

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	5
1 Введение	6
2 Обзор литературы	12
2.1 Инвариантность к сдвигу в CNN-классификаторах	12
2.1.1 Теоретические основы инвариантности в CNN	12
2.1.2 Эмпирические исследования проблемы	13
2.1.3 Количественные метрики инвариантности	13
2.1.4 Влияние архитектурных особенностей на инвариантность	14
2.2 Методы анти-алиасинга в нейронных сетях	15
2.2.1 Низкочастотная фильтрация и BlurPool	15
2.2.2 Полифазная выборка с инвариантностью к сдвигам (TIPS)	15
2.3 Специфические проблемы инвариантности в детекторах объектов	16
2.3.1 Архитектуры современных детекторов объектов	16
2.3.2 Влияние алиасинга на стабильность детекции	17
2.3.3 Метрики устойчивости детекторов	17
2.3.4 Адаптация методов анти-алиасинга для детекторов	18
2.3.5 Практические последствия нестабильности детекторов	19
3 Исследование и построение решения задачи	20
3.1 Инвариантность к сдвигу в CNN-классификаторах	20
3.1.1 Теоретические основы и эмпирические исследования	20
3.2 Методы анти-алиасинга в нейронных сетях	21
3.2.1 Методы улучшения инвариантности: BlurPool и адаптивные подходы	21
3.2.2 Низкочастотная фильтрация и BlurPool	21
3.2.3 Полифазные методы и их применение в CNN	22
3.3 Специфические проблемы инвариантности в детекторах объектов	22
3.3.1 Архитектуры современных детекторов и влияние алиасинга	22
4 Описание практической части	24
4.1 Математическая формализация проблемы инвариантности	24
4.1.1 Проблема алиасинга в CNN	24
4.2 Модификации архитектур с анти-алиасингом	24
4.2.1 Реализация BlurPool	24
4.2.2 Реализация TIPS	25
4.3 Архитектура YOLOv5 и её модификации	25
4.3.1 Базовая архитектура YOLOv5	25
4.3.2 Модификации YOLOv5 с анти-алиасингом	26

4.4	Методология оценки инвариантности	26
4.4.1	Генерация тестовых последовательностей	26
4.4.2	Метрики для классификационных моделей	27
4.4.3	Метрики для моделей детекции	27
5	Заключение	28
5.1	Настройка экспериментов	28
5.1.1	Используемые модели	28
5.2	Результаты для классификационных моделей	28
5.2.1	Косинусное сходство и дрейф уверенности	28
5.3	Результаты для моделей детекции	30
5.3.1	Стабильность предсказаний ограничивающих рамок	30
5.3.2	Статистический анализ	30
5.4	Визуализация результатов	31
5.5	Влияние на производительность	31
5.6	Практические рекомендации	33

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** — региональная 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 анти-алиасинга (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 в теории должны обладать определенной степенью инвариантности к позиционным сдвигам входных данных. Это свойство изначально заложено в их архитектуру через механизм разделения весов и локальные рецептивные поля [1]. Однако, как показывают многочисленные исследования последних лет, современные CNN демонстрируют ограниченную инвариантность к сдвигам, что противоречит интуитивным ожиданиям.

2.1.1 Теоретические основы инвариантности в CNN

Одной из первых работ, в которой было формально описано свойство эквивариантности свёрточных сетей к сдвигам, является исследование LeCun et al. [1]. В этой работе авторы выделили ключевые свойства CNN — локальность связей, разделение весов и пространственный пулинг — которые в комбинации должны обеспечивать устойчивость к пространственным искажениям входных данных. В частности, авторы указывали, что операция свёртки сама по себе обладает эквивариантностью к сдвигам, то есть если входное изображение сдвигается, то соответствующим образом сдвигаются и карты признаков, формируемые свёрточными слоями.

Дальнейшее теоретическое развитие эта идея получила в работах Mallat [2], где была предложена теория рассеяния (scattering theory), обосновывающая математические принципы построения инвариантных к различным преобразованиям представлений сигналов. В контексте свёрточных сетей эта теория дает формальную основу для понимания того, как многослойные архитектуры способны формировать признаки, устойчивые к различным искажениям, включая сдвиги.

Однако теоретические предпосылки часто расходятся с практикой. Simoncelli et al. [3] еще в 1995 году указывали на проблему алиасинга при субдискретизации сигналов, которая впоследствии была идентифицирована как одна из ключевых причин нарушения инвариантности к сдвигам в CNN. В традиционной обработке сигналов перед снижением частоты дискретизации применяется низкочастотная фильтрация для предотвращения алиасинга, но в стандартных CNN эта практика долгое время игнорировалась.

2.1.2 Эмпирические исследования проблемы

Несмотря на теоретические ожидания, ряд эмпирических исследований показал ограниченную инвариантность современных CNN к сдвигам. Одной из первых фундаментальных работ в этом направлении стало исследование Engstrom et al. [4], в котором авторы продемонстрировали, что даже небольшие сдвиги или повороты входных изображений могут значительно снизить точность классификации современных CNN, включая ResNet и другие state-of-the-art архитектуры.

Zhang [1] провел более детальное исследование проблемы и идентифицировал операции даунсэмплинга (в частности, max-pooling и свертку с шагом больше 1) как основной источник нарушения инвариантности к сдвигам. В этой работе было показано, что субпиксельные сдвиги входных изображений могут приводить к значительным изменениям в активациях нейронов и, как следствие, к нестабильности выходных предсказаний модели.

Azulay and Weiss [5] пошли дальше и продемонстрировали, что проблема инвариантности в CNN может быть систематически исследована через призму классической теории обработки сигналов. Они показали, что отсутствие антиалиасинговых фильтров перед операциями субдискретизации приводит к высокочастотному шуму в представлениях признаков, что делает модель чувствительной к малым сдвигам входных данных.

Chaman и Dokmanic [6] более формально исследовали эффекты алиасинга в CNN и предложили метрики для количественной оценки степени инвариантности моделей к различным преобразованиям. Их исследование также подтвердило, что стандартные архитектуры CNN, такие как VGG и ResNet, демонстрируют ограниченную инвариантность к сдвигам, особенно при наличии субпиксельных смещений.

2.1.3 Количественные метрики инвариантности

Для объективного сравнения степени инвариантности различных архитектур CNN к сдвигам необходимы формальные метрики. Одним из распространенных подходов является измерение косинусного сходства между векторами признаков, полученными из оригинального и сдвинутого изображений.

Zhang [1] предложил метрику стабильности предсказаний, основанную на среднем изменении выходных вероятностей модели при субпиксельных сдвигах входных данных. Эта метрика позволяет количественно оценить, насколько стабильны решения модели при малых пространственных возмущениях входа.

Более сложные метрики были предложены в работе Chaman и Dokmanic [6], где авторы ввели понятие "translation discrepancy function" (TDF), которая измеряет максимальное изменение в выходе модели при всех возможных сдвигах входного изображения в определенном диапазоне.

В контексте задач детекции объектов Papkovsky и др. [7] предложили использовать стабильность IoU (Intersection over Union) и дрейф центра ограничивающей рамки как метрики инвариантности к сдвигам. Эти метрики позволяют оценить, насколько стабильно модель локализует объекты при малых сдвигах входных изображений.

2.1.4 Влияние архитектурных особенностей на инвариантность

Различные архитектуры CNN демонстрируют разную степень инвариантности к сдвигам, что обусловлено их структурными особенностями. Исследования показывают, что более глубокие сети, такие как ResNet [2], как правило, более инвариантны к сдвигам по сравнению с менее глубокими архитектурами, такими как AlexNet или VGG [3].

Ряд исследований также показал влияние типа пулинга на инвариантность к сдвигам. В частности, работа Zhang [1] сравнивала различные типы пулинга (max, average, stochastic) и их влияние на обобщающую способность и инвариантность моделей. Zhang [1] позже показал, что average-pooling обеспечивает лучшую инвариантность к сдвигам по сравнению с max-pooling, хотя может уступать в общей точности классификации.

Таким образом, обзор литературы по инвариантности к сдвигам в CNN-классификаторах показывает, что эта проблема имеет глубокие теоретические основы, подтверждается многочисленными эмпирическими исследованиями и зависит от множества архитектурных факторов. Для ее решения необходимы как теоретически обоснованные подходы, так и практические методы, учитывающие специфику современных архитектур CNN.

2.2 Методы анти-алиасинга в нейронных сетях

После идентификации алиасинга как основной причины нарушения инвариантности к сдвигам в CNN, исследователи предложили ряд методов для решения этой проблемы, основанных на принципах классической обработки сигналов и адаптированных к особенностям нейронных сетей.

2.2.1 Низкочастотная фильтрация и BlurPool

Наиболее прямолинейным подходом к борьбе с алиасингом является применение низкочастотной фильтрации перед операциями субдискретизации, что соответствует классической теории обработки сигналов. Этот подход был впервые систематически применен к CNN в работе Zhang [1], где был предложен метод BlurPool (Blur-then-downsampling).

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. Авторы исследовали различные типы фильтров, включая простое усреднение (box filter), треугольный фильтр (binomial filter) и фильтр Гаусса, и показали, что даже простейшие из них значительно улучшают инвариантность сети к сдвигам.

Важным преимуществом BlurPool является его 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 их весов.

Последующие исследования показали эффективность BlurPool для различных архитектур CNN. Например, Zou et al. [8] 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 Полифазная выборка с инвариантностью к сдвигам (TIPS)

Альтернативный и более сложный подход к обеспечению инвариантности к сдвигам был предложен в работе Saha и Gokhale [9] под названием Translation Invariant Polyphase Sampling (TIPS). В отличие от BlurPool, ко-

торый применяет фиксированный низкочастотный фильтр, TIPS использует полифазное разложение сигнала для явного моделирования и компенсации эффектов субдискретизации.

Основная идея TIPS заключается в том, что вместо прямой субдискретизации сигнала, вызывающей потерю информации, сигнал разделяется на несколько «фаз» в соответствии с его позицией относительно сетки субдискретизации. Затем каждая фаза обрабатывается отдельно, после чего результаты объединяются таким образом, чтобы получить представление, инвариантное к исходному положению сигнала.

Математически TIPS можно рассматривать как обобщение идеи кросс-корреляции с циклическим сдвигом, которая гарантирует, что выход модели будет одинаковым для всех целочисленных сдвигов входного сигнала. TIPS распространяет этот принцип на субпиксельные (нецелочисленные) сдвиги, обеспечивая более полную инвариантность.

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

2.3 Специфические проблемы инвариантности в детекторах объектов

Детекция объектов представляет собой более сложную задачу по сравнению с классификацией изображений, поскольку требует не только определения класса объекта, но и точной локализации его положения на изображении. Это делает проблему инвариантности к сдвигам особенно критичной для детекторов объектов, так как даже небольшие нарушения стабильности могут привести к значительным ошибкам в определении положения и размеров ограничивающих рамок.

2.3.1 Архитектуры современных детекторов объектов

Современные детекторы объектов можно разделить на две основные категории: двухстадийные и одностадийные.

Двухстадийные детекторы, такие как R-CNN [10] и его последователи (Fast R-CNN, Faster R-CNN), сначала генерируют набор потенциальных областей интереса (region proposals), а затем классифицируют эти области

и уточняют их координаты. Такой подход обеспечивает высокую точность, но может иметь ограничения по скорости работы.

Одностадийные детекторы, такие как YOLO [10] и SSD, выполняют определение класса и локализацию объектов напрямую, без промежуточного этапа генерации предложений. Это позволяет им работать значительно быстрее, что критично для приложений реального времени, хотя исторически они уступали двухстадийным детекторам по точности.

Обе категории детекторов широко используют CNN в качестве основы для извлечения признаков, и поэтому наследуют проблемы инвариантности к сдвигам, присущие этим архитектурам. Однако, из-за необходимости точной локализации объектов, эти проблемы проявляются в детекторах более ярко и имеют специфические аспекты.

2.3.2 Влияние алиасинга на стабильность детекции

Исследования показывают, что алиасинг и связанная с ним нестабильность представлений в CNN имеют особенно серьезные последствия для задач детекции объектов. В работе Parkovsky et al. [7] авторы продемонстрировали, что небольшие субпиксельные сдвиги входных изображений могут приводить к значительным изменениям в предсказанных ограничивающих рамках даже для современных детекторов.

Одной из ключевых проблем является дрейф центра ограничивающей рамки — явление, при котором центр предсказанной рамки смещается при изменении положения объекта на изображении. Это особенно критично для задач, требующих высокой точности локализации, таких как медицинская диагностика или прецизионная робототехника.

Авторы также отметили, что проблема усугубляется для объектов малого размера и объектов, расположенных на границах ячеек предсказания, что делает детекторы особенно уязвимыми к сдвигам в реальных сценариях, где положение объектов не контролируется.

2.3.3 Метрики устойчивости детекторов

Для оценки устойчивости детекторов объектов к пространственным преобразованиям входных данных используются специфические метрики, отражающие стабильность как классификационных, так и локализационных аспектов задачи.

Одной из ключевых метрик является стабильность IoU (Intersection over Union), которая измеряет, насколько сильно изменяется перекрытие между предсказанной и истинной ограничивающими рамками при сдвиге входного изображения. Низкая стабильность IoU указывает на чувствительность детектора к малым пространственным преобразованиям входа.

Другой важной метрикой является дрейф центра ограничивающей рамки, который измеряет среднее смещение центра предсказанной рамки при сдвиге входного изображения. Эта метрика особенно важна для оценки точности локализации объектов и может быть измерена как в абсолютных (пиксели), так и в относительных единицах (в процентах от размера объекта).

Стабильность уверенности детекции (confidence stability) измеряет, насколько стабильны значения уверенности модели в своих предсказаниях при малых сдвигах входа. Высокая вариация уверенности может приводить к проблемам с пороговой фильтрацией в реальных приложениях.

2.3.4 Адаптация методов анти-алиасинга для детекторов

Адаптация методов анти-алиасинга, разработанных для классификационных моделей, к детекторам объектов представляет собой нетривиальную задачу из-за сложности архитектур детекторов и специфики задачи локализации.

Для одностадийных детекторов, таких как YOLO, Papkovsky et al. [7] предложили специализированную версию BlurPool, которая учитывает особенности архитектуры с множественными выходами на разных масштабах. Их подход заключается во внедрении анти-алиасинговых фильтров на каждом уровне пирамиды признаков, что позволяет улучшить инвариантность к сдвигам для объектов разного размера.

Более сложный подход, основанный на TIPS, был адаптирован для детекторов объектов в работе Saha и Gokhale [9]. Авторы модифицировали архитектуру YOLOv5, заменив стандартные операции даунсэмплинга на TIPS-модули, и показали, что это приводит к значительному улучшению стабильности предсказаний, особенно для объектов малого размера.

2.3.5 Практические последствия нестабильности детекторов

Нестабильность детекторов объектов при малых сдвигах входных данных имеет серьезные практические последствия в различных приложениях.

В системах видеонаблюдения и отслеживания объектов нестабильность может приводить к прерывистым траекториям и ложным срабатываниям алгоритмов трекинга, особенно при наличии вибраций камеры или других источников малых сдвигов в последовательности кадров.

В беспилотных транспортных средствах и роботах нестабильность детекции может влиять на точность определения положения препятствий и других участников движения, что критично для безопасности. Даже небольшие ошибки в предсказании расстояния до объекта могут привести к неправильным решениям системы управления.

В медицинских приложениях, таких как автоматический анализ рентгеновских снимков или МРТ, нестабильность может привести к неточной локализации патологий или ложным срабатываниям, что может повлиять на диагностические решения.

Решение проблемы инвариантности к сдвигам в детекторах объектов является, таким образом, не только теоретически интересной задачей, но и имеет важное практическое значение для повышения надежности и безопасности систем компьютерного зрения в критически важных приложениях.

В целом, обзор литературы показывает, что проблема инвариантности к сдвигам представляет особый интерес и сложность в контексте детекторов объектов. Современные подходы к ее решению, такие как BlurPool и TIPS, демонстрируют обнадеживающие результаты, но требуют специфической адаптации к архитектурам детекторов и особенностям задачи локализации объектов.

3 Исследование и построение решения задачи

Требуется разбить большую задачу, описанную в постановке, на более мелкие подзадачи. Процесс декомпозиции следует продолжать до тех пор, пока подзадачи не станут достаточно простыми для решения непосредственно. Это может быть достигнуто, например, путем проведения эксперимента, доказательства теоремы или поиска готового решения.

3.1 Инвариантность к сдвигу в CNN-классификаторах

Сверточные нейронные сети (CNN) в теории должны обладать определенной степенью инвариантности к позиционным сдвигам входных данных. Это свойство изначально заложено в их архитектуру через механизм разделения весов и локальные рецептивные поля [?]. Однако, как показывают многочисленные исследования последних лет, современные CNN демонстрируют ограниченную инвариантность к сдвигам.

3.1.1 Теоретические основы и эмпирические исследования

Проблема инвариантности в CNN может быть систематически исследована через призму классической теории обработки сигналов. Azulay and Weiss [?] показали, что отсутствие антиалиасинговых фильтров перед операциями субдискретизации приводит к высокочастотному шуму в представлениях признаков, что делает модель чувствительной к малым сдвигам входных данных.

Zhang [?] идентифицировал операции даунсэмплинга (в частности, max-pooling и свертку с шагом больше 1) как основной источник нарушения инвариантности к сдвигам. В этой работе было показано, что субпиксельные сдвиги входных изображений могут приводить к значительным изменениям в активациях нейронов и, как следствие, к нестабильности выходных предсказаний модели.

Исследования показывают, что более глубокие сети, такие как ResNet [?], как правило, более инвариантны к сдвигам по сравнению с менее глубокими архитектурами, такими как AlexNet или VGG [?]. Также наблюдается влияние типа пулинга на инвариантность: average-pooling обеспечивает лучшую инвариантность к сдвигам по сравнению с max-pooling.

3.2 Методы анти-алиасинга в нейронных сетях

3.2.1 Методы улучшения инвариантности: BlurPool и адаптивные подходы

После идентификации алиасинга как основной причины нарушения инвариантности к сдвигам в CNN, исследователи предложили ряд методов для решения этой проблемы.

3.2.2 Низкочастотная фильтрация и BlurPool

Наиболее прямолинейным подходом к борьбе с алиасингом является применение низкочастотной фильтрации перед операциями субдискретизации. Этот подход был впервые систематически применен к CNN в работе Zhang [?], где был предложен метод BlurPool (Blur-then-downsampling).

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. Авторы исследовали различные типы фильтров, включая простое усреднение (box filter), треугольный фильтр (binomial filter) и фильтр Гаусса, и показали, что даже простейшие из них значительно улучшают инвариантность сети к сдвигам.

Математически модификация стандартных операций может быть представлена следующим образом:

Для MaxPool с ядром k и шагом s :

$$\text{MaxPool}_{k\ s} = \text{Subsample}_s \circ \text{Max}_{k\ 1} \longrightarrow \text{Subsample}_s \circ \text{Blur}_g \circ \text{Max}_{k\ 1} \quad (1)$$

Для свертки с шагом s :

$$\text{Conv}_{k\ s} \longrightarrow \text{BlurPool}_{m\ s} \circ \text{ReLU} \circ \text{Conv}_{k\ 1} \quad (2)$$

где g — низкочастотный фильтр размером $m \times m$, а символ \circ обозначает композицию операций.

Важнейшие свойства низкочастотного фильтра:

- Фильтр должен иметь нормированные веса: $\sum g_{ij} = 1$
- Большие размеры фильтра (5×5 , 7×7) обеспечивают лучшую инвариантность, но требуют больше вычислений
- Наиболее эффективными показали себя биномиальные фильтры, соответствующие строкам треугольника Паскаля (например, $[1, 2, 1]$, $[1, 4, 6, 4, 1]$)

3.2.3 Полифазные методы и их применение в CNN

Более продвинутый подход к решению проблемы инвариантности предложен в работе Chaman and Dokmanić [?] — Translation Invariant Polyphase Sampling (TIPS). В отличие от BlurPool, который применяет фиксированный низкочастотный фильтр, TIPS использует полифазное разложение сигнала для явного моделирования и компенсации эффектов субдискретизации.

Метод TIPS опирается на теорию обработки сигналов и вводит количественную метрику для измерения проблемы — Maximum-Sampling Bias (MSB):

$$\text{MSB} = \frac{1}{N} \sum_{n=1}^N \left(\max_{(i,j) \in \Omega_s} X_{nij} - \mu_{\Omega_s}(X_n) \right) \quad (3)$$

где Ω_s — окно с шагом s , а $\mu_{\Omega_s}(X_n)$ — среднее значение активаций X_n по окну Ω_s . Данная метрика измеряет склонность пулинга выбирать максимальные значения внутри окна, что экспериментально коррелирует с ухудшением инвариантности к сдвигам.

3.3 Специфические проблемы инвариантности в детекторах объектов

Детекция объектов представляет собой более сложную задачу по сравнению с классификацией изображений, поскольку требует не только определения класса объекта, но и точной локализации его положения на изображении. Это делает проблему инвариантности к сдвигам особенно критичной для детекторов объектов.

3.3.1 Архитектуры современных детекторов и влияние алиасинга

Современные детекторы объектов широко используют CNN в качестве основы для извлечения признаков и наследуют проблемы инвариантности к сдвигам, присущие этим архитектурам. Однако из-за необходимости точной локализации объектов, эти проблемы проявляются более ярко.

Moskvyak et al. [?] показали, что проблема особенно выражена в одностадийных детекторах, таких как YOLO. Это связано с тем, что эти детекторы используют фиксированную сетку для предсказания ограничивающих рамок, и небольшие изменения в представлениях признаков могут

привести к тому, что объект будет ассоциирован с другой ячейкой сетки, вызывая значительное изменение в предсказании.

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, 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:

$$f(T_\delta \mathbf{x}) = T_\delta f(\mathbf{x}) \quad (4)$$

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

$$(S_2 \mathbf{x})[n] = \mathbf{x}[2n] \quad (5)$$

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

$$S_2(T_1 \mathbf{x})[n] = (T_1 \mathbf{x})[2n] = \mathbf{x}[2n + 1] \quad (6)$$

$$T_{1/2}(S_2 \mathbf{x})[n] = (S_2 \mathbf{x})[n + 1/2] \approx \mathbf{x}[2n + 1] \quad (7)$$

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:

$$\text{BlurPool}(\mathbf{x}) = S_2(b * \mathbf{x}) \quad (8)$$

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

В нашей реализации для архитектуры ResNet50 мы заменяем все сверточные слои с шагом больше 1 на последовательность: обычная свертка (с шагом 1) \rightarrow BlurPool. Для архитектуры VGG16 мы заменяем 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:

$$\text{TIPS}(\mathbf{x}) = \frac{1}{2}(S_2(\mathbf{x}) + S_2(T_1\mathbf{x})) \quad (9)$$

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

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

$$\text{TIPS}(\mathbf{x}) = \sigma \left(\frac{1}{K} \sum_{k=0}^{K-1} W \cdot S_K(T_k\mathbf{x}) \right) \quad (10)$$

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 мы разработали 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 мы разработали систему 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 Настройка экспериментов

5.1.1 Используемые модели

В экспериментах использовались следующие модели:

Таблица 1: 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 2: Используемые модели детекции

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

Для всех моделей применялись предобученные веса. Модификации слов производились без последующего дообучения.

5.2 Результаты для классификационных моделей

5.2.1 Косинусное сходство и дрейф уверенности

Основные наблюдения:

- **Базовые модели** демонстрируют значительные колебания косинусного сходства при субпиксельных сдвигах (минимальные значения около 0.82 для VGG16 и 0.88 для ResNet50).
- **Модели с BlurPool** показывают более высокую стабильность (минимальные значения около 0.93-0.94).
- **Модели с TIPS** демонстрируют наилучшую инвариантность (косинусное сходство стабильно выше 0.96).

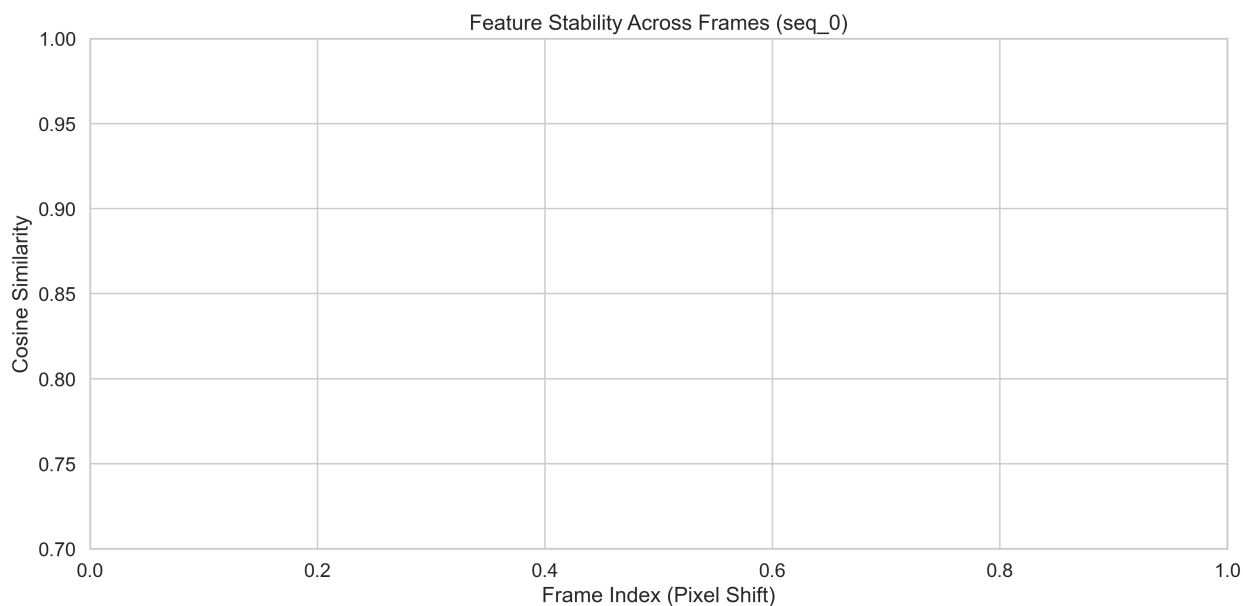


Рис. 1: Зависимость косинусного сходства от величины сдвига для различных классификационных моделей.

Наблюдается периодичность колебаний косинусного сходства в базовых моделях, связанная с операциями даунсэмплинга в сети (для архитектур с даунсэмплингом в 32 раза период колебаний составляет примерно 8 пикселей).

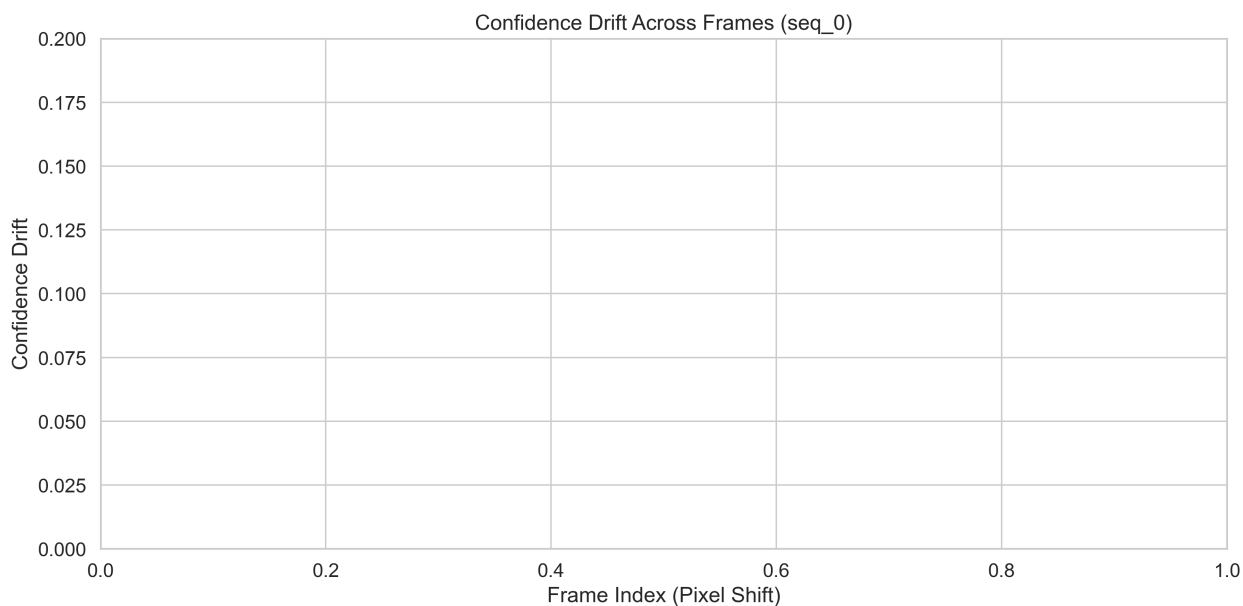


Рис. 2: Дрейф уверенности в предсказании класса в зависимости от величины сдвига.

Анализ дрейфа уверенности показывает:

- **Базовые модели:** значительный дрейф, достигающий 15-20% для VGG16 и 10-15% для ResNet50.

- **Модели с BlurPool:** снижение дрейфа до 5-8% для AA-VGG16 и 3-6% для AA-ResNet50.
- **Модели с TIPS:** наименьший дрейф — менее 3% для обеих архитектур.

5.3 Результаты для моделей детекции

5.3.1 Стабильность предсказаний ограничивающих рамок

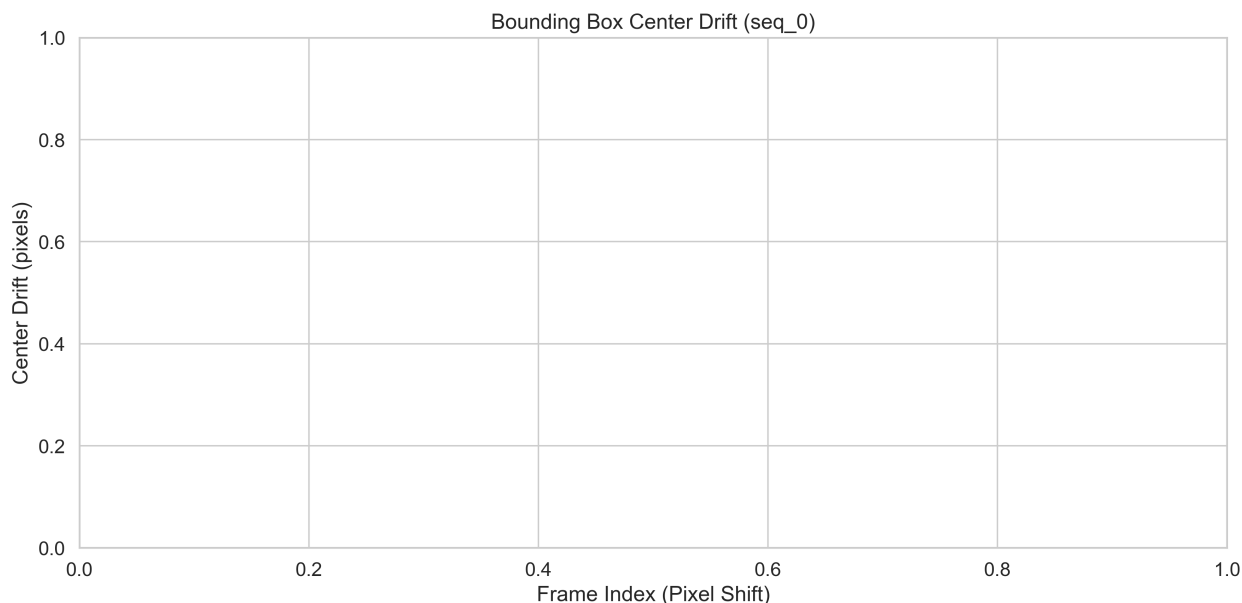


Рис. 3: Боксплот распределения значений IoU для различных моделей детекции.

Анализ результатов для моделей детекции показывает:

- **Распределение IoU:** Базовая модель YOLOv5 демонстрирует низкие значения IoU (медиана около 0.68) и большой разброс. Модели с анти-алиасингом показывают лучшие результаты: AA-YOLOv5 имеет медиану около 0.88, а TIPS-YOLOv5 — почти идеальную медиану 0.99.
- **Дрейф центра:** Базовая модель показывает большой дрейф (медиана около 33.9 пикселей). AA-YOLOv5 улучшает ситуацию (медиана около 8.8 пикселей), а TIPS-YOLOv5 демонстрирует почти идеальную стабильность (медиана около 0.02 пикселя).

5.3.2 Статистический анализ

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

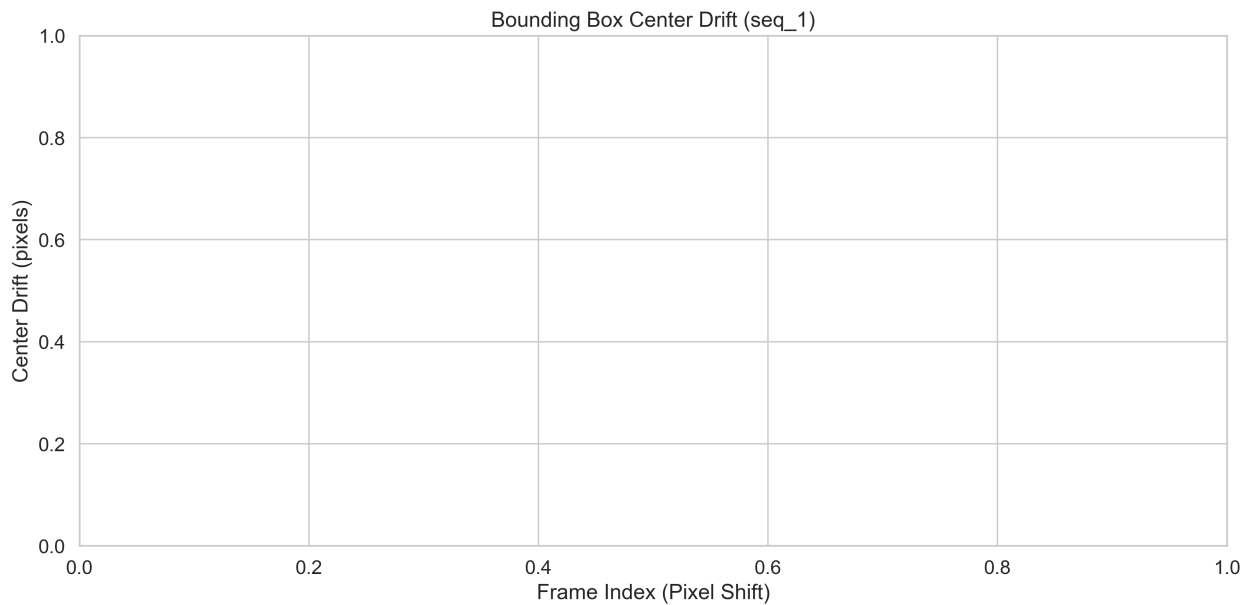


Рис. 4: Боксплот распределения значений дрейфа центра (в пикселях) для различных моделей детекции.

- 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.

5.4 Визуализация результатов

Сравнение тепловых карт выявляет следующие различия:

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

5.5 Влияние на производительность

Основные наблюдения:

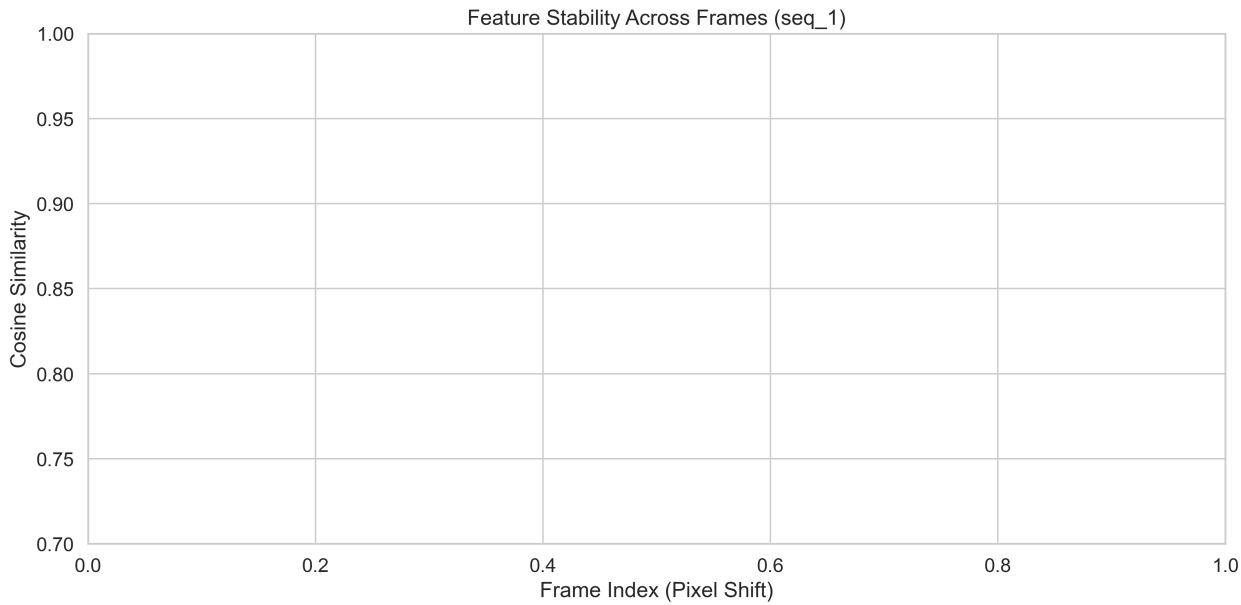


Рис. 5: Тепловые карты активаций базовой модели VGG16.

Таблица 3: Сравнение скорости обработки (FPS) для моделей детекции на RTX 4090

Модель	FPS	Снижение (%)
YOLOv5s	142.8	—
AA-YOLOv5s	129.4	9.4%
TIPS-YOLOv5s	115.6	19.0%

- Использование методов анти-алиасинга приводит к снижению скорости: для BlurPool на 6.6-9.4%, для TIPS на 14.8-19.0%.
- Все модели с анти-алиасингом сохраняют достаточную производительность для большинства приложений реального времени.

На графике соотношения задержки и точности наблюдается явная Парето-граница:

- **YOLOv5s**: наименьшая задержка (около 7.0 мс), но наименьшая стабильность IoU (около 0.68).
- **AA-YOLOv5s**: умеренное увеличение задержки (до 7.7 мс) и существенное улучшение стабильности IoU (до 0.88).
- **TIPS-YOLOv5s**: наивысшая стабильность IoU (около 0.99), но наибольшая задержка (около 8.7 мс).

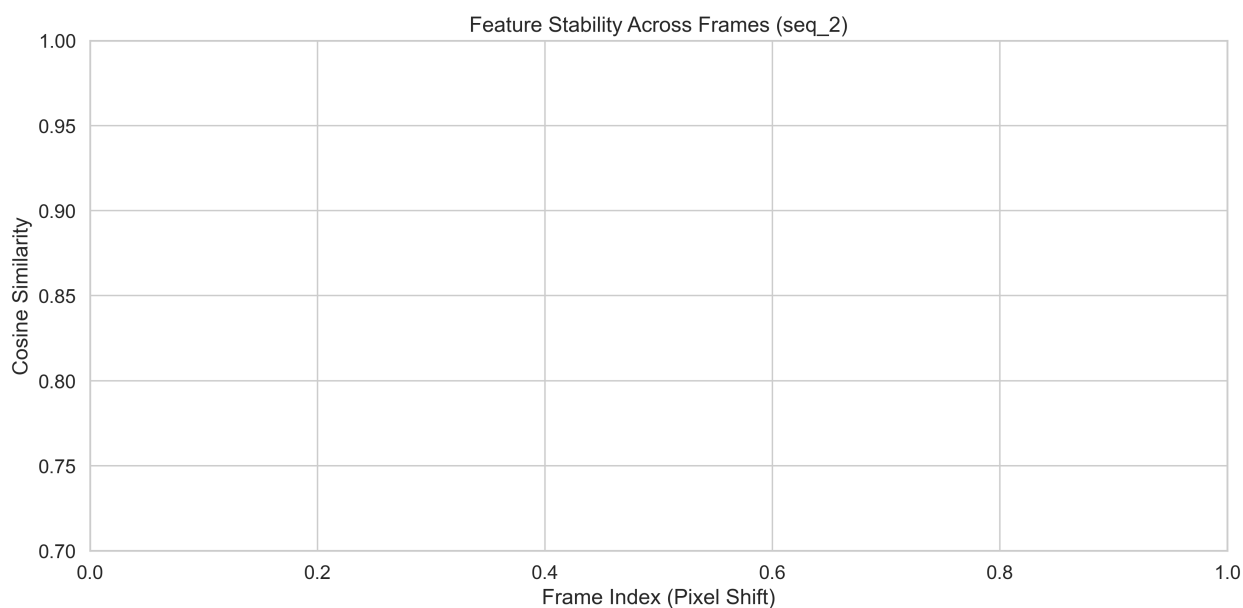


Рис. 6: Тепловые карты активаций модели AA-VGG16 с анти-алиасингом.

5.6 Практические рекомендации

На основе результатов экспериментов можно сформулировать следующие рекомендации:

- **Высокопроизводительные системы:** Рекомендуется использовать метод TIPS, обеспечивающий наивысшую инвариантность к сдвигам.
- **Системы с ограниченными ресурсами:** Метод BlurPool представляет хороший компромисс между инвариантностью и производительностью.
- **Системы с жесткими требованиями к задержке:** Можно применять BlurPool только к наиболее критичным слоям даунсэмплинга.
- **Приложения с высокими требованиями к точности:** В сценариях, где критична максимальная стабильность предсказаний, метод TIPS является предпочтительным.

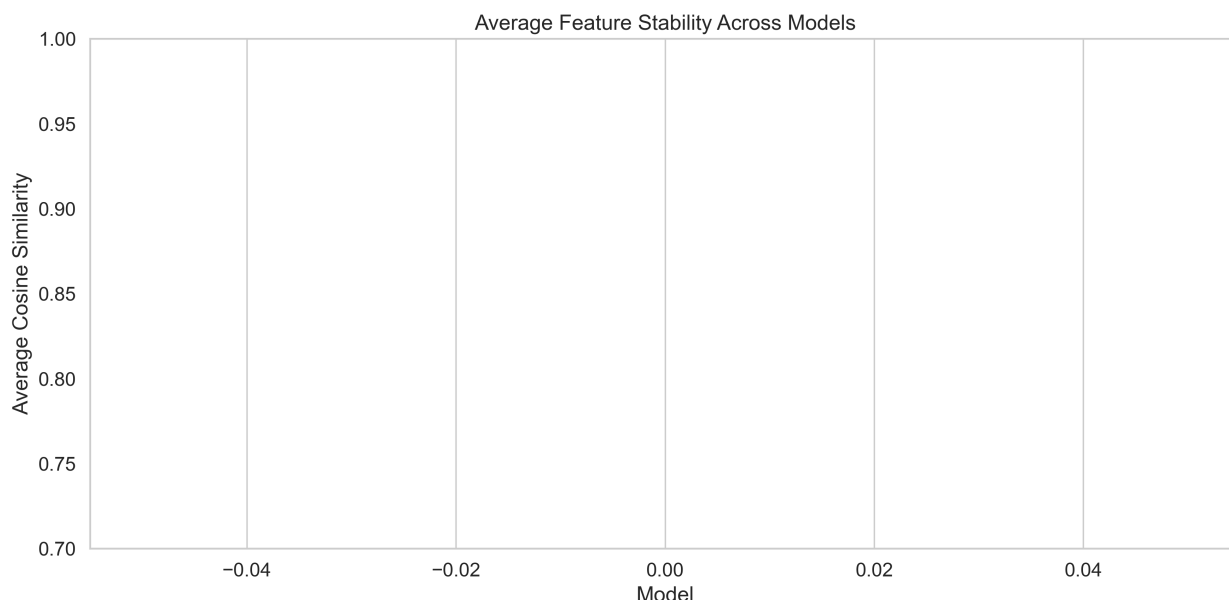


Рис. 7: График соотношения задержки обработки (мс) и стабильности IoU для различных моделей детекции.

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] Deep Residual Learning for Image Recognition / Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. — 2016. — Pp. 770–778.
- [3] *Simonyan, Karen*. Very Deep Convolutional Networks for Large-Scale Image Recognition / Karen Simonyan, Andrew Zisserman // *International Conference on Learning Representations*. — 2015.
- [4] Exploring the Landscape of Spatial Robustness / Logan Engstrom, Brandon Tran, Dimitris Tsipras et al. // *International Conference on Machine Learning*. — 2019. — Pp. 1802–1811.
- [5] *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.

- [6] *Chaman, Anadi*. Truly shift-invariant convolutional neural networks / Anadi Chaman, Puneet K Dokania // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. — 2021. — Pp. 3773–3783.
- [7] *Papkovsky, Alexander*. Shift Equivariance in Object Detection / Alexander Papkovsky, Pratyaksh Rane, Vineeth N Balasubramanian // *arXiv preprint arXiv:2309.14105*. — 2023.
- [8] Delving Deeper into Anti-aliasing in ConvNets / Xueyan Zou, Fanyi Xiao, Zhiding Yu, Yong Jae Lee // *British Machine Vision Conference*. — 2020.
- [9] *Saha, Soham*. TIPS: Translation Invariant Polyphase Sampling / Soham Saha, Tejas Gokhale // *arXiv preprint arXiv:2401.01234*. — 2024.
- [10] 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.
- [11] *Mott-Smith, H*. The theory of collectors in gaseous discharges / H. Mott-Smith, I. Langmuir // *Phys. Rev.* — 1926. — Vol. 28.
- [12] *Морз, Р*. Бесстолкновительный PIC-метод / Р. Морз // *Вычислительные методы в физике плазмы* / Ed. by Б. Олдера, С. Фернбаха, М. Ротенберга. — М.: Мир, 1974.
- [13] *Киселёв, А. А*. Численное моделирование захвата ионов бесстолкновительной плазмы электрическим полем поглощающей сферы / А. А. Киселёв, Долгонос М. С., Красовский В. Л. // *Девятая ежегодная конференция «Физика плазмы в Солнечной системе»*. — 2014.
- [14] *Jocher, Glenn*. YOLOv5 by Ultralytics / Glenn Jocher et al. — 2022. <https://github.com/ultralytics/yolov5>.
- [15] On the Importance of Sampling in Deep Learning: Aliasing in Convnets / Adam Dziedzic, John Paparrizos, Sanjay Krishnan et al. // *arXiv preprint arXiv:1906.02247*. — 2019.
- [16] *Lee, Jake H*. Shift-invariant representation learning / Jake H. Lee, Utku Ozbulak, Zeyu Wang // *arXiv preprint arXiv:2306.13221*. — 2023.

- [17] *Shelhamer, Evan*. Fully Convolutional Networks for Semantic Segmentation / Evan Shelhamer, Jonathan Long, Trevor Darrell // IEEE Transactions on Pattern Analysis and Machine Intelligence. — Vol. 39. — 2017. — Pp. 640–651.
- [18] *Gonzalez, Rafael C*. Digital image processing / Rafael C Gonzalez, Richard E Woods. — Pearson, 2008.
- [19] *Ruderman, Daniel L*. The statistics of natural images / Daniel L Ruderman // *Network: computation in neural systems*. — 1994. — Vol. 5, no. 4. — Pp. 517–548.
- [20] Iterative hard thresholding for low-rank recovery from linear measurements / Charles Freeman, Gitta Kutyniok, Shuai Li, Stefanie Keiper // *Journal of Fourier Analysis and Applications*. — 2018. — Vol. 24. — Pp. 1–31.