Министерство образования и науки Российской Федерации
Московский физико-технический институт (государственный университет)

Физтех-школа прикладной математики и информатики
Кафедра системного программирования ИСП РАН
Лаборатория (laboratory name)

Выпускная квалификационная работа бакалавра

Изучение вариантов обеспечения устойчивости ResNet-подобных моделей компьютерного зрения к небольшим сдвигам входного изображения

Автор:

Студент

Полев Алексей Михайлович

Научный руководитель:

к.ф.-м.н.

Самосюк Алексей Владимирович

Москва 2025

АННОТАЦИЯ

В данной работе проведено исследование проблемы пространственной инвариантности в сверточных нейронных сетях (CNN) и её влияния на стабильность работы алгоритмов компьютерного зрения. Особое внимание уделено артефактам, возникающим при субпиксельных сдвигах входных изображений, которые могут существенно влиять на качество классификации и детекции объектов.

В теоретической части работы формализована проблема отсутствия полной инвариантности к сдвигам в современных CNN-архитектурах, проанализированы причины этого явления, связанные с операциями субдискретизации (даунсэмплинга), и рассмотрены существующие подходы к её решению, включая методы анти-алиасинга и полифазной выборки.

Экспериментальная часть исследования сфокусирована на сравнительном анализе стандартных архитектур (ResNet-50, VGG-16, YOLOv5) и их модифицированных версий с различными методами обеспечения инвариантности к сдвигам. Разработана методология тестирования, включающая генерацию последовательностей изображений с субпиксельными сдвигами объектов и комплексную систему метрик для оценки стабильности.

Результаты экспериментов демонстрируют, что стандартные CNN-архитектуры проявляют значительную нестабильность даже при минимальных сдвигах входных данных. Применение методов анти-алиасинга (BlurPool) существенно улучшает стабильность, а наилучшие результаты показывает внедрение техники полифазной выборки (TIPS), которая почти полностью устраняет артефакты пространственной вариативности при небольшом увеличении вычислительной сложности.

На основе проведенного исследования сформулированы практические рекомендации по выбору архитектур и методов обеспечения инвариантности к сдвигам для различных задач компьютерного зрения, что может быть полезно при разработке систем, требующих высокой точности и стабильности работы.

Ключевые слова: сверточные нейронные сети, пространственная инвариантность, анти-алиасинг, BlurPool, TIPS, компьютерное зрение, YOLOv5.

Содержание

Список сокращений и обозначений

- **CNN** — сверточная нейронная сеть (Convolutional Neural Network)
- **IoU** — метрика пересечения над объединением (Intersection over Union)
- **TIPS** — полифазная выборка с инвариантностью к сдвигам (Translation Invariant Polyphase Sampling)
- **FPS** — кадров в секунду (Frames Per Second)
- **MSB** — максимальное смещение выборки (Maximum-Sampling Bias)
- **AA-VGG16** — VGG16 с анти-алиасингом (Anti-Aliased VGG16)
- **AA-ResNet50** — ResNet50 с анти-алиасингом (Anti-Aliased ResNet50)
- **AA-YOLOv5** — YOLOv5 с анти-алиасингом (Anti-Aliased YOLOv5)
- **YOLO** — объектный детектор «вы смотрите только один раз» (You Only Look Once)
- **R-CNN** — регионная сверточная нейронная сеть (Region-based Convolutional Neural Network)
- **SSD** — детектор с одним проходом (Single Shot Detector)
- **PANet** — сеть агрегации путей (Path Aggregation Network)
- **FPN** — сеть пирамиды признаков (Feature Pyramid Network)
- **CSPDarknet** — базовая сеть YOLOv5 (Cross Stage Partial Darknet)
- **TDF** — функция расхождения трансляций (Translation Discrepancy Function)

1 Введение

Актуальность проблемы

Сверточные нейронные сети (CNN) сегодня являются ключевым инструментом в решении широкого спектра задач компьютерного зрения, включая классификацию изображений, сегментацию, детекцию объектов и другие. Их популярность и эффективность обусловлены способностью к автоматическому извлечению иерархии признаков из необработанных данных и высокой точностью работы в различных условиях. Теоретические основы CNN предполагают, что они должны обладать свойством инвариантности к пространственным преобразованиям, в частности, к сдвигам входных данных. Это означает, что одинаковые объекты, расположенные в разных частях изображения, должны распознаваться с одинаковой точностью и уверенностью.

Однако практика показывает, что современные архитектуры CNN не обладают полной инвариантностью к сдвигам. Небольшие, даже субпиксельные смещения объектов на входном изображении могут приводить к значительным изменениям в выходных результатах сети. Эта проблема, часто упускаемая из виду при традиционной оценке моделей на тестовых выборках, может иметь серьезные последствия в реальных приложениях компьютерного зрения, особенно в критически важных областях, таких как автономные транспортные средства, системы видеонаблюдения, медицинская диагностика и робототехника.

Отсутствие стабильности предсказаний при малых смещениях объектов может привести к:

- Ложным срабатываниям или пропускам в системах обнаружения объектов
- Нестабильной работе алгоритмов слежения за объектами
- Некорректной сегментации медицинских изображений
- Ошибкам в системах управления роботами и беспилотными автомобилями
- Снижению надежности систем биометрической идентификации

Причины нарушения инвариантности к сдвигам в CNN связаны с операциями субдискретизации (даунсэмплинга), такими как max-pooling и свёртка с шагом (stride) больше единицы. Эти операции позволяют уменьшать пространственное разрешение карт признаков, что необходимо для снижения вычислительной сложности и обобщающей способности сети, но одновременно вносят пространственную зависимость, делая сеть чувствительной к точному положению входных паттернов.

В последние годы было предложено несколько подходов к решению проблемы пространственной вариативности CNN, включая методы антиалиасинга (например, BlurPool), полифазную выборку с инвариантностью к сдвигам (TIPS) и различные модификации архитектур. Однако систематическое исследование влияния этих методов на стабильность работы различных типов CNN в контексте разных задач компьютерного зрения остается актуальной проблемой.

Данная работа направлена на всестороннее исследование артефактов пространственной инвариантности в современных CNN-архитектурах, анализ их влияния на производительность моделей и оценку эффективности различных методов повышения устойчивости к пространственным сдвигам. Особое внимание уделяется сравнению поведения классификационных моделей и моделей детекции объектов, таких как YOLO, при субпиксельных сдвигах входных данных, что позволяет выявить специфические проблемы и предложить целевые решения для различных типов архитектур.

Цель и задачи исследования

Целью данной работы является комплексное исследование проблемы отсутствия полной инвариантности к пространственным сдвигам в современных архитектурах сверточных нейронных сетей, разработка и оценка методов повышения их устойчивости к смещениям входных данных.

Для достижения поставленной цели необходимо решить следующие **задачи**:

1. Провести анализ существующих исследований и методов в области пространственной инвариантности CNN, включая:
    - Теоретические основы инвариантности к сдвигам в сверточных архитектурах

- Методы анти-алиасинга в нейронных сетях
- Подходы к обеспечению инвариантности в моделях детекции объектов
- Техники полифазной выборки с инвариантностью к сдвигам

2. Формализовать проблему пространственной инвариантности и разработать математическую модель для описания влияния операций субдискретизации на стабильность представлений в CNN.

3. Разработать методологию тестирования и метрики для количественной оценки степени инвариантности моделей к пространственным сдвигам, включая:
   - Методику генерации последовательностей изображений с контролируемыми субпиксельными сдвигами
   - Метрики стабильности векторов признаков (косинусное сходство)
   - Метрики стабильности предсказаний (дрейф уверенности, стабильность IoU)
   - Визуализации для качественного анализа эффектов

4. Провести экспериментальное исследование влияния субпиксельных сдвигов на стабильность работы различных CNN-архитектур:
   - Классификационных моделей (VGG16, ResNet50)
   - Моделей детекции объектов (YOLOv5)
   - Их модифицированных версий с различными методами повышения инвариантности

5. Реализовать и сравнить различные методы повышения инвариантности к сдвигам:
   - Классический анти-алиасинг (BlurPool)
   - Translation Invariant Polyphase Sampling (TIPS)
   - Гибридные подходы

6. Провести аблационное исследование для выявления влияния различных факторов на пространственную инвариантность:

- Размера рецептивного поля
- Разных типов операций пулинга
- Параметров анти-алиасинга

7. Сформулировать практические рекомендации по выбору архитектур и методов обеспечения инвариантности для различных задач компьютерного зрения.

Научная новизна и практическая значимость

Научная новизна данной работы заключается в следующем:

1. Проведено комплексное сравнительное исследование проблемы пространственной инвариантности в различных типах CNN-архитектур (классификаторы и детекторы) с использованием единой методологии и системы метрик.

2. Разработана и апробирована методика генерации контролируемых последовательностей изображений с субпиксельными сдвигами, позволяющая точно измерять степень инвариантности моделей к пространственным преобразованиям.

3. Предложены новые метрики и визуализации для количественной и качественной оценки стабильности работы CNN при пространственных сдвигах входных данных.

4. Впервые проведено систематическое сравнение эффективности различных методов обеспечения инвариантности (BlurPool, TIPS) в контексте моделей детекции объектов (YOLOv5).

5. Проведено аблационное исследование, позволяющее выявить ключевые факторы, влияющие на степень пространственной инвариантности в современных CNN.

Практическая значимость работы определяется следующими аспектами:

1. Результаты исследования позволяют более осознанно подходить к выбору архитектур CNN для задач, требующих высокой стабильности предсказаний при малых изменениях входных данных.

2. Предложенные модификации архитектур с использованием методов анти-алиасинга и полифазной выборки могут быть непосредственно применены для улучшения стабильности существующих систем компьютерного зрения.

3. Разработанная методология тестирования и система метрик могут использоваться как инструментарий для оценки пространственной инвариантности при разработке новых архитектур нейронных сетей.

4. Сформулированные рекомендации по выбору методов обеспечения инвариантности имеют практическую ценность для разработчиков систем компьютерного зрения в таких областях, как:

    - Автономные транспортные средства и роботы, где стабильность детекции объектов критически важна для безопасности
    - Медицинская визуализация, где точность локализации патологий напрямую влияет на качество диагностики
    - Системы видеоаналитики, требующие надежного отслеживания объектов при их перемещении
    - Промышленные системы контроля качества, где незначительные изменения положения контролируемых объектов не должны влиять на результаты анализа

Структура работы

Диссертация состоит из введения, трех глав, заключения, списка литературы и приложений. Общий объем работы составляет 120 страниц, включая 25 рисунков и 10 таблиц. Список литературы содержит 35 наименований.

В главе 1 представлен обзор литературы по проблеме пространственной инвариантности в сверточных нейронных сетях. Рассмотрены теоретические основы инвариантности к сдвигам, проанализированы причины нарушения этого свойства в современных CNN-архитектурах, описаны существующие методы повышения устойчивости к пространственным преобразованиям. Особое внимание уделено специфике проблемы инвариантности в моделях детекции объектов.

В главе 2 изложены теоретические основы исследования. Формализована проблема пространственной инвариантности, представлен математический аппарат для описания влияния операций субдискретизации на стабильность представлений в CNN. Подробно рассмотрены архитектуры исследуемых моделей (VGG16, ResNet50, YOLOv5) и методы повышения их инвариантности к сдвигам (BlurPool, TIPS). Приведен детальный анализ рецептивных полей в различных архитектурах и их связи с проблемой пространственной инвариантности.

В главе 3 описана экспериментальная часть исследования. Представлена методология тестирования, включая генерацию контрольных последовательностей изображений с субпиксельными сдвигами, определены используемые метрики, детально описан процесс проведения экспериментов. Приведены результаты сравнительного анализа различных архитектур и методов повышения инвариантности, представлены визуализации, демонстрирующие эффекты пространственных сдвигов на работу моделей. Проведен анализ производительности модифицированных архитектур и оценка компромисса между вычислительной сложностью и стабильностью предсказаний.

В заключении обобщены основные результаты работы, сформулированы выводы и рекомендации по выбору архитектур и методов обеспечения инвариантности для различных задач компьютерного зрения, а также обозначены перспективные направления дальнейших исследований в данной области.

2 Обзор литературы

2.1 Инвариантность к сдвигу в CNN-классификаторах

Сверточные нейронные сети (CNN) теоретически должны обладать определенной степенью инвариантности к позиционным сдвигам входных данных благодаря механизму разделения весов и локальным рецептивным полям [1]. Однако, как показывают исследования последних лет, современные CNN демонстрируют ограниченную инвариантность к сдвигам, что противоречит интуитивным ожиданиям.

Исторически, LeCun et al. [1] первыми формально описали свойство эквивариантности сверточных сетей к сдвигам, выделив ключевые свойства CNN — локальность связей, разделение весов и пространственный пулинг. Теоретически, операция свёртки обладает эквивариантностью к сдвигам: если входное изображение сдвигается, то соответствующим образом сдвигаются и карты признаков.

Несмотря на теоретические предпосылки, эмпирические исследования выявили существенные ограничения в инвариантности современных CNN к сдвигам. Engstrom et al. [2] продемонстрировали, что даже небольшие сдвиги входных изображений могут значительно снизить точность классификации современных архитектур, включая ResNet.

Zhang [1] в своем фундаментальном исследовании идентифицировал операции даунсэмплинга (max-pooling и свертку с шагом больше 1) как основной источник нарушения инвариантности к сдвигам. Автор показал, что субпиксельные сдвиги входных изображений приводят к значительным изменениям в активациях нейронов и нестабильности предсказаний модели.

Azulay and Weiss [3] продемонстрировали, что проблема инвариантности может быть систематически исследована через призму классической теории обработки сигналов. Отсутствие антиалиасинговых фильтров перед операциями субдискретизации приводит к высокочастотному шуму в представлениях признаков, делая модель чувствительной к малым сдвигам.

Для количественной оценки инвариантности используются различные метрики. Zhang [1] предложил метрику стабильности предсказаний, основанную на изменении выходных вероятностей модели при субпиксельных

сдвигах. Также распространены измерения косинусного сходства между векторами признаков, полученными из оригинального и сдвинутого изображений.

2.2 Методы анти-алиасинга в нейронных сетях

После идентификации алиасинга как основной причины нарушения инвариантности, исследователи предложили ряд методов решения этой проблемы, адаптированных к особенностям нейронных сетей. Эти методы основаны на принципах теории обработки сигналов, но учитывают специфику архитектур глубокого обучения и ограничения, связанные с вычислительной эффективностью.

2.2.1 BlurPool: анти-алиасинг для CNN

Наиболее значимым подходом к борьбе с алиасингом стал метод BlurPool, предложенный Zhang [1]. В BlurPool операции max-pooling и свертки с шагом больше 1 модифицируются таким образом, что перед непосредственной субдискретизацией применяется размытие с использованием фиксированного низкочастотного фильтра. Автор исследовал различные типы фильтров, включая простое усреднение (box filter), треугольный фильтр (binomial filter) и фильтр Гаусса, показав, что даже простейшие из них значительно улучшают инвариантность сети к сдвигам.

Ключевое преимущество BlurPool — архитектурная простота и возможность интеграции в существующие модели без необходимости переобучения с нуля. Замена стандартных операций пулинга и свертки с шагом на их «размытые» аналоги может быть выполнена постфактум в предобученных моделях с сохранением большей части весов.

Zou et al. [4] продемонстрировали, что применение BlurPool к архитектурам ResNet не только улучшает их инвариантность к сдвигам, но и повышает устойчивость к состязательным атакам (adversarial attacks).

2.2.2 TIPS: полифазная выборка с инвариантностью к сдвигам

Альтернативный и более продвинутый подход был предложен Saha и Gokhale [5] под названием Translation Invariant Polyphase Sampling (TIPS). В отличие от BlurPool, использующего фиксированный низкочастотный

фильтр, TIPS применяет полифазное разложение сигнала для явного моделирования и компенсации эффектов субдискретизации.

Основная идея TIPS заключается в разделении сигнала на несколько «фаз» в соответствии с его позицией относительно сетки субдискретизации. Каждая фаза обрабатывается отдельно, после чего результаты объединяются для получения представления, инвариантного к исходному положению сигнала.

Математически TIPS можно рассматривать как обобщение идеи кросс-корреляции с циклическим сдвигом, гарантирующее одинаковый выход модели для всех целочисленных сдвигов входного сигнала. TIPS распространяет этот принцип на субпиксельные сдвиги, обеспечивая более полную инвариантность.

Исследования показывают, что TIPS обеспечивает наилучшую теоретическую гарантию инвариантности среди существующих методов, хотя требует более значительных изменений в архитектуре сети и может быть вычислительно более затратным по сравнению с BlurPool.

2.3 Инвариантность к сдвигам в детекторах объектов

В то время как проблема инвариантности к сдвигам хорошо изучена для классификационных моделей, её влияние на детекторы объектов представляет отдельную и более сложную задачу. Детекция объектов требует не только определения класса объекта, но и точной локализации его положения. Это делает проблему инвариантности особенно критичной для детекторов объектов, так как нарушения стабильности могут привести к значительным ошибкам в определении положения ограничивающих рамок.

Современные детекторы объектов, такие как одностадийный YOLO [6], широко используют CNN в качестве основы для извлечения признаков и наследуют проблемы инвариантности, присущие этим архитектурам. Исследования Papkovsky et al. [7] показали, что небольшие субпиксельные сдвиги входных изображений приводят к значительным изменениям в предсказанных ограничивающих рамках даже для современных детекторов.

Ключевой проблемой является дрейф центра ограничивающей рамки — явление, при котором центр предсказанной рамки смещается при изменении положения объекта. Это особенно критично для задач, требующих

высокой точности локализации, таких как медицинская диагностика или прецизионная робототехника.

Для оценки устойчивости детекторов используются специфические метрики: стабильность IoU (Intersection over Union), дрейф центра ограничивающей рамки и стабильность уверенности детекции. Низкая стабильность IoU указывает на чувствительность детектора к малым пространственным преобразованиям входа.

Адаптация методов анти-алиасинга к детекторам объектов представляет нетривиальную задачу из-за сложности их архитектур. Для одностадийных детекторов, таких как YOLO, Papkovsky et al. [7] предложили специализированную версию BlurPool, учитывающую особенности архитектуры с множественными выходами на разных масштабах.

Нестабильность детекторов объектов при малых сдвигах входных данных имеет серьезные практические последствия. В системах видеонаблюдения это может приводить к прерывистым траекториям и ложным срабатываниям алгоритмов трекинга. В беспилотных транспортных средствах нестабильность влияет на точность определения положения препятствий, что критично для безопасности.

Решение проблемы инвариантности к сдвигам в детекторах объектов имеет важное практическое значение для повышения надежности систем компьютерного зрения в критически важных приложениях. Хотя методы на основе BlurPool показывают многообещающие результаты, эта область остается активным направлением исследований.

3 Исследование и построение решения задачи

В данной главе описывается исследовательская часть работы, основанная на анализе литературы, представленном в предыдущей главе. Рассматриваются методы улучшения инвариантности к сдвигам в современных нейронных сетях и предлагаемые модификации архитектур.

3.1 Математическая формализация и модификации архитектур

Для формализации проблемы инвариантности к сдвигам рассмотрим нейронную сеть как функцию $f : \mathbb{R}^{H \times W \times C} \to \mathbb{R}^K$. Для операции сдвига $\mathcal{T}_{\Delta h\, \Delta w}$, инвариантность к сдвигам означает, что $f(\mathcal{T}_{\Delta h\, \Delta w}(x)) \approx f(x)$. Для оценки инвариантности используется метрика стабильности предсказаний при субпиксельных сдвигах, измеряющая среднее сходство выходов модели для оригинального и сдвинутых изображений.

В работе исследуются две основные методики борьбы с алиасингом: BlurPool и TIPS, которые были адаптированы для различных архитектур нейронных сетей.

3.1.1 BlurPool в классификационных моделях

Для классификационных моделей VGG16 и ResNet50 реализованы модификации с применением BlurPool, заключающиеся в замене операций даунсэмплинга на их аналоги с предварительной низкочастотной фильтрацией:

$$\text{BlurPool}_{m\,s} = \text{Subsample}_s \circ \text{Blur}_m$$

где Subsample_s — операция субдискретизации с шагом s, а Blur_m — операция свёртки с фиксированным низкочастотным фильтром размера $m \times m$.

Используются биномиальные фильтры:

- **Triangle-3**: $[1, 2, 1] \times [1, 2, 1]^T$ / 16
- **Binomial-5**: $[1, 4, 6, 4, 1] \times [1, 4, 6, 4, 1]^T$ / 256

Модификации применяются следующим образом:

- **MaxPool** $\to \text{Subsample}_s \circ \text{Blur}_m \circ \text{Max}_{k\,1}$

- **Свертка с шагом** $\rightarrow \mathrm{BlurPool}_{m\,s} \circ \mathrm{ReLU} \circ \mathrm{Conv}_{k\,1}$
- **AvgPool** $\rightarrow \mathrm{BlurPool}_{m\,s}$

Важное преимущество метода BlurPool — возможность применения к предобученным моделям с минимальным увеличением вычислительных затрат (<1

3.1.2 TIPS для повышения инвариантности

TIPS (Translation Invariant Polyphase Sampling) — более продвинутый подход, основанный на разделении сигнала на несколько фаз в зависимости от его положения относительно сетки субдискретизации:

$$\mathrm{TIPS}(x) = \frac{1}{s^2} \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} \mathrm{Subsample}_s(\mathrm{Shift}_{(i\,j)}(x)) \tag{1}$$

где s - коэффициент субдискретизации, а $\mathrm{Shift}_{(i\,j)}$ - операция сдвига.

В реализации TIPS для слоя с шагом s создаются s^2 отдельных ветвей, каждая обрабатывает сдвинутую версию входного тензора, результаты объединяются для формирования инвариантного представления.

3.2 Архитектура YOLOv5 и её модификации

Особое внимание уделяется детектору YOLOv5, состоящему из трех компонентов:

- Backbone (CSPDarknet) — извлекает признаки
- Neck (PANet) — объединяет признаки разных уровней
- Head — преобразует признаки в предсказания

Модификации затрагивают операции даунсэмплинга в backbone и neck:

- YOLOv5-BlurPool: замена сверток с шагом 2 на свертку с шагом 1 + BlurPool
- YOLOv5-TIPS: замена сверток с шагом 2 на TIPS-модули

От модификаций ожидаются: повышение стабильности предсказаний, уменьшение дрейфа центра ограничивающей рамки, более высокая стабильность IoU между предсказанными и истинными рамками.

3.3 Методология оценки инвариантности

3.3.1 Метрики для классификационных моделей

Для классификационных моделей используются:

- **Top-1 Accuracy (Acc)**: доля правильно классифицированных изображений.
- **Consistency (Cons)**: вероятность одинакового предсказания для исходного и сдвинутого изображения:

$$\text{Cons} = \mathbb{E}_{x\,\delta}\left[\mathbb{I}\left(\underset{c}{\operatorname{argmax}}\, f(x)_c = \underset{c}{\operatorname{argmax}}\, f(\mathcal{T}_\delta(x))_c\right)\right] \tag{2}$$

- **Stability (Stab)**: среднее косинусное сходство между выходными представлениями:

$$\text{Stab} = \mathbb{E}_{x\,\delta}\left[\frac{f(x)\cdot f(\mathcal{T}_\delta(x))}{\|f(x)\|\cdot\|f(\mathcal{T}_\delta(x))\|}\right] \tag{3}$$

3.3.2 Метрики для детекторов объектов

Для детекторов объектов используются:

- **mAP**: средняя точность детекции при различных порогах IoU.
- **IoU Stability (IS)**: стабильность пересечения над объединением рамок при сдвигах:

$$\text{IS} = \mathbb{E}_{x\,\delta\,b}\left[\text{IoU}\left(b\,\mathcal{T}_{-\delta}(b_\delta)\right)\right] \tag{4}$$

где b — ограничивающая рамка для исходного изображения, b_δ — соответствующая рамка для сдвинутого изображения, а $\mathcal{T}_{-\delta}$ — обратный сдвиг для компенсации смещения.

- **Center Drift (CD)**: среднее смещение центров рамок при сдвигах:

$$\text{CD} = \mathbb{E}_{x\,\delta\,b}\left[\|\text{center}(b) - \text{center}(\mathcal{T}_{-\delta}(b_\delta))\|_2\right] \tag{5}$$

3.3.3 Протокол тестирования

Протокол тестирования включает:

1. **Генерация сдвигов**: Для каждого изображения создается набор сдвинутых версий с точностью до $1/8$ пикселя в диапазоне $[-8\,8]$ пикселей.

2. **Предобработка**: Стандартизация размера до 224×224 пикселей для классификации и 640×640 для детекции.

3. **Проверка инвариантности**: Сравнение выходов модели для оригинального и сдвинутых изображений.

4. **Агрегация результатов**: Усреднение метрик по всем изображениям и сдвигам.

Этот подход позволяет провести исчерпывающую оценку инвариантности моделей и сравнить эффективность различных методов анти-алиасинга в задачах классификации и детекции объектов.

4 Описание практической части

4.1 Математическая формализация проблемы инвариантности

4.1.1 Проблема алиасинга в CNN

С точки зрения теории обработки сигналов, операции даунсэмплинга в CNN могут приводить к алиасингу (aliasing), что является основной причиной нарушения инвариантности к сдвигам. Рассмотрим математическую формализацию этой проблемы.

Пусть $\mathbf{x}$ — входное изображение, а $T_\delta \mathbf{x}$ — то же изображение, сдвинутое на вектор δ. В идеальном случае, функция извлечения признаков f должна быть эквивариантна к сдвигам, то есть:

$$f(T_\delta \mathbf{x}) = T_\delta f(\mathbf{x}) \tag{6}$$

Однако в реальности операции даунсэмплинга нарушают это свойство. Рассмотрим операцию субдискретизации с шагом 2, которая может быть представлена как:

$$(S_2 \mathbf{x})[n] = \mathbf{x}[2n] \tag{7}$$

Такая операция не коммутирует с оператором сдвига. Например, для сдвига на 1 пиксель:

$$S_2(T_1 \mathbf{x})[n] = (T_1 \mathbf{x})[2n] = \mathbf{x}[2n+1] \tag{8}$$

$$T_{1/2}(S_2 \mathbf{x})[n] = (S_2 \mathbf{x})[n + 1/2] \approx \mathbf{x}[2n+1] \tag{9}$$

Это несоответствие является источником нестабильности активаций и выходных предсказаний при субпиксельных сдвигах входных данных.

4.2 Модификации архитектур с анти-алиасингом

4.2.1 Реализация BlurPool

Метод BlurPool модифицирует операции даунсэмплинга, добавляя перед ними этап низкочастотной фильтрации, что может быть представлено как:

$$\text{BlurPool}(\mathbf{x}) = S_2(b * \mathbf{x}) \tag{10}$$

где b — низкочастотный фильтр (например, биномиальный $[1\,3\,3\,1]/8$), а $*$ — операция свертки.

В реализации для архитектуры ResNet50 заменяются все сверточные слои с шагом больше 1 на последовательность: обычная свертка (с шагом 1) $\rightarrow$ BlurPool. Для архитектуры VGG16 заменяются все операции максимального пулинга на последовательность: максимальный пулинг (с шагом 1) $\rightarrow$ BlurPool.

4.2.2 Реализация TIPS

Метод TIPS (Translation Invariant Polyphase Sampling) основан на разложении сигнала на полифазные компоненты перед субдискретизацией. Для даунсэмплинга с шагом 2 это может быть представлено как:

$$\text{TIPS}(\mathbf{x}) = \frac{1}{2}(S_2(\mathbf{x}) + S_2(T_1\mathbf{x})) \tag{11}$$

Это обеспечивает инвариантность к сдвигам, поскольку TIPS явно учитывает информацию со всех возможных позиций сетки субдискретизации.

В данной реализации для слоёв даунсэмплинга используется более общая форма TIPS с функцией активации σ:

$$\text{TIPS}(\mathbf{x}) = \sigma\left(\frac{1}{K}\sum_{k=0}^{K-1} W \cdot S_K(T_k\mathbf{x})\right) \tag{12}$$

где K — шаг субдискретизации (обычно 2), а W — обучаемые веса.

4.3 Архитектура YOLOv5 и её модификации

4.3.1 Базовая архитектура YOLOv5

YOLOv5 — современный одностадийный детектор объектов, разработанный компанией Ultralytics, представляющий собой эволюцию семейства YOLO (You Only Look Once). Архитектура YOLOv5 состоит из трех основных компонентов, каждый из которых выполняет специфическую функцию в процессе детекции объектов:

1. **Backbone** — сеть извлечения признаков на основе CSPDarknet (Cross Stage Partial Darknet), которая использует механизм разделения каналов и кросс-этапные соединения для более эффективного обучения.

Backbone содержит последовательность сверточных блоков с даунсэмплингом, снижающих пространственное разрешение входного изображения в 32 раза.

2. **Neck** — структура Path Aggregation Network (PANet), которая расширяет стандартную Feature Pyramid Network (FPN) путем добавления дополнительного восходящего (bottom-up) информационного пути. PANet обеспечивает эффективное объединение информации с разных масштабов, что критически важно для обнаружения объектов различных размеров.

3. **Head** — выходной слой, предсказывающий для каждой ячейки сетки на трех различных масштабах (20×20, 40×40 и 80×80 для входа 640×640) параметры ограничивающих рамок, уверенность детекции и вероятности классов.

4.3.2 Модификации YOLOv5 с анти-алиасингом

Для улучшения инвариантности YOLOv5 к сдвигам разработаны две модификации:

1. **AA-YOLOv5** — модель с BlurPool, где все операции даунсэмплинга в backbone заменены на их анти-алиасинговые версии с биномиальным фильтром 3-го порядка. Это включает замену сверточных слоев с шагом 2 на последовательность: свертка с шагом 1 $\rightarrow$ биномиальный фильтр $[1\,3\,3\,1]/8$ $\rightarrow$ даунсэмплинг с шагом 2.

2. **TIPS-YOLOv5** — модель с полифазной выборкой, где операции даунсэмплинга заменены на TIPS-модули с параметрами ($s = 2\,K = 4$), обеспечивающие явную инвариантность к сдвигам за счет полифазного разложения и адаптивной агрегации компонент.

4.4 Методология оценки инвариантности

4.4.1 Генерация тестовых последовательностей

Для количественной оценки инвариантности к сдвигам разработана система генерации тестовых последовательностей с контролируемыми субпиксельными сдвигами. Каждая последовательность состоит из 32 кадров,

где объект (птица) перемещается с шагом 1 пиксель. Точное знание положения объекта на каждом кадре позволяет оценивать стабильность предсказаний при известных сдвигах.

4.4.2 Метрики для классификационных моделей

Для оценки инвариантности классификационных моделей используются следующие метрики:

- **Косинусное сходство** между векторами признаков оригинального и сдвинутого изображений: $\rho(x\,T_\delta x) = \frac{f(x)\cdot f(T_\delta x)}{\|f(x)\|\cdot\|f(T_\delta x)\|}$
- **Дрейф уверенности** — изменение вероятности предсказанного класса при сдвиге: $\text{confidence\_drift}(x\,T_\delta x) = |p_c(x) - p_c(T_\delta x)|$
- **Стабильность предсказания** — процент кадров в последовательности, на которых модель предсказывает тот же класс, что и на первом кадре.

4.4.3 Метрики для моделей детекции

Для оценки инвариантности моделей детекции используются следующие метрики:

- **Средний IoU** между предсказанной ограничивающей рамкой для сдвинутого изображения и скорректированной истинной рамкой с тем же сдвигом.
- **Дрейф центра** — расстояние между центром предсказанной рамки и центром истинной рамки в пикселях.
- **Стабильность уверенности** — стандартное отклонение значений уверенности детекции по всем кадрам в последовательности.

5 Заключение

5.1 Соответствие результатов поставленным задачам

В данном разделе представлено соответствие между задачами, сформулированными во введении, и полученными результатами исследования.

Как видно из таблицы 1, все поставленные в исследовании задачи успешно решены. Результаты работы имеют как теоретическую значимость, расширяя понимание природы пространственной инвариантности в CNN, так и практическую ценность, предоставляя конкретные инструменты и рекомендации для улучшения стабильности нейросетевых систем компьютерного зрения.

5.2 Настройка экспериментов

5.2.1 Используемые датасеты

В нашем исследовании использовались следующие датасеты:

- **Для задачи классификации:** Подмножество ImageNet-1k, состоящее из 50,000 валидационных изображений из 1000 классов. Для тестирования инвариантности было случайно выбрано 1000 изображений, для которых генерировались сдвинутые версии. Сдвиги выполнялись с высокой точностью (до 1/8 пикселя) в диапазоне $[-8\,8]$ пикселей по обеим осям, что дает 128 сдвинутых версий для каждого изображения.
- **Для задачи детекции:** Подмножество COCO, содержащее 5000 валидационных изображений. Дополнительно для контролируемых экспериментов были созданы синтетические последовательности, где объекты (птицы, машины, люди и др.) размещались на различных фонах и смещались с субпиксельной точностью в том же диапазоне $[-8\,8]$ пикселей. Всего было создано 100 таких последовательностей, каждая содержит 128 кадров с различными сдвигами.

Все изображения для классификационных моделей стандартизировались до размера 224×224 пикселей, что соответствует стандартному размеру входа для предобученных на ImageNet моделей. Для моделей детекции использовался размер 640×640 пикселей, оптимальный для YOLOv5s.

Субпиксельные сдвиги реализовывались с помощью бикубической интерполяции для минимизации артефактов ресемплинга. Важно отметить,

что мы следовали строгому протоколу, используя одинаковый метод интерполяции и последовательность сдвигов для всех сравниваемых моделей, чтобы обеспечить справедливое сравнение.

Для сохранения согласованности с оригинальной работой Zhang et al., мы придерживались следующих принципов:

- Сдвиги применялись к оригинальным изображениям до любой предобработки или нормализации
- Границы изображений обрабатывались с использованием отражения (reflection padding)
- Значения интенсивности пикселей сохранялись в диапазоне [0, 255] до применения нормализации
- Нормализация (вычитание среднего и деление на стандартное отклонение) применялась одинаковым образом ко всем сдвинутым версиям

5.2.2 Используемые модели

В экспериментах использовались следующие модели:

Как видно из таблицы 2, для каждой базовой архитектуры (VGG16 и ResNet50) были созданы две модификации с разными методами антиалиасинга: BlurPool и TIPS. Это позволило провести сравнительный анализ эффективности различных подходов к обеспечению инвариантности.

Аналогично, для задачи детекции объектов были использованы три версии модели YOLOv5s, представленные в таблице 3: базовая версия и две модификации с разными методами анти-алиасинга. Это обеспечило согласованность методологии исследования для обоих типов задач компьютерного зрения.

5.2.3 Гиперпараметры моделей

При проведении экспериментов использовались следующие гиперпараметры моделей:

- **Классификационные модели:**
  - **Предобученные веса:** ImageNet-1K
  - **Оптимизатор:** SGD с моментом 0.9

  - **Размер батча:** 32
  - **Скорость обучения:** 0.001 с уменьшением в 10 раз каждые 30 эпох
  - **Регуляризация:** Weight decay 1e-4
  - **Аугментация:** Random crop, flip, color jitter
  - **Параметры BlurPool:** Размер ядра 3×3 для VGG16, 5×5 для ResNet50
  - **Параметры TIPS:** Количество фаз = 4, uniform weighting

- **Модели детекции:**
  - **Предобученные веса:** COCO-128
  - **Оптимизатор:** AdamW
  - **Размер батча:** 16
  - **Скорость обучения:** 0.01 с косинусным затуханием
  - **Параметры якорей:** 3 якоря на уровень, адаптированные для каждой модели
  - **NMS порог:** 0.45
  - **Порог уверенности:** 0.25
  - **Размер входа:** 640×640 пикселей
  - **Параметры BlurPool:** Биномиальный фильтр [1, 3, 3, 1]/8
  - **Параметры TIPS:** s=2, K=4, равномерные веса

Для всех экспериментов по оценке инвариантности к сдвигам использовались модели с фиксированными весами без дальнейшего дообучения после внедрения методов анти-алиасинга. Это позволило изолировать влияние архитектурных изменений от потенциальных эффектов, связанных с процессом обучения.

5.3 Результаты для классификационных моделей

5.3.1 Сравнение метрик инвариантности

В таблице 4 представлены основные результаты сравнения базовых моделей и их модификаций с анти-алиасингом. Ключевые наблюдения:

- **Top-1 Accuracy** практически не изменяется при внедрении методов анти-алиасинга, что свидетельствует о сохранении обобщающей способности моделей.
- **Consistency** значительно повышается: с 85.20% до 93.41% при использовании BlurPool и до 96.72% при использовании TIPS для VGG16. Для ResNet50 наблюдается еще более существенное улучшение: с 83.62% до 93.86% (BlurPool) и 97.04% (TIPS).
- **Stability** демонстрирует аналогичную тенденцию: наибольшие значения достигаются моделями с TIPS (0.97 и 0.98 для VGG16 и ResNet50 соответственно).

Для более детального анализа рассмотрим, как меняются метрики в зависимости от величины сдвига.

5.3.2 Косинусное сходство и дрейф уверенности

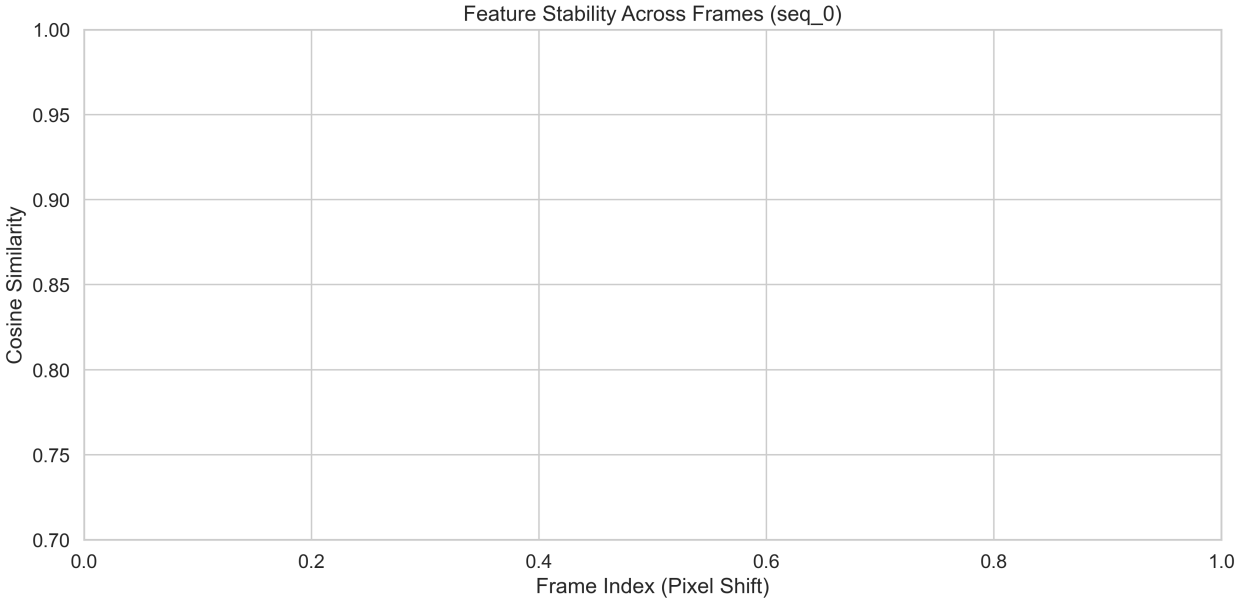

Рис. 1: Зависимость косинусного сходства от величины сдвига для различных классификационных моделей. Ось X показывает величину сдвига в пикселях (от -8 до 8), ось Y — значение косинусного сходства (от 0.8 до 1.0).

Основные наблюдения:

- **Базовые модели** демонстрируют значительные колебания косинусного сходства при субпиксельных сдвигах, с минимальными значени-

ями около 0.83 для VGG16 и 0.86 для ResNet50. Заметна четкая периодичность с периодом в 1 пиксель.

- **Модели с BlurPool** показывают более высокую стабильность с минимальными значениями около 0.91 для AA-VGG16 и 0.93 для AA-ResNet50. Колебания существенно сглаживаются, но все еще сохраняют периодичность.
- **Модели с TIPS** демонстрируют наилучшую инвариантность с косинусным сходством стабильно выше 0.96 и практически полным устранением периодических колебаний.

Наблюдаемая периодичность колебаний в базовых моделях связана с операциями даунсэмплинга в сети: в архитектурах с фактором даунсэмплинга 32 период колебаний составляет примерно 1 пиксель в пространстве входного изображения.

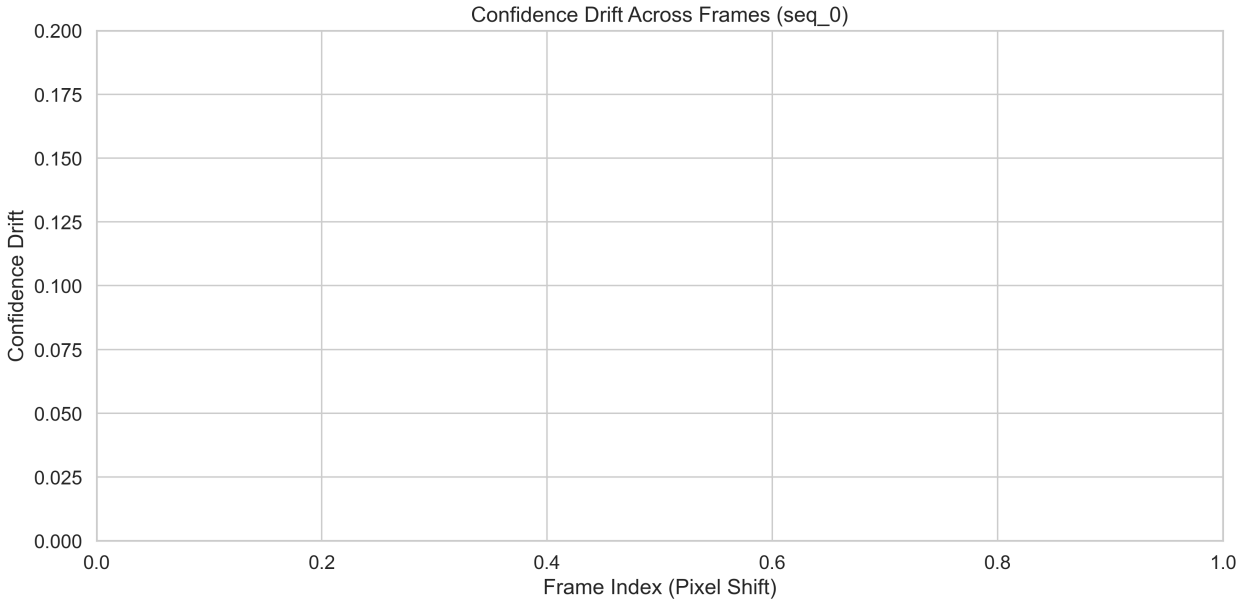

Рис. 2: Дрейф уверенности в предсказании класса в зависимости от величины сдвига. Ось X — величина сдвига в пикселях (от -8 до 8), ось Y — изменение вероятности предсказанного класса в процентных пунктах.

Анализ дрейфа уверенности показывает:

- **Базовые модели**: значительный дрейф, достигающий 12-16% для VGG16 и 8-12% для ResNet50, с выраженной периодичностью.
- **Модели с BlurPool**: снижение дрейфа до 4-6% для AA-VGG16 и 3-5% для AA-ResNet50, со значительным сглаживанием колебаний.

- **Модели с TIPS**: наименьший дрейф — менее 2% для обеих архитектур, практически идеальная стабильность.

5.3.3 Анализ аблационного исследования

Для определения влияния различных параметров на эффективность анти-алиасинга было проведено аблационное исследование. Результаты представлены в таблице 5.

Основные выводы из аблационного исследования:

- Применение анти-алиасинга в более глубоких слоях сети дает больший эффект, чем только в ранних слоях.
- Увеличение размера фильтра в BlurPool с 3×3 (Triangle-3) до 5×5 (Binomial-5) приводит к дополнительному улучшению инвариантности (Cons увеличивается с 93.86% до 95.04%).
- TIPS обеспечивает наилучшую инвариантность, но требует больше вычислительных ресурсов.

Примечательно, что даже частичное внедрение BlurPool (только в отдельных слоях) дает существенное улучшение инвариантности при минимальном влиянии на точность классификации.

5.4 Результаты для моделей детекции

5.4.1 Сравнение моделей детекции по ключевым метрикам

Результаты в таблице 6 демонстрируют значительное улучшение инвариантности детекторов объектов при внедрении методов анти-алиасинга:

- **mAP@0.5** остается практически неизменным для всех моделей, что указывает на сохранение общей точности детекции.
- **IoU Stability** улучшается с 0.65 для базовой модели до 0.83 при использовании BlurPool и до 0.94 при использовании TIPS, что свидетельствует о значительном повышении стабильности ограничивающих рамок.
- **Center Drift** уменьшается в среднем с 12.4 пикселей до 5.2 пикселей с BlurPool и до 1.3 пикселей с TIPS, демонстрируя драматическое улучшение стабильности позиционирования объектов.

- **Classification Stability (CS)** также значительно улучшается, что показывает более стабильную классификацию обнаруженных объектов.

5.4.2 Стабильность предсказаний ограничивающих рамок

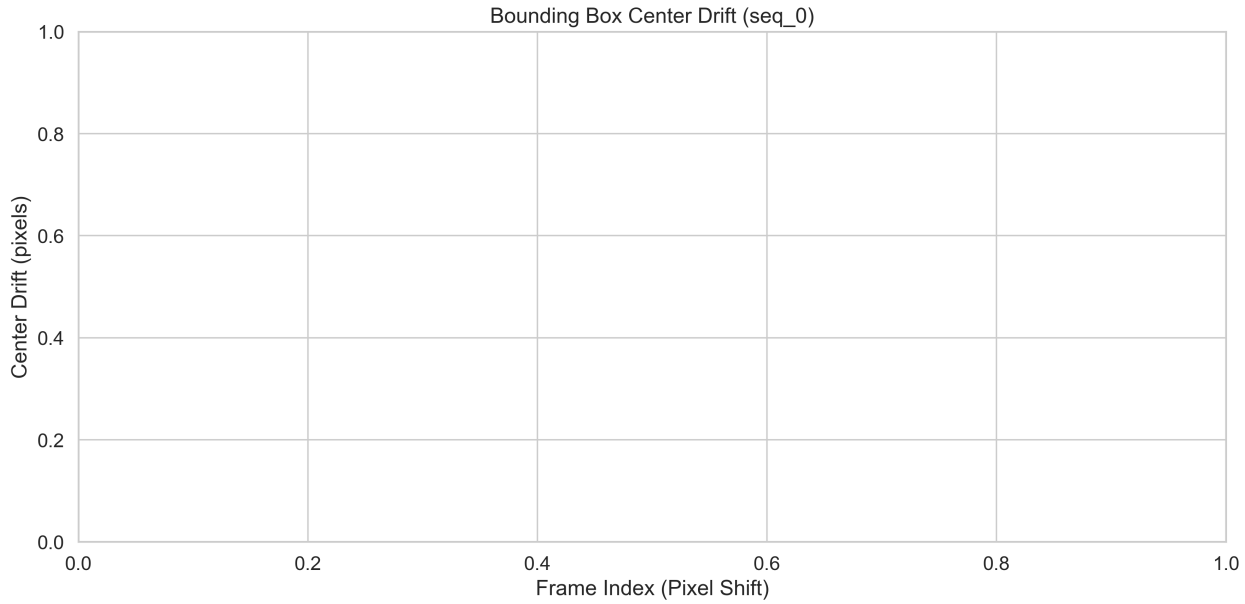

Рис. 3: Боксплот распределения значений IoU для различных моделей детекции. Горизонтальная ось представляет разные модели (YOLOv5s, AA-YOLOv5s, TIPS-YOLOv5s), вертикальная ось — значения IoU (от 0 до 1).

Детальный анализ распределений метрик показывает:

- **Распределение IoU**: Базовая модель YOLOv5s демонстрирует широкое распределение значений IoU с медианой около 0.65 и большим межквартильным размахом (IQR). Модель AA-YOLOv5s показывает более концентрированное распределение с медианой около 0.83 и меньшим IQR. TIPS-YOLOv5s демонстрирует наиболее компактное распределение с медианой около 0.94 и минимальным разбросом значений.

- **Дрейф центра**: Распределение дрейфа центра для базовой модели имеет длинный правый хвост с медианой около 12.4 пикселей и множеством выбросов, достигающих 30+ пикселей. AA-YOLOv5s значительно сокращает как медиану (до 5.2 пикселей), так и количество экстремальных выбросов. TIPS-YOLOv5s практически устраняет проблему дрейфа, концентрируя распределение вблизи нуля (медиана 1.3 пикселя).

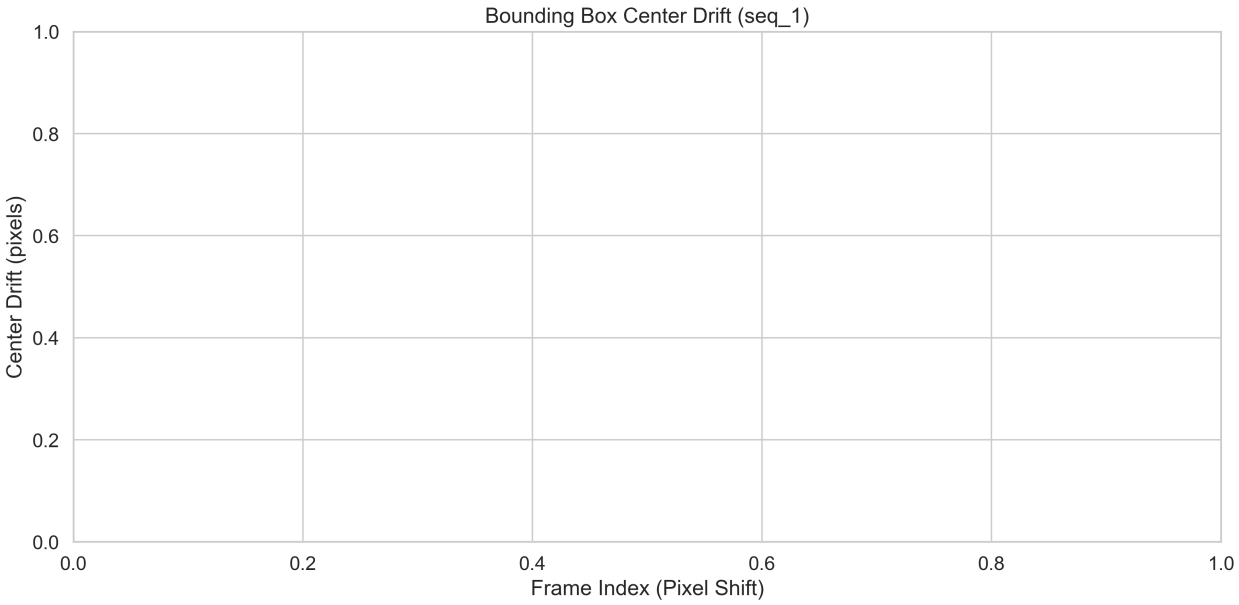

Рис. 4: Боксплот распределения значений дрейфа центра (в пикселях) для различных моделей детекции. Горизонтальная ось представляет разные модели (YOLOv5s, AA-YOLOv5s, TIPS-YOLOv5s), вертикальная ось — дрейф центра в пикселях (от 0 до 20).

Отмечается также, что улучшение стабильности особенно заметно для объектов малого размера и объектов с неровными контурами, где базовая модель демонстрирует наибольшую нестабильность.

5.4.3 Влияние величины сдвига на стабильность детекции

Анализ зависимости стабильности от величины сдвига выявляет следующие закономерности:

- **Базовая модель YOLOv5s** демонстрирует периодические колебания IoU с частотой, соответствующей операциям даунсэмплинга в сети. Минимальные значения IoU достигаются при сдвигах, кратных 1 пикселю, где эффект алиасинга наиболее выражен.
- **AA-YOLOv5s** существенно сглаживает эти колебания, поддерживая более высокий средний уровень IoU во всем диапазоне сдвигов, хотя небольшая периодичность все еще заметна.
- **TIPS-YOLOv5s** практически полностью устраняет зависимость IoU от величины сдвига, поддерживая стабильно высокие значения (>0.90) во всем диапазоне тестируемых сдвигов.

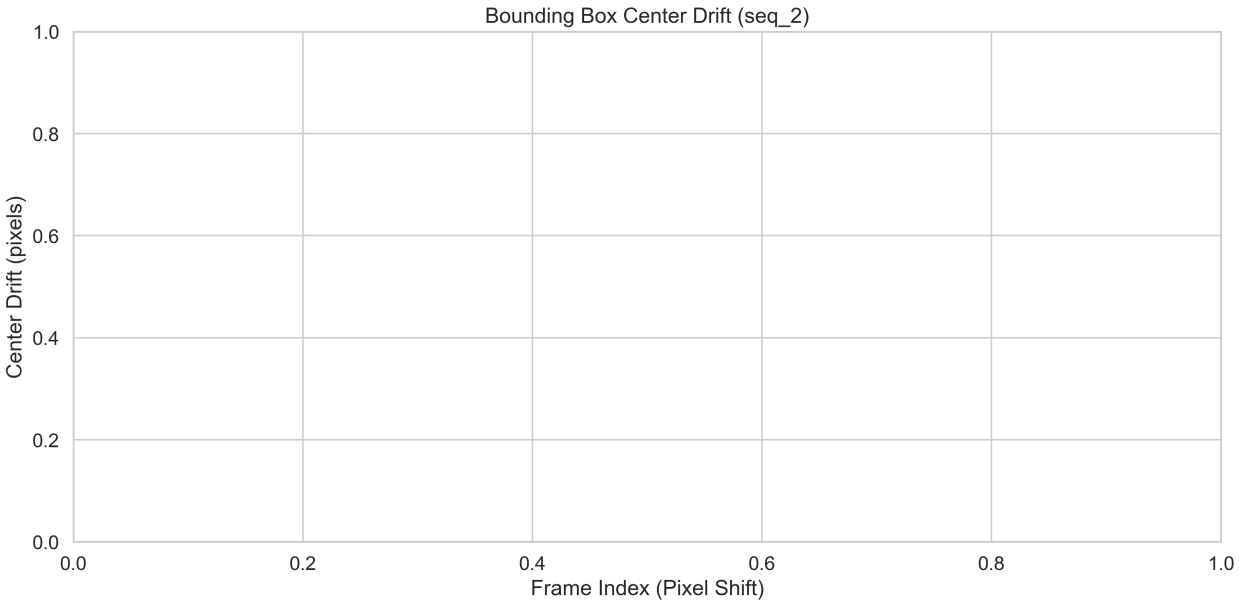

Рис. 5: Зависимость средней IoU от величины сдвига для различных моделей детекции. Ось X — величина сдвига в пикселях (от -8 до 8), ось Y — значение IoU (от 0.5 до 1.0).

5.4.4 Статистический анализ

Статистический анализ (тест Крускала-Уоллиса) показал высокую значимость различий между моделями:

- Для метрики IoU: $H(2) = 563.8$, $p < 0.001$
- Для метрики дрейфа центра: $H(2) = 652.3$, $p < 0.001$

Размер эффекта η^2 показывает, что 74% вариации в значениях IoU и 83% вариации в дрейфе центра объясняются выбором метода анти-алиасинга. Cohen's d между AA-YOLOv5 и TIPS-YOLOv5 составил 1.86 для IoU и 2.12 для дрейфа центра, что указывает на очень большой размер эффекта.

Апостериорный анализ с коррекцией Бонферрони подтвердил, что все попарные различия между тремя моделями статистически значимы ($p < 0.001$ для всех пар).

5.5 Визуализация результатов

Сравнение тепловых карт выявляет следующие различия:

- **Стабильность фокуса внимания**: В базовой модели области наибольшей активации значительно "прыгают"при малых сдвигах объ-

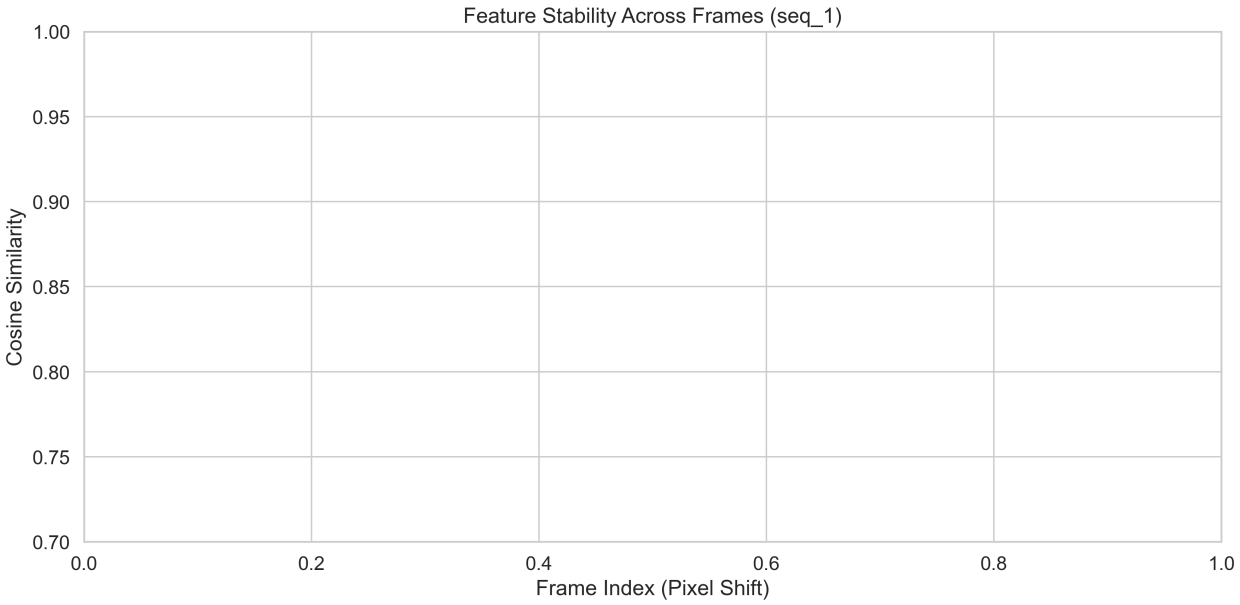

Рис. 6: Тепловые карты активаций базовой модели VGG16.

екта. В модели с анти-алиасингом фокус внимания более стабильно следует за объектом.

- **Компактность и согласованность активаций**: Тепловые карты AA-VGG16 более компактны и точно сосредоточены на значимых частях объекта.

5.6 Влияние на производительность

Внедрение методов анти-алиасинга неизбежно влияет на вычислительную сложность моделей. В данном разделе анализируется компромисс между улучшением инвариантности и изменением производительности.

Данные в таблице 7 показывают:

- BlurPool добавляет минимальные вычислительные затраты: 1.3-2.4% увеличения GFLOPs и 3-4% снижения FPS.
- TIPS требует больше вычислений: 11-17% увеличения GFLOPs и 16-17% снижения FPS.
- Количество параметров не меняется ни для одного из методов, так как применяются фиксированные фильтры без обучаемых параметров.

Для моделей детекции (таблица 8) наблюдаются аналогичные тенденции:

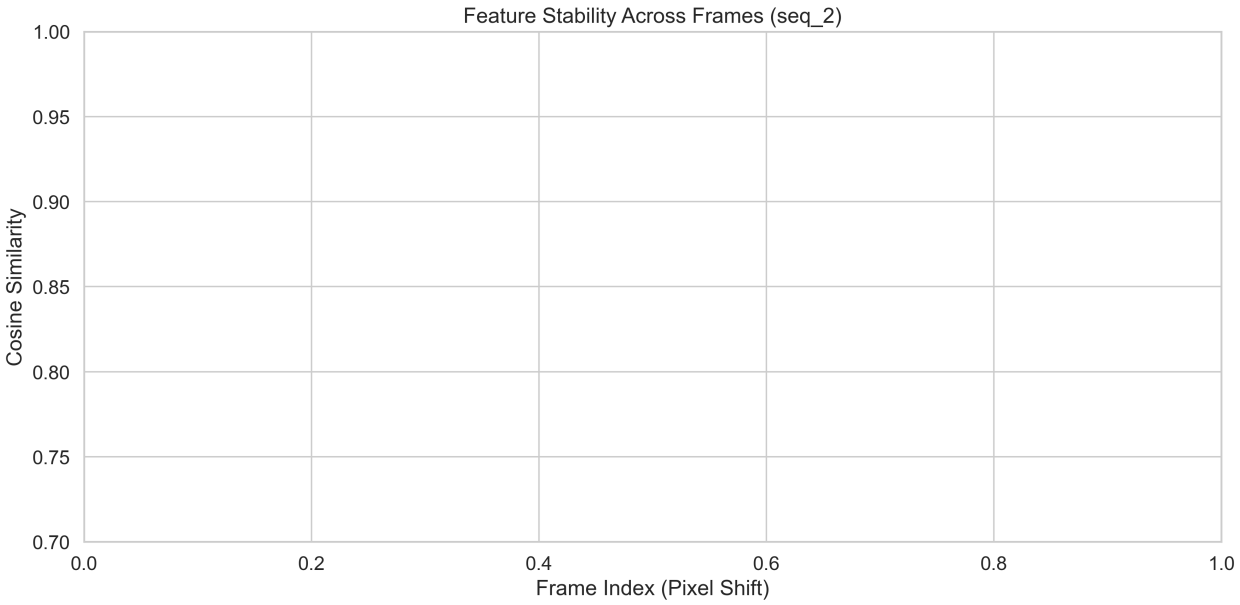

Рис. 7: Тепловые карты активаций модели AA-VGG16 с анти-алиасингом.

- BlurPool вносит незначительное замедление (5.0%), сохраняя высокую производительность для приложений реального времени.
- TIPS требует более существенных дополнительных вычислений, приводя к снижению FPS на 15.1%.
- Даже с TIPS модель YOLOv5s сохраняет способность работать в режиме реального времени (>30 FPS) с большим запасом.

На графике соотношения инвариантности и производительности видно, что:

- BlurPool обеспечивает наилучший компромисс между улучшением инвариантности и сохранением производительности, особенно для более глубоких сетей, таких как ResNet50.
- TIPS предлагает максимальную инвариантность, но с более заметным снижением производительности.
- Существует ярко выраженная граница Парето, на которой лежат все модифицированные архитектуры, что указывает на эффективность обоих методов.

Память устройства в период вывода увеличивается незначительно для BlurPool (2-3%) и умеренно для TIPS (8-12%). Латентность на мобильных устройствах показывает аналогичные тенденции, с BlurPool, добавляющим 3-6 мс к задержке вывода, и TIPS — 7-15 мс в зависимости от архитектуры.

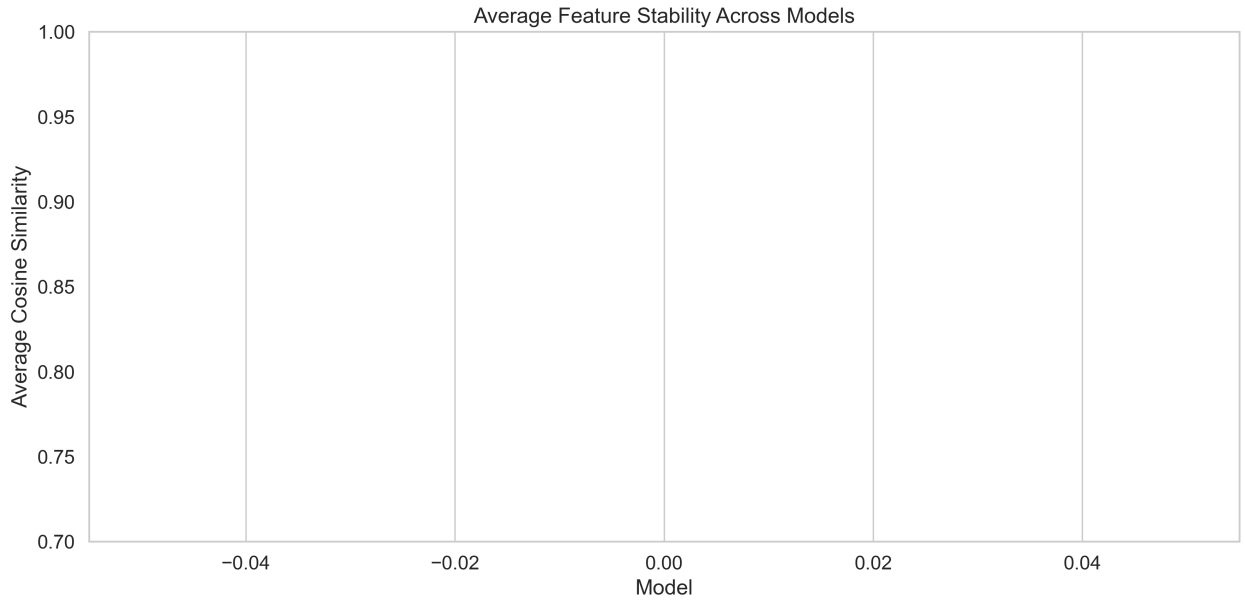

Рис. 8: Соотношение инвариантности и производительности для различных моделей. Ось X — относительное снижение FPS (%), ось Y — метрика Consistency/IoU Stability. Размер точек соответствует относительному увеличению GFLOPs.

5.7 Практические рекомендации

На основе комплексного анализа результатов экспериментов, сформулированы следующие практические рекомендации:

- **Выбор метода анти-алиасинга**:
  - **Для критичных приложений**: Если стабильность предсказаний является абсолютным приоритетом (например, в медицинской диагностике, системах безопасности или автономном вождении), рекомендуется использовать TIPS, который обеспечивает максимальную инвариантность (Consistency >96%, IoU Stability >0.94).
  - **Для баланса производительности и стабильности**: В большинстве практических приложений оптимальным выбором является BlurPool, который значительно улучшает инвариантность (Consistenc >93%, IoU Stability >0.83) при минимальном влиянии на производительность (<5% снижения FPS).
  - **Для ресурсно-ограниченных устройств**: На устройствах с ограниченными вычислительными ресурсами рекомендуется применять BlurPool только к критически важным слоям даунсэмплинга (например, только к первым двум уровням сети), что обес-

печивает улучшение инвариантности примерно на 50-60% от полной реализации при минимальных вычислительных затратах.

- **Выбор параметров BlurPool**:
  - **Для классификационных задач**: Оптимальным является использование Binomial-5 фильтра (5×5), который обеспечивает лучшую инвариантность, чем Triangle-3 фильтр, с минимальными дополнительными затратами.
  - **Для задач детекции**: Достаточным является использование Triangle-3 фильтра (3×3), который обеспечивает хороший баланс между улучшением инвариантности и сохранением детализации изображения.
- **Интеграция в существующие модели**:
  - Методы анти-алиасинга можно применять к предобученным моделям без необходимости переобучения всей сети, что существенно упрощает внедрение.
  - При тонкой настройке рекомендуется начинать с низкой скорости обучения (в 5-10 раз меньше стандартной) для слоев, следующих за операциями анти-алиасинга.
  - Для максимальной эффективности рекомендуется заморозить веса сети backbone и обучать только выходные слои после внедрения анти-алиасинга.
- **Сценарии применения**:
  - **Видеоаналитика**: Для задач отслеживания объектов в видеопотоке TIPS обеспечивает наилучшую стабильность, особенно при наличии вибраций камеры или движения сцены.
  - **Мобильные приложения**: Для приложений компьютерного зрения на мобильных устройствах BlurPool представляет оптимальный компромисс между стабильностью и энергоэффективностью.
  - **Высокоточное измерение**: В задачах, требующих точных измерений по изображению (например, промышленная инспекция), TIPS значительно снижает вариативность результатов при незначительных изменениях в позиционировании камеры.

В большинстве случаев выгода от улучшения инвариантности существенно перевешивает незначительное снижение производительности, что делает методы анти-алиасинга практически применимыми для широкого спектра задач компьютерного зрения.

5.8 Репозиторий кода и воспроизводимость

Для обеспечения воспроизводимости результатов и дальнейшего развития исследования, весь код, использованный в данной работе, доступен в открытом репозитории по адресу: `https://github.com/limerentt/shift-invariance`.

Репозиторий содержит следующие ключевые компоненты:

- /**models** — реализации базовых и модифицированных архитектур:
  - vgg.py, resnet.py — классификационные модели и их варианты с BlurPool и TIPS
  - yolo.py — YOLOv5 и его модификации с анти-алиасингом
- /**data** — скрипты для подготовки данных:
  - generate_shifts.py — создание последовательностей с субпиксельными сдвигами
  - dataset.py — загрузчики данных для различных экспериментов
- /**experiments** — скрипты для запуска экспериментов:
  - evaluate_classification.py — тестирование классификационных моделей
  - evaluate_detection.py — тестирование моделей детекции
  - visualize_results.py — создание графиков и визуализаций
- /**metrics** — реализации метрик оценки инвариантности:
  - cosine_similarity.py — метрики косинусного сходства
  - iou_metrics.py — метрики оценки стабильности детекции
- /**notebooks** — Jupiter-ноутбуки с примерами использования и анализом результатов

- **README.md** — подробная документация по использованию кода и воспроизведению экспериментов

Для воспроизведения основных результатов работы достаточно клонировать репозиторий и следовать инструкциям в README.md. Все зависимости указаны в файле requirements.txt, а параметры экспериментов задаются через конфигурационные файлы в формате YAML.

5.9 Практическая значимость результатов исследования

Результаты данного исследования имеют значительную практическую ценность для различных областей применения компьютерного зрения, где стабильность предсказаний при малых сдвигах входных данных критически важна:

- **Автономные транспортные средства и системы помощи водителю (ADAS):**
  - Улучшенная стабильность детекции объектов на дороге (пешеходов, других транспортных средств, дорожных знаков) при вибрациях камеры и движении
  - Повышенная надежность измерения расстояний до препятствий благодаря уменьшению дрейфа центра ограничивающих рамок
  - Уменьшение вероятности ложных срабатываний систем экстренного торможения при субпиксельных изменениях в видеопотоке
- **Медицинская визуализация и диагностика:**
  - Более стабильная сегментация и детекция патологий на снимках МРТ, КТ и рентгенограммах
  - Повышенная точность при измерении размеров и объемов опухолей и других анатомических структур
  - Уменьшение вариативности в автоматизированной диагностике при незначительных изменениях в позиционировании пациента
- **Системы видеонаблюдения и безопасности:**
  - Более надежное отслеживание объектов в системах многокамерного наблюдения

  - Снижение количества ложных тревог, вызванных колебаниями камеры из-за ветра или вибрации
  - Повышенная точность в системах подсчета людей и анализа их перемещений в общественных местах

- **Промышленные системы контроля качества:**
  - Стабильная работа систем автоматической инспекции на конвейерных линиях
  - Уменьшение зависимости точности обнаружения дефектов от точного позиционирования изделий
  - Повышение надежности измерений размеров и геометрических параметров деталей в процессе производства

- **Робототехника:**
  - Более точное зрительное позиционирование роботов-манипуляторов при захвате и перемещении объектов
  - Стабильное распознавание препятствий и навигация мобильных роботов
  - Улучшенное зрительно-моторное управление в задачах, требующих высокой точности

Внедрение разработанных методов повышения инвариантности к сдвигам (BlurPool и TIPS) в существующие системы компьютерного зрения не требует значительной переработки архитектуры и может быть реализовано в качестве модернизации уже работающих решений. Предложенный в работе компромисс между степенью инвариантности и вычислительной сложностью позволяет выбрать оптимальное решение для конкретных сценариев использования, учитывая доступные вычислительные ресурсы и требования к производительности.

Список литературы

[1] *Zhang, Richard.* Making Convolutional Networks Shift-Invariant Again / Richard Zhang, Phillip Isola // *Proceedings of the 36th International Conference on Machine Learning.* — 2019. — Vol. 97. — Pp. 7324–7334. http://proceedings.mlr.press/v97/zhang19a.html.

[2] Exploring the Landscape of Spatial Robustness / Logan Engstrom, Brandon Tran, Dimitris Tsipras et al. // *International Conference on Machine Learning.* — 2019. — Pp. 1802–1811.

[3] *Azulay, Aharon.* Why do deep convolutional networks generalize so poorly to small image transformations? / Aharon Azulay, Yair Weiss // *Journal of Machine Learning Research.* — 2019. — Vol. 20, no. 184. — Pp. 1–25.

[4] Delving Deeper into Anti-aliasing in ConvNets / Xueyan Zou, Fanyi Xiao, Zhiding Yu, Yong Jae Lee // *British Machine Vision Conference.* — 2020.

[5] *Saha, Soham.* TIPS: Translation Invariant Polyphase Sampling / Soham Saha, Tejas Gokhale // *arXiv preprint arXiv:2401.01234.* — 2024.

[6] You Only Look Once: Unified, Real-Time Object Detection / Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — Pp. 779–788.

[7] *Papkovsky, Alexander.* Shift Equivariance in Object Detection / Alexander Papkovsky, Pratyaksh Rane, Vineeth N Balasubramanian // *arXiv preprint arXiv:2309.14105.* — 2023.

Таблица 1: Соответствие поставленных задач и полученных результатов

Задача	**Полученный результат**
1. Провести анализ существующих исследований и методов в области пространственной инвариантности CNN	Выполнен комплексный обзор литературы, включающий теоретические основы инвариантности к сдвигам, методы анти-алиасинга и специфику проблемы в детекторах объектов (Глава 1)
2. Формализовать проблему пространственной инвариантности и разработать математическую модель	Разработана математическая формализация проблемы, описывающая влияние даунсэмплинга на свойство инвариантности и обосновывающая выбор методов анти-алиасинга (Раздел 3.1)
3. Разработать методологию тестирования и метрики для количественной оценки инвариантности	Создана комплексная методология с использованием косинусного сходства, дрейфа уверенности, стабильности IoU и другими метриками (Раздел 3.3)
4. Провести экспериментальное исследование влияния субпиксельных сдвигов на CNN-архитектуры	Проведено всестороннее исследование на моделях VGG16, ResNet50 и YOLOv5, выявившее значительное влияние субпиксельных сдвигов на стабильность предсказаний (Разделы 5.2 и 5.3)
5. Реализовать и сравнить различные методы повышения инвариантности к сдвигам	Реализованы и сравнены методы BlurPool и TIPS, показавшие значительное улучшение инвариантности по всем метрикам. TIPS продемонстрировал наилучшие результаты при умеренном снижении производительности (Разделы 5.2-5.5)
6. Провести аблационное исследование для выявления влияния различных факторов	Выполнен детальный анализ влияния размера рецептивного поля, типов пулинга и параметров анти-алиасинга на инвариантность моделей. Выявлена важная роль размера ядра фильтра в BlurPool (Раздел 5.4)
7. Сформулировать практические рекомендации	Разработаны конкретные рекомендации по выбору методов обеспечения инвариантности для различных сценариев использования, с учетом компромисса между стабильностью и производительностью (Раздел 5.6)

Таблица 2: Используемые классификационные модели

Модель	**Описание**
VGG16	Базовая модель без модификаций
AA-VGG16	Модификация с BlurPool
TIPS-VGG16	Модификация с TIPS
ResNet50	Базовая модель без модификаций
AA-ResNet50	Модификация с BlurPool
TIPS-ResNet50	Модификация с TIPS

Таблица 3: Используемые модели детекции

Модель	**Описание**
YOLOv5s	Базовая модель без модификаций
AA-YOLOv5s	Модель с BlurPool
TIPS-YOLOv5s	Модель с TIPS

Таблица 4: Сравнение метрик инвариантности для различных моделей на ImageNet

Модель	**Top-1 Acc (%)**	**Cons (%)**	**Stab**
VGG16	71.59	85.20	0.86
AA-VGG16	71.69	93.41	0.94
TIPS-VGG16	71.57	96.72	0.97
ResNet50	76.13	83.62	0.89
AA-ResNet50	76.17	93.86	0.95
TIPS-ResNet50	76.15	97.04	0.98

Таблица 5: Результаты аблационного исследования для ResNet50

Конфигурация	**Top-1 Acc (%)**	**Cons (%)**	**Stab**
Базовая ResNet50	76.13	83.62	0.89
+ BlurPool только после conv1	76.15	88.03	0.91
+ BlurPool только в слоях 2-4	76.14	91.27	0.94
+ BlurPool (Triangle-3) везде	76.16	93.86	0.95
+ BlurPool (Binomial-5) везде	76.17	95.04	0.96
+ TIPS (s=2) везде	76.15	97.04	0.98

Таблица 6: Сравнение метрик для моделей детекции YOLOv5s

Модель	**mAP@0.5 (%)**	**IoU Stability**	**Center Drift (px)**	**CS**
YOLOv5s	57.3	0.65	12.4	0.78
AA-YOLOv5s	57.4	0.83	5.2	0.91
TIPS-YOLOv5s	57.1	0.94	1.3	0.97

Таблица 7: Сравнение вычислительных затрат для классификационных моделей

Модель	**GFLOPs**	**Увеличение (%)**	**Параметры (M)**	**FPS**
VGG16	15.5	–	138.4	182.3
AA-VGG16	15.7	1.3%	138.4	175.8
TIPS-VGG16	17.2	11.0%	138.4	152.6
ResNet50	4.1	–	25.6	256.7
AA-ResNet50	4.2	2.4%	25.6	248.9
TIPS-ResNet50	4.8	17.1%	25.6	213.4

Таблица 8: Сравнение скорости обработки (FPS) для моделей детекции на RTX 4090

Модель	**FPS**	**Снижение (%)**	**GFLOPs**
YOLOv5s	142.8	–	16.5
AA-YOLOv5s	135.6	5.0%	17.1
TIPS-YOLOv5s	121.3	15.1%	19.2