

## ОТЗЫВ

официального оппонента Спектора Александра Аншелевича на диссертацию Свитова Давида Вячеславовича «Оптимизация производительности свёрточных нейронных сетей в системе распознавания лиц», представленную на соискание учёной степени кандидата технических наук по специальности 1.2.2 - «Математическое моделирование, численные методы и комплексы программ»

### **1. Актуальность темы исследования и соответствие требованиям Положения ВАК РФ по специальности**

К настоящему моменту нейросетевые модели являются эффективным инструментом для решения задач автоматизации. Однако их применимость ограничена высокими требованиями к вычислительным ресурсам. Это приводит к удорожанию использования нейросетевых систем и повышению сложности их внедрения из-за необходимости поддержки вычислительных серверов.

Востребованной и актуальной задачей автоматизации с использованием нейронных сетей являются технологии распознавания лиц. В диссертационной работе Свитова Д.В. рассмотрена задача оптимизации производительности нейронных сетей в системе биометрической идентификации по лицу. В работе предложены и исследованы новые подходы к улучшению соотношения качества и скорости нейронных сетей в ключевых узлах такой системы. Это позволяет повысить практическую применимость нейронных сетей для биоидентификации, что обеспечивает актуальность данной работы.

Тема диссертационного исследования соответствует следующим пунктам паспорта специальности 1.2.2 - «Математическое моделирование, численные методы и комплексы программ»:

Пункт 2. «Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий».

Пункт 3. «Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента».

Пункт 6. «Разработка систем компьютерного и имитационного моделирования, алгоритмов и методов имитационного моделирования на основе анализа математических моделей».

Пункт 9. «Постановка и проведение численных экспериментов, статистический анализ их результатов, в том числе с применением современных компьютерных технологий».

## 2. Структура диссертации

Диссертация состоит из введения, четырёх глав, заключения и списка литературы из 106 источников. Общий объем работы составляет 109 страниц машинописного текста, иллюстрированного 23 рисунками и 18 таблицами.

Во *введении* показана актуальность темы, сформулированы цель и задачи диссертационного исследования и представлены основные положения, выносимые на защиту.

В *первой главе* приведён обзор методов оптимизации производительности нейросетевых моделей и проанализирована их применимость к системе распознавания лиц.

Во *второй главе* описан предложенный автором новый алгоритм снижения числа ложноположительных срабатываний детектора лиц. Суть предложенного подхода заключается в новом устойчивом к зашумлению сигнала алгоритме обнаружения движения в видеопотоке. Алгоритм, основанный на использовании промежуточных карт признаков предобученного детектора, позволяет повышать точность некоторой имеющейся системы. В качестве дополнительного преимущества данного подхода выделяется возможность ранней остановки вычисления нейронной сети детектора. Остановка производится на основе анализа карт признаков промежуточного слоя активации. В конце главы приведены описание и результаты экспериментов, подтверждающих эффективность предложенного подхода.

В *третьей главе* описан предложенный автором новый подход к дистилляции моделей для построения биометрических векторов. Описанный метод позволяет повысить точность нейронной сети с малым числом обучаемых параметров за счёт передачи информации от нейронной сети с большим числом параметров. В работе предложен подход, при котором для инициализации сети-ученика используются веса последнего слоя сети-учителя. Данный способ инициализации весов открывает возможность к разработке новых методов дистилляции. Так в представленной работе предложен подход к дистилляции на основе анализа угла между векторами признаков и центром кластера соответствующего класса в сетях учителя и ученика.

В ряде экспериментов осуществлено сравнение предложенного подхода с существующими методами дистилляции на открытых наборах данных и



продемонстрирована его эффективность. Также в главе приведён анализ различных элементов подхода и их влияние на итоговый результат.

В *четвёртой главе* описана реализация системы распознавания лиц в виде программного модуля и соответствующих нейронных сетей. Приведено описание архитектур используемых сетей и методов их обучения. Обоснован выбор программных библиотек для их обучения и описаны методы их конвертации для использования на целевом устройстве (как правило, умном домофоне). В главе приведены результаты тестирования разработанной системы на реальных данных. Описаны используемые метрики, способы сбора данных, постановка и результаты экспериментов. Также в главе приводятся и анализируются результаты тестирования системы независимым институтом стандартов.

В *заключении* приведены основные результаты, полученные в диссертационной работе.

### **3. Основные научные результаты, их новизна и ценность для науки и практики**

В диссертационной работе Свитова Д.В. предложен и реализован новый метод обнаружения движения в видео потоке на основе анализа промежуточных карт признаков. Подход позволяет производить устойчивое к зашумлению входных изображений обнаружение движения без существенного увеличения времени работы системы. Данный метод удобен к применению в существующем и внедренном детекторе, что повышает его практическую применимость.

Разработан и реализован новый метод инициализации сети-ученика в задаче дистилляции нейронных сетей, обученных с применением функции Софтмакс с отступом. Инициализация весов копированием параметров последнего слоя из сети-учителя позволяет улучшить разделимость классов в пространство векторов при обучении. Также предложенный метод позволяет расширить пространство возможных каналов передачи информации между нейронными сетями при дистилляции. Это открывает возможность развития новых подходов к дистилляции знаний.

В работе предложен и реализован новый метод дистилляции знаний для задачи биометрической идентификации сетей, обученных с функцией Софтамакс с отступом. Предложенный метод основывается на описанном ранее подходе к инициализации весов сети-ученика и осуществляет передачу информации путём добавления функции ошибки на основе анализа углов между векторами в сети-учителе. При этом в отличие от подходов на основе триплетов, используется новый метод на основе оценки угла между вектором и

центром класса. Данный метод в различных экспериментах показал эффективность по сравнению с существующими подходами.

Предложенные подходы объединены в программной реализации системы распознавания лиц для встраиваемых устройств. В работе приведено описание реализации системы и результаты её тестирования на реальных данных. Это позволяет судить о высокой практической применимости разработанных подходов.

Предлагаемые в работе Свитова Д.В. подходы являются новыми и реализованы в виде программных модулей, имеющих высокую практическую ценность, что подтверждается актом о внедрении разработанной системы.

#### **4. Достоверность и обоснованность выводов и результатов диссертации**

Автором проведены тесты на открытых наборах данных, что позволило провести прямое сравнение результатов с существующими подходами. В работе продемонстрированы результаты работы на лабораторных и реальных данных (собранных в неконтролируемой среде). Полученные результаты не противоречат друг другу и теоретическим выкладкам. Дополнительную достоверность полученных результатов осуществляет проведённое независимым институтом стандартов тестирование разработанной системы.

#### **5. Основные замечания по диссертации**

5.1. Обнаружение лица для предъявления к классификации опирается на признак его перемещения в последовательности кадров. Исключается принятие решения относительно объектов неподвижных. Всегда ли это оправдано? Как это соответствует, например, задаче распознавания лица в домофоне?

5.2. В работе неоднократно подчеркивается «устойчивость к шуму» предложенных алгоритмов. Например, алгоритма обнаружения объекта. Говорится о вероятности ложного обнаружения. Существующие в локации (радио-, гидро-, сейсмо-) методы ориентированы на применение критерия Неймана-Пирсона, предполагающего жесткую стабилизацию вероятности ложного обнаружения (ложной тревоги) и, при этом, максимизацию вероятности правильного обнаружения объекта наблюдения. Жесткое отношение к вероятности ложного обнаружения способствует стабилизации требований к вычислительным ресурсам, что, как для локационных систем, важно и для обсуждаемых в диссертации.

5.3. При использовании дистилляции для передачи данных сети-учителя в сеть-ученика, имеющую значительно меньшую ресурсную базу, применяется ранее



предложенная функция softmax, 
$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^N e^{z_k}}$$
, которой придается смысл распределения вероятностей. Далее говорится, что «*после такого преобразования все элементы вектора  $z$  имеют значения в интервале  $[0,1]$  и их сумма равна 1. Это позволяет обучать значения весов модели, используя вероятностные методы, такие как метод максимального правдоподобия (МП)*». На самом деле, свойством неотрицательности и нормировки к единице обладает не исходный вектор  $z$ , а вектор преобразованных значений  $\sigma(z)$ .

Придание этой функции вероятностного смысла было бы полезно обосновать хотя бы качественно. Почему она имеет монотонный характер (чем больше  $z_i$ , тем больше вероятность)? Не ясно, каким образом на этой основе можно использовать метод МП в решаемой задаче. Тем более что в дальнейшем вместо нормированной функции  $\sigma(z)_i$  используется ее логарифм, ликвидирующий свойство нормировки...

5.4. Рисунок 1.7 показывает различие распределений параметров не квантованной и квантованной нейронной сети. Из их сравнения видно, что квантование параметров приводит к выпадению целых областей в квантованном пространстве параметров. Разве это не прунинг, рассматриваемый как отдельный способ упрощения технологии?

5.5. Применение функции softmax с отступами направлено на увеличение расстояний между классами наблюдаемых объектов. Было бы полезно воспользоваться достижениями существующих методов статистической классификации сигналов, эффективность которых определяется степенью ортогональности сигналов.

5.6. В диссертации большое (и оправданное) внимание уделяется описанию известных типов нейронных сетей. Было бы полезно наглядное представление их характерных архитектурных особенностей. Тем более что одним из подходов к оптимизации, развиваемых в работе, является архитектурный поиск.

## **6. Публикация и апробация результатов диссертации**

Материалы диссертации опубликованы в 5 печатных работах в рецензируемых журналах индексируемых Scopus и ВАК РФ, в том числе относящихся к первой и второй квартилям. Основные положения и результаты диссертационной работы докладывались на конференциях в России и Европе.

## 7. Содержание автореферата

Автореферат диссертации достаточно полно отражает основные идеи, результаты, выводы и положения диссертационной работы.

## 8. Оценка диссертации в целом

Диссертационная работа Свитова Д.В. «Оптимизация производительности свёрточных нейронных сетей в системе распознавания лиц» представляет собой законченную научно-квалификационную работу. По уровню проведённых исследований, актуальности, новизне и практической ценности полученных результатов работа удовлетворяет требованиям пункта 9 «Положения о присуждении учёных степеней» ВАК при Минобрнауки России, а её автор Свитов Давид Вячеславович заслуживает присуждения учёной степени кандидата технических наук по специальности 1.2.2 - «Математическое моделирование, численные методы и комплексы программ».

Спектор Александр Аншелевич,  
доктор технических наук, профессор,  
профессор кафедры теоретических основ радиотехники Федерального  
государственного бюджетного образовательного учреждения высшего  
образования «Новосибирский государственный технический  
университет»

Россия, 630073, г. Новосибирск, пр-т К. Маркса, 20

Телефон: +7-953-769-1272

Эл. почта: spectora@mail.ru

Подпись \_\_\_\_\_ А.А. Спектор

11.05.2023

Подпись Спектора А.А. заверяю

