют теоретическую и практическую базу для разработки ПО компьютерных систем для решения задач идентификации движений человека.

2. Предложенная композиция из нейронных сетей может применяться в системах биометрической идентификации, системах детекции с идентификацией личности, системах анализа видеоконтента за счет надежности, а также высокой скорости идентификации.

3. Результаты работы могут использоваться при подготовке персонала для разработки и обслуживания компьютерных систем, решающих задачи детекции и(или) распознавания человека на изображении или видео, а также анализа движений.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

Статьи в рецензируемых журналах

1. Дацик А.А., Стасюк А.В. Алгоритмы обнаружения объектов. Образование и наука в России и за рубежом выпуск N15(2019) Vol.63.