

На правах рукописи



Вольнова Яна Сергеевна

**Метод обучения классификаторов изображений с учетом
отношений между классами**

2.3.5 Математическое и программное обеспечение вычислительных систем,
комплексов и компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой степени кандидата технических наук

Москва – 2026

Работа выполнена на кафедре «Компьютерные системы и сети» Федерального государственного автономного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана).

Научный руководитель: **ИВАНОВА Галина Сергеевна**
доктор технических наук, профессор, профессор
кафедры «Компьютерные системы и сети»

Официальные оппоненты: **МАТВЕЕВ Иван Алексеевич**
доктор технических наук, главный научный
сотрудник Отделения 1 «Математические методы
анализа данных и прогнозирования» Федерального
государственного учреждения «Федеральный
исследовательский центр «Информатика и
управление» Российской академии наук»
МОСЕВА Марина Сергеевна
кандидат технических наук, доцент кафедры
«Программная инженерия» Ордена Трудового
Красного Знамени федерального государственного
бюджетного образовательного учреждения
высшего образования «Московский технический
университет связи и информатики»

Ведущая организация: Общество с ограниченной ответственностью
«РТСофт-Смарт Грид» (г. Москва)

Защита состоится «1» октября 2026 года в 15:00 на заседании диссертационного совета 24.2.331.19 в МГТУ им. Н.Э. Баумана по адресу: 105005, г. Москва, ул. 2-я Бауманская, д. 5, стр. 1, зал Учёного совета ГУК МГТУ им. Н.Э. Баумана.

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана и на сайте bmstu.ru.

Отзывы на автореферат, заверенные печатью учреждения, просьба направлять по адресу: 105005, г. Москва, ул. 2-я Бауманская, д.5, стр.1, Учёному секретарю диссертационного совета 24.2.331.19, кафедра ИУЗ.

Автореферат разослан «__» _____ 2026 года.

Ученый секретарь
диссертационного совета 24.2.331.19
кандидат технических наук, доцент



С.А. Сакулин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Задача классификации – одна из основных в компьютерном зрении. Под классификацией понимается соотнесение изображения объекта с одним из заранее заданных классов в зависимости от представленного на нем визуального образа. Современные классификаторы, как правило, создаются с помощью машинного обучения с учителем, в качестве модели обычно используют сверточные нейронные сети или трансформеры зрения, так как они позволяют получить высокую точность распознавания и применимы к широкому кругу предметных областей. Несмотря на распространение больших языковых моделей, потребность в локальных узкоспециализированных классификаторах сохраняется в сферах, где требуются конфиденциальность, скорость обработки, автономность, надежность, таких как медицина, сельское хозяйство, розничная торговля и других.

В общепринятом подходе к обучению классификаторов изображений классы по умолчанию полагаются в равной степени отличными друг от друга, хотя в реальном мире они часто связаны отношениями – похожи или отличаются по некоторому признаку, составляют иерархию. К примеру, разные растения может объединить общая категория «сорняки», а их виды – подкатегория «многолетние». Эти отношения игнорируются, в результате чего классификаторы допускают грубые ошибки, когда предсказанный моделью и истинный класс совершенно разные по смыслу. К примеру, если классификатор перепутал два растения – это объяснимо, а если он посчитал за растение зеленое насекомое на земле, то это грубая ошибка, которая подрывает доверие к решениям модели. Грубые ошибки случаются, потому что визуальные признаки, на которые опирается модель (цвет, фон), не всегда совпадают с важными для человека (форма). Поэтому использование информации об отношениях между классами в обучении классификаторов может снизить количество грубых ошибок, а также повысить обобщающую способность модели при ограниченном числе обучающих примеров. Кроме того, классификаторы, обученные таким образом, демонстрируют большую прозрачность в принятии решений. В них можно отследить, почему объект был отнесен к тому или иному классу, какие классы модель считает похожими.

Решением задачи классификации изображений в рамках машинного обучения занимались такие ученые как Афанасьев Г.И., Николаев Д.П., Султанов Р.О., Филист С.А., Aziere N., Todorovic S., Chen T., Dozat T., LeCun Y., Li L., Luo C., Liu Z., Mao A., Moayeri M., Mohri M., Vaswani A., Zhong Y. и другие. В результате был сформирован основной подход к созданию классификаторов, использующий в качестве основы обучения функцию потерь под названием кросс-энтропия. Развитию методов обучения классификаторов и альтернативных функций потерь, в частности триплетной функции потерь, посвящены работы Арлазарова В.Л., Зацаринного А.А., Нестерова Ю.Е., Новиковой К.Н., Сулоева К.К., Толстых А.А., Шешкуса А.В., Beyer L., Chopra S., Hadsell R., He K., Hermans A., Kalenichenko D.,

Leibe B., Lin T. Y., Oh Song H., Philbin J., Schroff F., Tian Y., Wang X., Xuan H., Zhang S. и других. Однако в их исследованиях внимание сфокусировано на самих изображениях и дополнительных визуальных признаках, а не на классах. Потенциал использования знаний о классах во время обучения был частично реализован в работах Csányi G. M., Dao S.D., Deng W., Kharbanda S., Sajedi A., Shaik S., Sumbul G., Sun Y., Zakharov S. и других. Однако способы применения отношений между классами, предложенные ими, существенно повышают вычислительную сложность обучения. Кроме того, результаты исследований получены на авторских датасетах, что затрудняет их сравнение с общепринятыми подходами и воспроизводимость.

Эти недостатки мешают использовать существующие разработки в прикладных задачах, потому что получаемого прироста в точности распознавания недостаточно для компенсации усложнения процесса обучения и архитектуры самих классификаторов. Кроме того, существующие исследования фокусируются на альтернативных функциях потерь, а не на процессе обучения в целом. В связи с этим актуальна разработка метода обучения классификаторов изображений с учетом отношений между классами.

Объектом исследования является машинное обучение нейронных сетей для классификации изображений.

Предметом – способ внесения информации об отношениях между классами в этот процесс.

Целью диссертационной работы является увеличение точности и снижение количества грубых ошибок классификации изображений за счет учета отношений между классами в методе обучения при сохранении порядка вычислительной сложности на уровне общепринятых методов.

Для разработки метода, учитывающего отношения между классами, решены следующие основные **задачи**:

1) предложен способ внесения информации об отношениях между классами в обучающий набор данных;

2) разработана модифицированная триплетная функция потерь с гибкой границей, лежащая в основе обучения классификаторов, и алгоритм ее вычисления, что позволило использовать отношения между классами в процессе обучения;

3) разработаны основные положения метода обучения классификаторов изображений с учетом отношений между классами и алгоритм его реализации, которые позволяют улучшить качество итоговой классификации;

4) разработан программный инструмент для обучения классификаторов, реализующий алгоритм метода обучения с учетом отношений между классами и позволяющий получить более точные классификаторы изображений, уменьшить количество их грубых ошибок;

5) выполнены экспериментальные исследования применения разработанного программного инструмента в прикладных задачах и оценены его эффективность

для распознавания образов, а также применимость для иных задач компьютерного зрения: кластеризации и поиска похожих изображений.

Область исследования соответствует п. 4 и 7 паспорта научной специальности 2.3.5 в части создания инструментальных средств разработки цифровых продуктов (п.4) и разработки алгоритмов обработки изображений и видеоданных (п.7).

Методы исследования. В работе использованы классические методы исследования: декомпозиция, формализация, сравнительный анализ, обобщение и эксперимент. Также для реализации метода и оценки результатов были использованы методы машинного обучения, разведочный анализ данных, математические основы программирования.

Научная новизна работы заключается в разработке метода обучения классификаторов с учетом отношений между классами, для этого:

1) разработана модифицированная триплетная функция потерь с гибкой границей и алгоритм ее вычисления, отличающиеся учетом дополнительных меток, которая позволяет использовать отношения между классами в процессе разработки классификаторов (п.7 паспорта научной специальности);

2) разработаны основные положения метода обучения классификаторов изображений и алгоритм его реализации, отличающиеся учетом отношений между классами, что позволяет более точно настраивать параметры нейронных сетей и улучшить качество распознавания образов (п. 7 паспорта научной специальности);

3) разработан программный инструмент для обучения классификаторов, реализующий алгоритм метода обучения с учетом отношений между классами и позволяющий получить более точные классификаторы изображений, снизить количество их грубых ошибок (п. 4 паспорта научной специальности).

Теоретическая значимость диссертации определяется разработкой и обоснованием нового подхода к обучению классификаторов изображений, использующего отношения между классами. В работе предложена модификация триплетной функции потерь с гибкой границей, которая формализует влияние связей между классами на процесс формирования векторных представлений. Это позволяет по-новому рассматривать задачу метрического обучения – не только как моделирование сходств и различий между отдельными изображениями, но и как структурирование пространства векторных представлений в соответствии с семантическими отношениями между классами. Полученные результаты могут служить основой для последующих исследований в области создания интерпретируемых векторных представлений.

Практическая значимость работы заключается в создании программного инструмента для обучения классификаторов изображений. Инструмент может быть использован для обучения классификаторов изображений в широком спектре прикладных задач. В предметных областях, где известны отношения сходства-различия или иерархии между классами, метод помогает увеличить точность распознавания. Это подтверждается успешным применением метода для решения

прикладных задач: распознавания продуктов и заболеваний желудочно-кишечного тракта, обе решены с точностью распознавания выше 85%, а также классификации изображений руды. В прочих предметных областях инструмент позволяет ускорить процесс обучения классификаторов, что повышает эффективность работ по созданию и поддержке систем распознавания. Это подтверждается его практическим применением для ускорения обучения модели распознавания лиц в системе контроля доступа. Разработанная функция потерь с гибкой границей совместима с современным фреймворком для работы с нейронными сетями TensorFlow и подходит как для обучения с нуля, так и для дообучения. Реализация поддерживает ускорение вычислений за счет параллелизма. Разработанный инструмент подходит и для небольших мобильных моделей, и для серверных решений. Векторные представления, полученные с помощью модели, которая обучена в соответствии с разработанным методом, могут применяться для других задач компьютерного зрения: многометочной классификации, кластеризации, поиска похожих изображений.

Достоверность результатов работы подтверждается корректностью математических преобразований и статистической значимостью результатов экспериментов, а также их воспроизводимостью, так как для проверки метода использовались общедоступные обучающие выборки, в том числе открытые исследовательские датасеты.

Апробация работы. Основные положения диссертационной работы обсуждались на международных научных конференциях «Математические методы в технике и технологиях ММТТ-36» и «Кибер-физические системы: проектирование и моделирование CYBERPHY:2023» (Нижний Новгород, 2023), а также на 5-ой международной конференции по математическому моделированию систем управления, автоматизации и энергоэффективности «SUMMA-2023» (Липецк, 2023), на 5-ой Межвузовской конференции аспирантов, соискателей и молодых ученых «Наука, технологии и бизнес» (Москва, 2023), на международной научно-практической конференции «Индустриальное программирование - 2025» (Москва, 2025) и всероссийской конференции «Студенческая научная весна», посвященной 85-летию со дня рождения академика И.Б. Фёдорова (Москва, 2025).

Разработанный метод обучения и его компоненты реализованы в виде программного инструмента для обучения классификаторов изображений, что подтверждается его регистрацией в реестре программ для ЭВМ. Инструмент использован для ускорения обучения модели распознавания лиц в системе контроля доступа в офис компании ООО «СНС Холдинг» и для снижения количества ошибок классификации при сортировке изображений руды в ООО «СН Голд Майнинг», о чем имеются соответствующие акты внедрения. Также материалы диссертации использованы в учебном процессе МГТУ им. Н.Э. Баумана.

Положения, выносимые на защиту. Соискателем лично получены следующие основные научные результаты, выносимые на защиту:

1) модифицированная триплетная функция потерь с гибкой границей и алгоритм ее вычисления;

2) основные положения метода обучения классификаторов изображений с учетом отношений между классами и алгоритм его реализации;

3) программный инструмент для обучения классификаторов.

Личный вклад соискателя состоит в исследовании подходов к обучению нейронных сетей для классификации изображений; разработке триплетной функции потерь с гибкой границей, положений метода обучения с учетом отношений между классами, инструмента для разработки классификаторов изображений; создании модели представления отношений между классами для параллельной обработки. Без соавторов опубликовано 3 работы.

Публикации. Основные результаты диссертационной работы опубликованы в 8 научных работах, из них: 3 научные статьи в журналах, рекомендованных ВАК РФ (К2); 2 научные статьи, входящие в международную базу Scopus; 2 – в научных журналах, материалах и трудах конференций; 1 свидетельство о государственной регистрации программы для ЭВМ. Общий объем 3,7 п.л. / 3,3 п.л.

Объем и структура диссертации. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы и двух приложений. Объем диссертации составляет 157 страниц. В ней содержится 29 рисунков и 22 таблицы, список литературы из 103 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертации, сформулированы цель, основные задачи, научная новизна, перечислены положения, выносимые на защиту. Отмечены практическая и теоретическая ценность, приведены сведения по апробации и внедрению результатов.

В первой главе проанализированы существующие общепринятые подходы к обучению классификаторов с точки зрения задачи: обучение с нуля и дообучение, и подхода: обучение с помощью *кросс-энтропии* или использование *метрических функций потерь*. В результате их сравнения выявлено, что разрабатываемый метод должен поддерживать и возможность обучения с нуля, и дообучение, так как оба сценария активно используются для создания современных классификаторов. Также сделан вывод, что метрическое обучение более гибкое, чем обучение с помощью кросс-энтропии. Оно позволяет не фиксировать состав и количество меток классов. Классификатор, обученный метрической функцией потерь, предсказывает для объекта не уверенность модели в его принадлежности к классу, а векторное представление. Векторы, соответствующие похожим изображениям, размещаются в пространстве близко, разным – далеко, так их можно классифицировать. Однако в обоих подходах к обучению выявлен общий недостаток – *предположение, что все классы одинаково различны*. В реальном мире многие классы схожи, и их принудительное разделение снижает эффективность

классификатора. Отказ от предположения и переход к моделированию отношений между классами может вести к созданию более интерпретируемых классификаторов, способных адекватно отражать структуру реального мира.

Попытки создать более информативное векторное пространство за счет дополнительной информации предпринимались ранее, однако дополнительные признаки извлекались из самих изображений, а не из классов, либо использовались в функциях, которые ограничивают возможности моделирования векторного пространства. Кроме того, исследования проводились на авторских датасетах, что затрудняет интерпретацию и воспроизводимость полученных результатов. Был сделан вывод о необходимости дальнейшего исследования возможностей моделирования отношений между классами и получения результатов на открытых академических датасетах. Таким образом была поставлена задача улучшения точности классификаторов изображений за счет создания метода обучения, учитывающего отношения между классами. Метод необходимо реализовать в виде программного инструмента. В качестве основы подхода используется метрическое обучение, так как оно позволяет использовать дополнительные данные. Метод предназначен для обучения сверточных нейронных сетей и трансформеров зрения. Классификация предполагает соотнесение изображения только с одним классом, поэтому для корректной работы на изображении должен присутствовать только один распознаваемый объект, изображения каждого класса должны быть статичными и обладать специфическими визуальными признаками.

Вторая глава посвящена разработке требуемого метода обучения классификаторов изображений. Для этого проанализированы входные и выходные данные процесса обучения: изображения, классы и информация об отношениях между ними. В процессе формализации выявлены два основных типа отношений между классами: иерархия (включение) и сходство/различие по признаку. Они представлены в виде дополнительных меток, которые добавляются к основной метке класса и формируют расширенный вектор класса. Метки отношений предложено приводить к спискам фиксированной длины, что упрощает их дальнейшую обработку и дает возможность использовать параллельные вычисления. К примеру, изначальные классы в датасете: «ангина», «глаукома», «катаракта», а расширенные векторы классов: «ангина, горло», «глаукома, глаз», «катаракта, глаз». К каждому классу добавлена дополнительная метка, обозначающая локализацию целевого класса – заболевания. Так в модель можно передать информацию, что «глаукома» и «катаракта» более схожи, так как имеют отношение к одному и тому же органу, в отличие от «ангины». Количество меток отношений должно быть меньше количества классов.

Тогда входные данные для метода – это множество изображений $X = \{x_1, x_2, \dots, x_N\}$, где $x_i \in \mathbb{R}^{h \times w \times c}$, h, w, c – это высота, ширина и количество каналов изображения и матрица расширенных векторов класса Y размерности $N \times (K + 1)$, где K – количество дополнительных меток, N – количество изображений в датасете. При этом дополнительные метки нужны только в

обучающем наборе данных, так как для проверки работы классификатора требуются лишь метки целевых классов. Добавление меток отношений к классам осуществляет специалист предметной области на основании их сходств и различий.

Для разработки основы обучения – функции потерь – исследованы способы обучения многометочных классификаторов, так как модель входных данных для метода предусматривает несколько меток для каждого изображения. В качестве базиса функции потерь была выбрана триплетная функция потерь. Она позволяет использовать метрическое обучение и успешно применяется в задачах классификации. Произведено экспериментальное сравнение модификаций триплетной функции потерь в задаче классификации датасета живописных полотен. В результате выбрана триплетная функция потерь со стратегией выбора полусложных триплетов. Принцип работы триплетной функции потерь: выбрать три изображения якорное, положительное, отрицательное. Якорное и положительное должны иметь один класс, отрицательное – другой. Триплетная функция потерь поощряет модель размещать положительное изображение x_p ближе к якорному x_a , а отрицательное x_n – дальше от x_a (1).

$$L(x_a, x_p, x_n) = \max(D(x_a, x_p) - D(x_a, x_n) + \alpha, 0), \quad (1)$$

где D – функция подсчета расстояния, например, евклидового, а α – параметр границы, определяющий, насколько далеко друг от друга должны находиться векторы изображений разных классов (Рисунок 1).

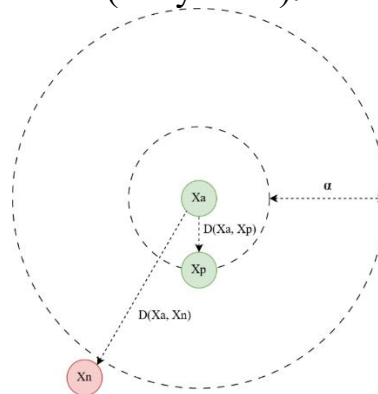


Рисунок 1. Расположение векторов для триплетной функции потерь

Входные данные для разрабатываемой функции потерь представляют матрицу меток и векторные представления изображений. Так как в метрическом обучении модели f с весами W осуществляют преобразование $f(x_i, W): X \rightarrow Z$, входными данными для функции потерь являются векторы $z_i \in \mathbb{R}^d$, где d – длина вектора, определяемая архитектурой модели.

Разработка триплетной функции потерь с учетом отношений между классами. Для расчета любой триплетной функции потерь нужно выбрать отрицательные и положительные векторы для каждого якорного. Их выбор производится на основе равенства классов изображений. Равенство классов двух изображений было решено отождествить с эквивалентностью множеств их меток. Тогда положительным вектором для якорного изображения может быть выбран

только тот, у которого с ним совпадают все метки: и целевого класса, и дополнительные.

Оригинальная формула триплетной функции потерь (1) не учитывает, что классы отрицательных изображений могут иметь разную степень отличия от класса якорного и соответственно располагаться ближе или дальше от него. Желаемый результат можно представить на примере двух триплетов. В первом участвуют объекты А, В и С, во втором А, В, D. А и В – объекты с полностью совпадающими метками. А – якорный, В – положительный в обоих триплетах. С – объект, у которого только часть меток совпадает с А, D – объект, у которого все метки отличаются от А. Получается, С и D – отрицательные объекты. Желаемое расположение: похожие объекты А и В находятся рядом, незначительно отличающийся С – чуть дальше, совсем иной D – на максимальном расстоянии от А и В. Для разработки соответствующей функции потерь были изучены модификации триплетной функции потерь. Выяснилось, что большинство из них построены на манипуляциях с параметром границы α , эта идея была применена. Чтобы иметь возможность настраивать степень влияния различий в дополнительных метках на формирование векторного пространства, в формулу оригинальной триплетной функции потерь (1) вводится функция гибкой границы M вместо константы α . В результате получается (2) – *триплетная функция потерь с гибкой границей*.

$$L(x_a, x_p, x_n, L_a, L_n) = \max(D(x_a, x_p) - D(x_a, x_n) + M(L_a, L_n), 0), \quad (2)$$

где L_a, L_n – наборы меток якорного и отрицательного экземпляров. M – функция расчета параметра границы для заданного триплета.

Желаемое положение векторов можно реализовать, если у триплета А, В, С будет одно значение границы $\alpha = M_1$, а у триплета А, В, D другое – $\alpha = M_2$. Тогда если правильно подобрать M_1 и M_2 , получится расположить векторы как на Рисунке 2. Триплетная функция потерь с гибкой границей будет равна нулю, если объекты размещены в пространстве так, как приведено на Рисунке 2, что и является ожидаемым поведением разрабатываемой функции потерь.

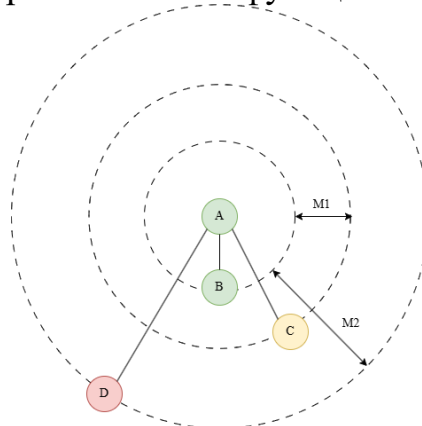


Рисунок 2. Расположение объектов и гибкие границы M1, M2

Формула вычисления гибкой границы $M(L_a, L_n)$. Значение границы должно быть тем больше, чем больше несовпадений между наборами меток L_a, L_n . Его можно рассчитать по (3).

$$N(L_a, L_n) = \alpha * |L_a \setminus L_n|, \quad (3)$$

где α – константа, определяющая минимальную границу между якорным элементом и самым похожим на него отрицательным.

Если несовпадения определенных признаков или уровней иерархии важнее, чем другие, то α можно заменить на вектор границ $A = \{\alpha_1, \dots, \alpha_K\}$, и добавить к (3) взятие максимума. Далее были произведены доработки функции потерь для повышения стабильности ее работы. Так как функция границы не зависит от весов модели, формула вычисления градиента функции потерь не меняется, но меняется ее значение. Поэтому функция вычисления границы (3) была доработана (4).

$$M(L_a, L_n) = \alpha * \ln(|L_a \setminus L_n|) \quad (4)$$

Логарифмическое изменение функции границы более предпочтительно, чем линейное, потому что значение функции потерь, а значит и ее градиента, при линейном изменении может отличаться в несколько раз между триплетами, что негативно влияет на процесс обновления весов. Предлагаемое логарифмическое изменение более стабильное.

В рамках улучшения эффективности функции потерь с гибкой границей в нее был добавлен *фокальный подсчет расстояний*, и (2) принимает вид (5).

$$L(x_a, x_p, x_n, L_a, L_n) = \max\left(e^{\frac{D(x_a, x_p)}{\sigma}} - e^{\frac{D(x_a, x_n)}{\sigma}} + M(L_a, L_n), 0\right) \quad (5)$$

Внедрения экспоненты в расчет расстояний позволяет усилить влияние триплетов, в которых расстояние между якорным изображением и положительным велико, а между якорным и отрицательным – мало. Увеличение вклада сложных триплетов в функцию потерь вкупе с выбранной стратегией выбора полу-сложных примеров позволяет ускорить обучение, но при этом не дестабилизировать его.

Далее проанализированы прочие компоненты и *гиперпараметры обучения*. В результате в завершении главы сформулированы *основные положения* разрабатываемого метода:

1) метод предназначен для обучения сверточных нейронных сетей и трансформеров зрения с нуля или дообучения и применения в задачах классификации;

2) отношения между классами изображений, используемых для обучения, задаются посредством фиксированного количества дополнительных меток, которые могут отражать иерархию или сходства и различия;

3) в основе метода лежит функция потерь с гибкой границей (5), значение которой необходимо минимизировать в процессе обучения за счет изменения весов – коэффициентов модели;

4) изменение весов осуществляется в соответствии с алгоритмом оптимизации Nadam на основе градиентного спуска, параметр скорости потерь

выставляется статическим и снижается только при достижении плато значения функции потерь;

5) для каждого шага алгоритма оптимизации следует использовать сбалансированный по количеству целевых классов фрагмент обучающего датасета среднего размера (128-512 изображений всего или по 2-5 примеров каждого класса);

6) параметр границы гибкой функции потерь выбирается по (4), то есть растет логарифмически по мере увеличения семантического различия между классами;

7) для повышения эффективности обучения применяется фокальный расчет расстояний и стратегия выбора полу-сложных триплетов для расчета функции потерь;

8) для контроля обучения рассчитывается значение функции потерь со стратегией выбора самых сложных триплетов.

Для использования метода необходим датасет изображений, размеченных целевыми классами и метками их отношений. Меток отношений должно быть меньше, чем меток классов. Количество меток, присвоенных классу, должно быть много меньше количества изображений, используемых на шаге обучения. Изображения разных классов должны быть визуально отличимы.

В третьей главе разработаны алгоритмы для реализации метода в *виде программного инструмента* с возможностью параллелизации вычислений. Создан классический *алгоритм предобработки*, включающий нормализацию входных изображений, представление меток в виде целых положительных чисел, стандартизацию их количества в соответствии с моделью входных данных.

Разработан алгоритм *генератора обучающих примеров*, обеспечивающий:

1) равномерное распределение классов в обучающих фрагментах, что предотвращает смещение в сторону более частотных классов;

2) гарантированное наличие положительного и отрицательного объектов для каждого класса;

3) совместимость с разрабатываемым методом (как при обучении с нуля, так и дообучении).

Функция потерь с гибкой границей *реализована в виде алгоритма*. Для параллелизации ее вычисления метки и изображения представлены в виде тензоров. Все операции обработки производятся над векторами и матрицами без циклов и ветвлений, чтобы ускорить вычисления. Реализация параллельных вычислений произведена инструментами библиотеки NumPy и фреймворка TensorFlow. Алгоритм вычисления функции потерь с гибкой границей представлен на Рисунке 3.

В конце главы к разработанному методу обучения добавляется алгоритм классификации *k ближайших соседей*, чтобы получить классификатор из обученной модели, генерирующей векторные представления изображений.

В результате разработан *программный инструмент* для обучения классификаторов изображений и зарегистрирован в реестре программ для ЭВМ.



Рисунок 3. Схема алгоритма функции потерь с гибкой границей

В четвертой главе для апробации метода обоснован выбор критериев его эффективности и метрик для их оценки:

- 1) точность распознавания обученного классификатора – метрика макро- F_1 ;
- 2) количество грубых ошибок обученного классификатора – метрика E , рассчитываемая по количеству изображений, для которых истинный и предсказанный класс имеют различные дополнительные метки;
- 3) качество полученных векторных представлений – значение триплетной функции потерь на валидационной выборке;
- 4) скорость сходимости – количество эпох для завершения обучения.

Для оценки метода было выбрано несколько прикладных задач:

- 1) распознавание продуктов (5421 фото, 81 класс),
- 2) распознавание заболеваний желудочно-кишечного тракта (10619 фото, 23 класса),
- 3) распознавание образов CIFAR (60000 фото, 100 классов).

Первые две – реальные задачи, последняя – академическая, широко используемая для оценки и сравнения методов распознавания образов. Метрики на открытом исследовательском датасете обеспечивают воспроизводимость результатов, полученных в данной работе. Оценки работы на датасетах с реальными фотографиями из разных предметных областей подтверждают применимость метода для решения прикладных задач. Каждый из датасетов содержит дополнительные метки классов. Для CIFAR это категории, для датасетов продуктов и болезней – категории и подкатегории.

Экспериментальная проверка метода. Для корректного сравнения классификаторы были обучены по разработанному методу и по общепринятым. Для проверки метода при обучении с нуля использовалась FaceNet-подобная модель классификатора. В результате изменения подхода к обучению на предлагаемый метод значение функции ошибки на датасете CIFAR сократилось на 13%. Это свидетельствует о его способности эффективнее располагать векторные представления изображений в пространстве.

Точность классификации обученных моделей на 100 классах датасета CIFAR была рассчитана по выбранной ранее формуле макро- F_1 и приведена в процентах в Таблице 1.

Таблица 1.

Точность распознавания классов CIFAR

Функция потерь	После обучения, %	После обучения с фокальными расстояниями, %	После обучения с фокальными расстояниями и логарифмической границей, %	После обучения по предлагаемому методу, %
Кросс-энтропия	43	не применимо	не применимо	не применимо
Триплетная функция	42	46	не применимо	не применимо
Триплетная функция с гибкой границей	46	48	50	52

Согласно данным в Таблице 1, предлагаемый метод демонстрирует на 9% более высокую точность, чем базовый метод обучения – кросс-энтропия, и на 10% более высокую, чем метод с использованием оригинальной триплетной функции потерь. Предлагаемый метод так же показывает снижение количества грубых ошибок классификации на 28% в сравнении с лучшим из общепринятых подходов. Оценка времени обучения показала, что триплетная функция потерь с гибкой границей не замедляет процесс обучения – эпоха завершения обучения примерно одинаковая для всех обученных моделей.

Далее в Таблице 2 приведены результаты работы метода в задачах классификации продуктов и заболеваний при обучении с нуля. В результате анализа

результатов обучения классификаторов с нуля можно сделать вывод, что предлагаемый метод позволяет улучшить точность классификации – точность улучшилась на 7% при классификации продуктов и на 3% при классификации заболеваний. Количество грубых ошибок в этих задачах снизилось на 50% и 33% соответственно.

Таблица 2.

Результаты обучения классификаторов продуктов и заболеваний

Подход к обучению	Точность, продукты %	Точность, заболевания, %
Обучение с кросс-энтропией	83	83
Метрическое обучение с триплетной функцией потерь	83	82
Предлагаемый метод	90	86

Далее оценена применимость метода к *дообучению* предварительно обученных моделей, используя классификатор ConvNext, предобученный на ImageNet, результаты приведены в Таблице 3. Архитектура ConvNext выбрана в результате экспериментального сравнения с другими популярными моделями для классификации изображений: XCiT, Swin, ViT, EfficientNet.

Таблица 3.

Результаты дообучения классификаторов продуктов и заболеваний

Подход к обучению	Точность, продукты, %	Точность, заболевания %
Обучение с кросс-энтропией	95	73
Метрическое обучение с триплетной функцией потерь	92	70
Предлагаемый метод	98	76

Анализ Таблицы 3 показывает, что классификатор, обученный по предлагаемому методу, имеет на 6% более высокую точность, чем метрическое обучение с триплетной функцией потерь, и на 3% в сравнении с кросс-энтропией. Количество грубых ошибок в этих задачах сократилось на 50% при классификации продуктов и на 33% при распознавании болезней. Таким образом, разработанный метод подходит и для дообучения классификатора.

Кроме общедоступных датасетов, метод и его компоненты также были апробированы при решении коммерческих задач: для ускорения обучения классификатора для распознавания лиц и для снижения ошибок классификатора изображений руды. В первом случае были применены предлагаемые в методе способы стабилизации триплетной функции потерь, во втором обучение классификатора проводилось с помощью разработанного программного инструмента, что позволило снизить количество ошибок при распознавании на 4%. Полученные результаты внедрения зафиксированы в соответствующих актах.

Далее была исследована *применимость* полученных векторных представлений для *иных задач компьютерного зрения*. На Рисунке 4 изображены пространство признаков, полученных из обычного классификатора, и пространство

векторных представлений, полученных классификатором с учетом отношений между классами. Он наглядно демонстрирует возможность применения метода для кластеризации.

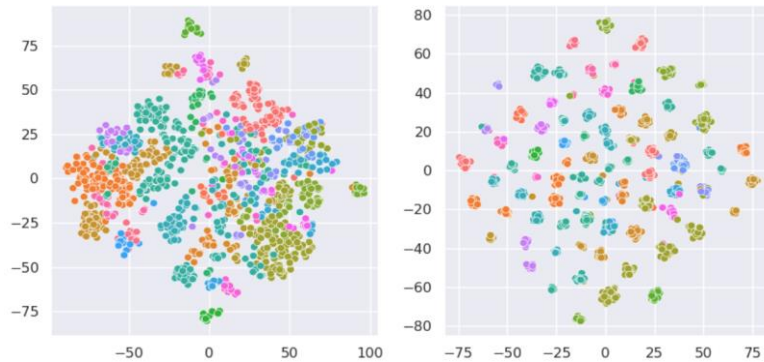


Рисунок 4. Пространства признаков (слева) и векторных представлений (справа) категорий

На Рисунке 5 изображен пример использования метода в задаче поиска похожих изображений по целевому (слева) в сравнении с поиском по признакам из классификатора, обученного с помощью кросс-энтропии. Метод помогает модели учитывать сразу несколько важных признаков: жанр, стиль, художника, а не только общий визуальный образ.

Также была проверена устойчивость метода к шуму. При добавлении ошибок в датасет метод все еще положительно влияет на качество распознавания, однако его эффективность снижается.

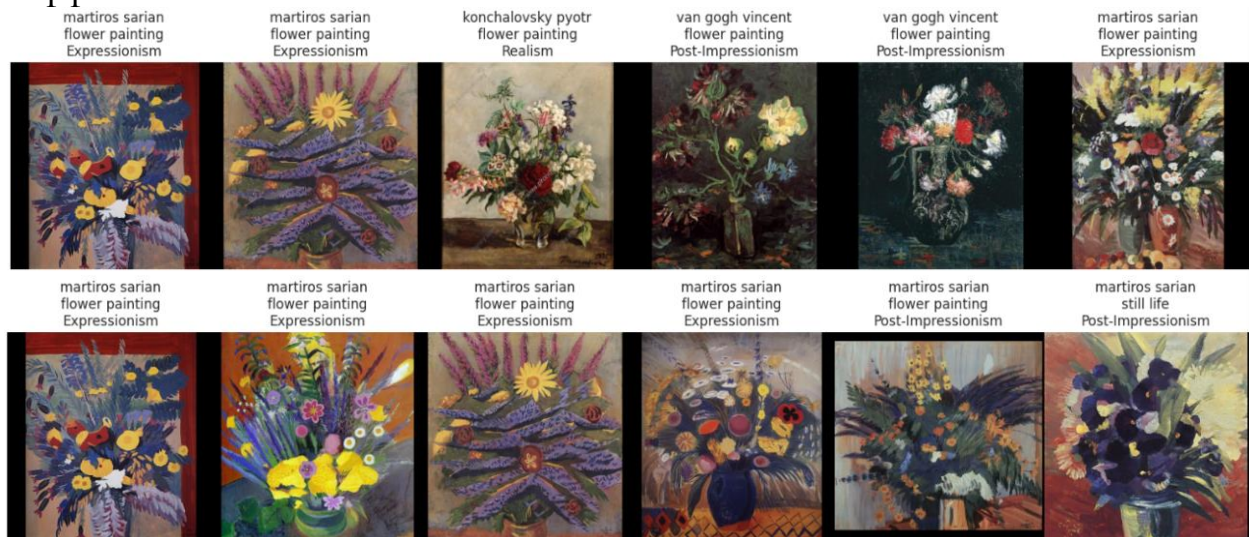


Рисунок 5. Ближайшие изображения к вектору признаков целевого (сверху) и к векторному представлению целевого (снизу)

В завершении главы была проведена оценка вычислительной сложности реализации метода. Так как количество меток изображения K обычно много меньше количества изображений в батче N , верхняя оценка вычислительной сложности разработанной функции потерь остается $O(N^3)$. Однако в задачах с высокой разреженностью меток, где K сопоставимо с N , верхняя оценка вычислительной сложности алгоритма будет равна $O(K * N^3)$. Поэтому применение метода

рекомендуется в задачах, где количество меток для одного изображения много меньше, чем количество изображений.

В заключении сформулированы основные результаты диссертационной работы:

1) выполнен анализ существующих подходов к обучению классификаторов изображений, определены основные компоненты процесса, сформулирована проблема предположения, что все классы одинаково различны, и задача разработки метода обучения классификаторов с учетом отношений между классами;

2) исследованы отношения между классами и предложен способ их формализации для параллелизации их обработки;

3) разработана модифицированная триплетная функция потерь с гибкой границей для учета дополнительных меток классов во время обучения, предложены формулы для вычисления оптимального параметра границы;

4) проведен анализ влияния гиперпараметров обучения на скорость и стабильность оптимизации, сформулированы основные положения метода обучения классификаторов с учетом отношений между классами;

5) разработан алгоритм функции потерь с гибкой границей с векторизацией операций и сохранением порядка вычислительной сложности на уровне традиционных функций потерь для обучения классификаторов;

6) сформулированы критерии эффективности метода обучения, обоснован выбор метрик для их оценки;

7) метод апробирован в прикладных задачах, его применение улучшило точность классификации: в задаче распознавания образов CIFAR на 10% (количество грубых ошибок снизилось на 28%), распознавания продуктов – на 7% (количество грубых ошибок снизилось на 75%), классификации заболеваний ЖКТ – на 6% (количество грубых ошибок снизилось на 55%), относительно метрического обучения с помощью триплетной функции потерь;

8) во всех рассмотренных задачах классификации разработанный метод позволил получить более высокую точность распознавания, чем обучение с помощью кросс-энтропии, что позволяет создавать классификаторы с возможностью менять количество классов после обучения без потери точности классификации.

9) метод помогает сформировать более качественные векторные представления – значение функции потерь в экспериментах снизилось на 5 – 36% относительно общепринятых подходов;

10) метод эффективен как для академических датасетов, так и для реальных несбалансированных данных с разнообразными визуальными признаками;

11) модели, обученные по предлагаемому методу, можно применять и в других задачах компьютерного зрения: кластеризации, поиске похожих изображений, с ограничениями – в многометочной классификации, что расширяет область практического применения разработки.

Поставленная цель и задачи диссертационного исследования выполнены в полном объеме. Решена актуальная научная задача создания метода обучения, обеспечивающего улучшение точности и снижение количества грубых ошибок классификации, имеющая существенное значение для развития компьютерного зрения. Результаты работы апробированы в нескольких открытых исследовательских задачах и внедрены в деятельность организаций, где применяются для повышения точности распознавания образов и снижения временных затрат на обучение нейронных сетей.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Петрова Я.С. Методика обучения классификаторов изображений с использованием дополнительных меток // Моделирование, оптимизация и информационные технологии. 2025. Т. 13. № 2. С. 1–13. (0,9 п.л.)
2. Петрова Я.С. Анализ методов улучшения сходимости триплетной функции потерь в многометочной классификации // Моделирование, оптимизация и информационные технологии. 2025. Т. 13. № 4. С. 1–9. (0,6 п.л.)
3. Иванова Г.С., Петрова Я.С. Анализ методов компьютерного зрения для систематизации живописных полотен // Нейрокомпьютеры: разработка, применение. 2022. Т. 24. № 6. С. 20–29. (0,6 п.л./0,5 п.л.)
4. Petrova I.S., Ivanova G.S. Comparison of Cross-Entropy Loss and Metric Losses in the Multi-Label Classification // 2023 5th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA): conference proceedings. IEEE, 2023. P. 222–225. (0,3 п.л./0,2 п.л.)
5. Petrova I.S., Ivanova G.S. A Comparison of Feature Extraction Models for Images with Multiple Annotations // Cyber-Physical Systems. Studies in Systems, Decision and Control. 2024. Vol. 554. P. 81–90. (0,6 п.л./0,5 п.л.)
6. Петрова Я.С. Сравнительный анализ моделей для извлечения признаков из изображений // Наука, технологии и бизнес: сб. ст. М.: Издательство МГТУ им. Н.Э. Баумана, 2023. С. 368–373. (0,4 п.л.)
7. Петрова Я. С., Иванова Г.С. Сравнение моделей извлечения признаков из изображений с несколькими аннотациями // Математические методы в технологиях и технике. 2023. № 2. С. 71–74. (0,3 п.л./0,2 п.л.)
8. Программа для создания многометочного классификатора изображений: свидетельство о государственной регистрации программы для ЭВМ № 2025688066 РФ. / Я.С. Петрова; заявл. 07.10.2025; опубл. 16.10.2025; Бюлл. № 2025687047.