

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

Полев Алексей Михайлович

**СИМУЛЯЦИЯ СДВИГА ДОМЕНА И АНАЛИЗ
ИЗМЕНЕНИЯ ПРОМЕЖУТОЧНЫХ
НИЗКОРАЗМЕРНЫХ ПРЕДСТАВЛЕНИЙ В
RESNET-ПОДОБНЫХ АРХИТЕКТУРАХ ДЛЯ
ЗАДАЧИ КОМПЬЮТЕРНОГО ЗРЕНИЯ**

Специальность 01.03.02 —

«Прикладная математика и информатика»

Научный руководитель:

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 инвари-

антности 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.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

- CNN** — Convolutional Neural Network, сверточная нейронная сеть.
- AA** — Anti-Aliasing, анти-алиасинг, техника устранения эффекта алиасинга.
- TIPS** — Translation Invariant Polyphase Sampling, метод полифазной выборки с инвариантностью к сдвигам.
- RF** — Receptive Field, рецептивное поле нейрона в CNN.
- IoU** — Intersection over Union, метрика, измеряющая отношение площади пересечения к площади объединения двух областей.
- FPS** — Frames Per Second, количество кадров в секунду, метрика производительности.
- ρ — Коэффициент косинусной схожести между векторами признаков, $\rho(a, b) = \frac{a \cdot b}{\|a\| \cdot \|b\|}$.
- σ — Стандартное отклонение, мера разброса величины.
- Δ_c — Дрейф центра, метрика смещения предсказанного центра ограничивающей рамки.
- VGG** — Visual Geometry Group, архитектура CNN, разработанная в Оксфордском университете.
- ResNet** — Residual Network, архитектура CNN с остаточными связями.
- YOLO** — You Only Look Once, архитектура модели для детекции объектов в реальном времени.
- BlurPool** — Метод анти-алиасинга с применением размытия перед операцией пулинга.

Содержание

	Стр.
Введение	2
Глава 1. Обзор литературы	9
1.1 Инвариантность к сдвигу в CNN-классификаторах	9
1.1.1 Теоретические основы инвариантности в CNN	9
1.1.2 Эмпирические исследования проблемы	10
1.1.3 Количественные метрики инвариантности	11
1.1.4 Влияние архитектурных особенностей на инвариантность	12
1.2 Методы анти-алиасинга в нейронных сетях	13
1.2.1 Низкочастотная фильтрация и BlurPool	14
1.2.2 Полифазная выборка с инвариантностью к сдвигам (TIPS)	15
1.2.3 Другие подходы к обеспечению инвариантности	16
1.2.4 Применение анти-алиасинга в различных задачах . .	17
1.3 Специфические проблемы инвариантности в детекторах объектов	18
1.3.1 Архитектуры современных детекторов объектов	18
1.3.2 Влияние алиасинга на стабильность детекции	19
1.3.3 Метрики устойчивости детекторов	20
1.3.4 Адаптация методов анти-алиасинга для детекторов .	21
1.3.5 Практические последствия нестабильности детекторов	22
1.4 Другие типы инвариантности в нейронных сетях	23
1.4.1 Инвариантность к масштабу	23
1.4.2 Инвариантность к повороту	24

1.4.3	Другие типы геометрических инвариантностей	25
1.4.4	Связь между различными типами инвариантности . .	26
1.5	Обучение с функцией потерь согласованности	27
1.5.1	Основные принципы обучения с согласованностью . .	28
1.5.2	Полуавтоматическое и самоконтролируемое обучение	29
1.5.3	Применение в детекции объектов	29
1.5.4	Преимущества и ограничения	30
1.6	Выводы по обзору литературы	31

Глава 2. Теоретические основы пространственной

	инвариантности в CNN	33
2.1	Формальное определение инвариантности к переносу	33
2.1.1	Инвариантность и эквивариантность функций	33
2.1.2	Операторы сдвига для дискретных и непрерывных сигналов	34
2.1.3	Инвариантность к переносу в контексте CNN	36
2.1.4	Квантификация инвариантности к сдвигу	38
2.1.5	Теоретические границы инвариантности в дискретных CNN	40
2.2	Вывод рецептивных полей в CNN	41
2.2.1	Определение и значимость рецептивных полей	41
2.2.2	Расчёт размера рецептивного поля	42
2.2.3	Связь размера рецептивного поля с инвариантностью к сдвигу	43
2.2.4	Эффективные рецептивные поля	44
2.3	Описание архитектур CNN	45
2.3.1	Архитектура VGG	46
2.3.2	Архитектура ResNet	47
2.3.3	Архитектура YOLO	49

2.3.4	Сравнение архитектур с точки зрения пространственной инвариантности	51
2.4	Теория анти-алиасинга в CNN	51
2.4.1	Причины алиасинга в операциях даунсэмплинга	52
2.4.2	Метод BlurPool	53
2.4.3	Метод TIPS	55
2.4.4	Сравнение методов анти-алиасинга	58
Глава 3.	Экспериментальная часть	60
3.1	Настройка экспериментов	60
3.1.1	Детали скриптов генерации данных	60
3.1.2	Описание чекпоинтов моделей	64
3.1.3	Определения метрик	66
3.2	Эксперименты со статическим сдвигом	69
3.2.1	Методология	69
3.2.2	Результаты косинусного сходства	70
3.2.3	Результаты дрейфа уверенности	71
3.2.4	Сравнительный анализ архитектур	72
3.2.5	Влияние величины сдвига	74
3.3	Динамические последовательности	74
3.3.1	Тепловые карты активаций для классификационных моделей	75
3.3.2	GIF-наложения для моделей детекции	76
3.3.3	Сводные таблицы по моделям	77
3.3.4	Наблюдения и промежуточные выводы	78
3.4	Аблация и устойчивость	79
3.4.1	Аблация размера рецептивного поля	80
3.4.2	Варианты BlurPool	81
3.4.3	Тесты статистической значимости	83

3.5	Профилирование производительности	84
3.5.1	Бенчмарки FPS	85
3.5.2	Соотношение задержки и точности	87
3.5.3	Потребление памяти и энергии	88
3.5.4	Практические рекомендации	89
Заключение		103

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 преобра-

зованиям. 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.

1 Обзор литературы

1.1 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 [?]. Однако, как показывают многочисленные исследования последних лет, современные CNN демонстрируют ограниченную инвариантность к сдвигам, что противоречит интуитивным ожиданиям.

1.1.1 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. [?]. 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 [?], где была предложена теория рассеяния (scattering theory), обосновывающая математические принципы построения инвариантных к раз-

личным преобразованиям представлений сигналов. В контексте сверточных 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. [?] 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.

1.1.2 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. [?], 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 [?] 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 [?] 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 [?] 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.

Работа Kayhan и van Gemert [?] обратила внимание еще на один интересный аспект проблемы — они показали, что современные CNN в процессе обучения могут "запоминать" абсолютные позиции объектов в кадре из обучающей выборки, что также противоречит желаемому свойству инвариантности к положению. Это явление они называли "позиционным кодированием" (position encoding) и предложили методы его смягчения.

1.1.3 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 [?] 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 [?], 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 Manfredi и Wang [?] 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.

1.1.4 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 [?], как правило, более инвариантны к сдвигам по сравнению с менее глубокими архитектурами, такими как AlexNet или VGG [?].

Sabour et al. [?] в своей работе по капсульным сетям указывали на фундаментальные ограничения CNN в части обеспечения инвариантности

к сдвигам и предложили альтернативную архитектуру, в которой явно моделируются пространственные отношения между частями объектов.

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

Исследование Blot et al. [?] обратило внимание на влияние размера рецептивного поля на инвариантность к сдвигам. Авторы продемонстрировали, что модели с большими рецептивными полями, как правило, более устойчивы к пространственным преобразованиям входных данных.

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.

1.2 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.

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

1.2.2 Полифазная выборка с инвариантностью к сдвигам (TIPS)

Альтернативный и более сложный подход к обеспечению инвариантности к сдвигам был предложен в работе Chaman и Dokmanic [?] под названием 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.

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.

1.2.3 Другие подходы к обеспечению инвариантности

Помимо BlurPool и TIPS, в литературе предложены и другие подходы к улучшению инвариантности CNN к сдвигам.

Одним из таких подходов является Deep Shift-Invariant Network (DSI), предложенный Zou et al. [?]. DSI основан на идее моделирования сдвига как дифференцируемой операции и использования глубокого обучения для адаптации параметров размытия к конкретной задаче. В отличие от BlurPool, где параметры фильтра фиксированы, в DSI они оптимизируются в процессе обучения модели.

Другой подход, предложенный в работе Anwar et al. [?], использует идею рандомизированного пулинга (Random Sampling Pooling), где позиции для субдискретизации выбираются случайным образом во время обучения. Это помогает модели не «привязываться» к фиксированным позициям при субдискретизации и лучше генерализоваться на сдвинутые входные данные.

Интересным направлением является также использование явной аугментации данных для улучшения инвариантности. Например, Cheng et al. [?] предложили метод Shift Equivariance Regularization (SER), который во время обучения явно стимулирует модель давать согласованные предсказания для оригинальных и сдвинутых версий входных изображений.

1.2.4 Применение анти-алиасинга в различных задачах

Методы анти-алиасинга нашли применение в различных задачах компьютерного зрения, выходящих за рамки простой классификации изображений.

В области сегментации изображений Zheng et al. [?] показали, что применение BlurPool к архитектурам U-Net и DeepLab значительно улучшает стабильность границ сегментации при малых сдвигах входных изображений, что особенно важно в медицинских приложениях.

В контексте генеративных моделей Karras et al. [?] продемонстрировали, что включение анти-алиасинговых фильтров в генеративно-состязательные сети (GAN) помогает устранить характерные артефакты в генерируемых изображениях и улучшает стабильность процесса обучения.

Для задач детекции объектов Lin et al. [?] адаптировали методы анти-алиасинга к популярным архитектурам детекторов, таким как Faster R-CNN и YOLO, и показали, что это значительно улучшает стабильность предсказаний ограничивающих рамок при малых сдвигах объектов.

В целом, обзор литературы по методам анти-алиасинга в нейронных сетях показывает, что применение принципов классической обработки сигналов к современным архитектурам CNN является эффективным подходом к улучшению их инвариантности к сдвигам. Различные методы, от простого BlurPool до более сложных подходов, основанных на полифазной выборке или адаптивных фильтрах, предоставляют широкий спектр инструментов для повышения устойчивости моделей к пространственным преобразованиям входных данных с различными компромиссами между вычислительной сложностью, архитектурными изменениями и степенью достигаемой инвариантности.

1.3 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.

1.3.1 Архитектуры современных детекторов объектов

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

Двухстадийные детекторы, такие как R-CNN [?] 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 [?] 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 инвариант-

ности к сдвигам, присущие этим архитектурам. Однако, из-за необходимости точной локализации объектов, эти проблемы проявляются в детекторах более ярко и имеют специфические аспекты.

1.3.2 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 Manfredi и Wang [?] 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.

Moskvyak et al. [?] исследовали влияние различных архитектурных компонентов детекторов на их устойчивость к сдвигам и 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.

Авторы также отметили, что проблема усугубляется для объектов малого размера и объектов, расположенных на границах ячеек предска-

зания, что делает детекторы особенно уязвимыми к сдвигам в реальных сценариях, где положение объектов не контролируется.

1.3.3 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.

Yang et al. [?] предложили комплексную метрику инвариантности детекторов, которая объединяет все эти аспекты и позволяет количественно

сравнивать различные архитектуры и методы по их устойчивости к пространственным преобразованиям.

1.3.4 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.

Lin et al. [?] предложили подход к интеграции BlurPool в архитектуру Faster R-CNN, модифицируя как базовую сеть извлечения признаков, так и модуль предложения регионов (Region Proposal Network, RPN). Авторы показали, что такая модификация значительно улучшает стабильность предсказаний ограничивающих рамок при малых сдвигах входных изображений без существенного влияния на общую точность детекции.

Для одностадийных детекторов, таких как YOLO, Wang et al. [?] 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 Chaman et al. [?]. Авторы модифицировали архитектуру YOLOv3, заменив стандартные операции даунсэмплинга на TIPS-модули, и показали, что это приводит к значительному улучшению стабильности предсказаний, особенно для объектов малого размера.

1.3.5 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 специ-

фической адаптации к архитектурам детекторов и особенностям задачи локализации объектов.

1.4 Другие типы инвариантности в нейронных сетях

Хотя проблема инвариантности к сдвигам является одной из наиболее изученных в контексте CNN, существуют и другие типы инвариантности, которые также важны для построения надежных систем компьютерного зрения. Эти типы инвариантности отражают различные преобразования, которым могут подвергаться объекты в реальном мире, такие как изменение масштаба, поворот, аффинные преобразования и другие.

1.4.1 Инвариантность к масштабу

Инвариантность к масштабу — это способность модели одинаково эффективно распознавать объекты независимо от их размера в кадре. Эта проблема особенно актуальна для приложений, где расстояние до объектов может значительно меняться.

Стандартным подходом к обеспечению инвариантности к масштабу в современных CNN является использование многоуровневых (многомасштабных) представлений. Архитектуры, такие как Feature Pyramid Network (FPN) [?], интегрируют информацию с разных уровней сети, создавая пирамиду признаков с разным разрешением и рецептивным полем.

Альтернативный подход — это явное моделирование преобразований масштаба. Например, в работе Kanazawa et al. [?] был предложен метод под названием "Locally Scale-Invariant Convolutional Neural Networks кото-

рый модифицирует свёрточные слои таким образом, чтобы они становились инвариантными к локальным изменениям масштаба.

Xu et al. [?] предложили Scale-Invariant Convolutional Neural Network (SiCNN), в которой входное изображение обрабатывается на нескольких масштабах параллельно, после чего результаты объединяются. Этот подход позволяет явно моделировать инвариантность к масштабу на уровне архитектуры.

В контексте детекции объектов проблема инвариантности к масштабу особенно важна, поскольку объекты могут появляться в широком диапазоне размеров. Современные архитектуры детекторов, такие как RetinaNet [?] и YOLOv3 [?], используют пирамиды признаков для обнаружения объектов разного размера.

1.4.2 Инвариантность к повороту

Инвариантность к повороту — это способность модели одинаково распознавать объекты независимо от их ориентации в пространстве. Эта проблема важна для многих приложений, где ориентация объектов может быть произвольной, например, в спутниковых снимках, медицинских изображениях и распознавании текстур.

Классические CNN не обладают встроенной инвариантностью к повороту, и для решения этой проблемы были предложены различные подходы. Один из них — аугментация данных через случайные повороты во время обучения. Хотя этот подход прост и эффективен, он может требовать значительного увеличения размера обучающей выборки и времени обучения.

Более элегантные архитектурные решения включают Rotation Equivariant Vector Field Networks (RotEqNet) [?], которые используют филь-

тры, способные явно моделировать вращения входных данных, и Harmonic Networks [?], которые основаны на использовании гармонических фильтров, инвариантных к вращению.

Cohen и Welling [?] предложили Group Equivariant Convolutional Networks (G-CNNs), обобщение CNN, которое обеспечивает эквивариантность к более широкому классу преобразований, включая повороты, отражения и другие действия определенных групп симметрий.

Weiler et al. [?] развили эти идеи дальше, предложив Steerable CNNs, которые используют теорию представлений групп для построения фильтров, естественным образом моделирующих вращательные симметрии.

В области детекции объектов инвариантность к повороту особенно важна для задач, где ориентация объектов может быть произвольной, таких как аэрофотосъемка или сканирование багажа в аэропортах. Zhou et al. [?] предложили Oriented Response Networks (ORN), которые расширяют стандартные CNN для обработки входных данных в различных ориентациях и показывают хорошие результаты в задачах детекции и сегментации объектов произвольной ориентации.

1.4.3 Другие типы геометрических инвариантностей

Помимо сдвигов, масштаба и поворотов, существуют и другие типы геометрических преобразований, к которым может быть полезно обеспечить инвариантность моделей.

Аффинные преобразования, которые включают в себя линейные деформации (сжатие, растяжение, сдвиг), часто встречаются в реальных сценариях из-за перспективных искажений и различных условий съемки. Jaderberg et al. [?] предложили архитектуру Spatial Transformer Networks

(STN), которая способна обучаться локализовать и выравнивать релевантные части входного изображения перед его дальнейшей обработкой, что позволяет достичь инвариантности к широкому классу пространственных преобразований.

Проективные преобразования, возникающие из-за перспективы, также могут значительно влиять на внешний вид объектов. Esteves et al. [?] предложили подход, основанный на сферических представлениях, который обеспечивает инвариантность к более широкому классу 3D-преобразований.

Инвариантность к деформациям, таким как изгибы и складки, важна для распознавания эластичных объектов, таких как одежда или биологические структуры. В работе Sun et al. [?] была предложена архитектура Deformation Convolutional Network (DCN), которая адаптирует форму рецептивных полей для моделирования деформаций и обеспечивает лучшую инвариантность к этому типу преобразований.

1.4.4 Связь между различными типами инвариантности

Интересным аспектом исследования является связь между различными типами инвариантности и возможность их совместного обеспечения в рамках одной архитектуры.

Lenc и Vedaldi [?] исследовали теоретическую связь между различными типами инвариантности и показали, что некоторые из них могут быть несовместимы или требовать компромиссов. Например, стремление к полной инвариантности к повороту может негативно влиять на способность модели различать некоторые классы объектов (например, '6' и '9').

Другой важный аспект — это различие между инвариантностью и эквивариантностью. В то время как инвариантность означает, что выход модели не меняется при преобразовании входа, эквивариантность означает, что выход модели преобразуется предсказуемым образом при преобразовании входа. В некоторых задачах, особенно связанных с локализацией, требуется именно эквивариантность, а не полная инвариантность.

Современные подходы, такие как Group Equivariant CNNs и Steerable CNNs, предлагают общую теоретическую основу для моделирования различных типов геометрических инвариантностей и эквивариантностей в рамках единой архитектуры. Эти подходы позволяют достичь баланса между различными типами инвариантности в зависимости от требований конкретной задачи.

1.5 Обучение с функцией потерь согласованности

Помимо архитектурных модификаций для обеспечения инвариантности к различным преобразованиям, важным направлением исследований является разработка специальных функций потерь, которые явно стимулируют модель давать согласованные результаты при преобразованиях входных данных. Этот подход, известный как обучение с функцией потерь согласованности (consistency loss training), представляет собой мощную альтернативу или дополнение к архитектурным модификациям.

1.5.1 Основные принципы обучения с согласованностью

Центральная идея обучения с согласованностью заключается в том, чтобы явно включить в функцию потерь требование, что выходы модели для оригинального и преобразованного входов должны быть связаны определенным образом, соответствующим примененному преобразованию.

Формально, если f — это модель, x — входное изображение, а T — преобразование (например, сдвиг, поворот или изменение масштаба), то можно определить потери согласованности как:

$$L_{consistency} = d(f(x) g(f(T(x)))) \quad (1.1)$$

где d — это некоторая мера расстояния (например, среднеквадратичная ошибка или дивергенция Кульбака-Лейблера), а g — это функция, которая преобразует выход модели для трансформированного входа таким образом, чтобы он был сопоставим с выходом для оригинального входа.

Cheng et al. [?] предложили метод Shift Equivariance Regularization (SER), который добавляет к стандартной функции потерь классификации дополнительный член, штрафующий модель за несогласованность предсказаний при сдвигах входного изображения. Авторы показали, что такой подход может значительно улучшить инвариантность CNN к сдвигам без необходимости внесения изменений в их архитектуру.

Подобный подход был расширен на другие типы преобразований в работе Zhang et al. [?], где авторы предложили общую структуру для обучения с согласованностью для различных геометрических преобразований.

1.5.2 Полуавтоматическое и самоконтролируемое обучение

Важным приложением обучения с согласованностью является полуавтоматическое и самоконтролируемое обучение, где ограниченное количество размеченных данных дополняется большим количеством неразмеченных данных.

Xie et al. [?] предложили метод Unsupervised Data Augmentation (UDA), который использует согласованность предсказаний между оригинальными и аугментированными версиями неразмеченных изображений как сигнал для обучения. Этот подход позволяет эффективно использовать неразмеченные данные для улучшения обобщающей способности модели.

Sohn et al. [?] развили эту идею в методе FixMatch, который сочетает согласованность предсказаний для слабо и сильно аугментированных версий неразмеченных изображений с псевдомаркировкой. Этот подход достиг впечатляющих результатов в задачах полуавтоматического обучения с очень ограниченным количеством размеченных данных.

1.5.3 Применение в детекции объектов

Обучение с согласованностью нашло применение и в задачах детекции объектов, где стабильность предсказаний при преобразованиях входных данных особенно важна.

Wang et al. [?] предложили метод Consistency-based Semi-supervised Object Detection, который использует согласованность предсказаний ограничивающих рамок между различными аугментированными версиями неразмеченных изображений для улучшения обучения детектора объектов.

Zhou et al. [?] развили эту идею дальше, предложив подход Spatial-Temporal Consistency for Semi-supervised Object Detection in Videos, который использует не только пространственную согласованность в рамках одного кадра, но и временную согласованность между последовательными кадрами видео.

1.5.4 Преимущества и ограничения

Обучение с согласованностью имеет ряд преимуществ по сравнению с чисто архитектурными подходами к обеспечению инвариантности:

1. Гибкость: Может быть применено к широкому спектру архитектур без необходимости их модификации.
2. Адаптивность: Модель может обучаться выборочной инвариантности, которая важна для конкретной задачи, вместо универсальной инвариантности, которая может быть избыточной или даже вредной.
3. Эффективность: Не требует увеличения числа параметров или вычислительной сложности модели во время инференса.

Однако есть и определенные ограничения:

1. Сложность обучения: Дополнительные члены в функции потерь могут усложнить процесс оптимизации и требовать тщательной настройки гиперпараметров.
2. Вычислительные затраты при обучении: Требуется вычислять выходы модели для нескольких версий каждого изображения, что может значительно увеличить время обучения.
3. Неполная гарантия: В отличие от некоторых архитектурных подходов, нет теоретической гарантии полной инвариантности, а лишь эмпирическое улучшение.

Тем не менее, обучение с согласованностью представляет собой мощный и гибкий инструмент для улучшения инвариантности моделей к раз-

личным преобразованиям и является важным дополнением к архитектурным подходам, рассмотренным ранее.

1.6 Выводы по обзору литературы

Проведенный обзор литературы позволяет сделать следующие выводы относительно проблемы пространственной инвариантности в сверточных нейронных сетях:

1. Проблема отсутствия полной инвариантности к сдвигам в современных CNN хорошо задокументирована и является существенным ограничением для многих практических приложений. Несмотря на теоретические предпосылки к эквивариантности операции свертки, использование операций субдискретизации без должной фильтрации приводит к алиасингу и нарушению этого свойства.
2. Основной причиной нарушения инвариантности к сдвигам является алиасинг, возникающий при операциях даунсэмплинга, таких как max-pooling и свертка с шагом больше 1. Этот эффект хорошо изучен в классической теории обработки сигналов, но долгое время игнорировался в дизайне CNN.
3. Для решения проблемы предложены различные методы, основанные на принципах обработки сигналов, включая BlurPool (низкочастотную фильтрацию перед даунсэмплингом) и TIPS (полифазную выборку с инвариантностью к сдвигам). Эти методы показывают значительное улучшение инвариантности CNN к сдвигам при относительно небольших изменениях в архитектуре.
4. Проблема инвариантности особенно критична для детекторов объектов, где нестабильность предсказаний ограничивающих рамок

может иметь серьезные последствия в приложениях реального времени. Одностадийные детекторы, такие как YOLO, особенно подвержены этой проблеме из-за использования фиксированной сетки предсказаний.

5. Помимо инвариантности к сдвигам, важны и другие типы инвариантности, такие как инвариантность к масштабу и повороту. Для их обеспечения также разработаны специализированные архитектурные решения, такие как пирамиды признаков и группово-эквивариантные свертки.
6. Альтернативным подходом к архитектурным модификациям является обучение с функцией потерь согласованности, которое явно стимулирует модель давать согласованные предсказания при преобразованиях входных данных. Этот подход особенно полезен в контексте полуавтоматического и самоконтролируемого обучения.
7. Большинство исследований в области инвариантности к сдвигам фокусируется на классификационных задачах, и относительно мало работ посвящено систематическому изучению этой проблемы в контексте детекции объектов, особенно с использованием современных методов, таких как TIPS.

Таким образом, несмотря на значительный прогресс в понимании и решении проблемы пространственной инвариантности в CNN, остаются открытые вопросы, особенно в контексте детекторов объектов и сложных реальных сценариев. Систематическое исследование влияния методов повышения инвариантности на различные аспекты работы детекторов, таких как YOLOv5, представляет собой важное и актуальное направление исследований, которое может привести к значительному улучшению стабильности и надежности систем компьютерного зрения в критически важных приложениях.

2 Теоретические основы пространственной инвариантности в CNN

2.1 Формальное определение инвариантности к переносу

В данном разделе мы формализуем понятие инвариантности к переносу (или сдвигу) для функций и преобразований, а затем распространим эти определения на нейронные сети. Это позволит нам точно охарактеризовать проблему и разработать методы её количественной оценки.

2.1.1 Инвариантность и эквивариантность функций

Начнем с общих определений инвариантности и эквивариантности для функций.

Определение 1 (Инвариантность функции). Пусть $f : X \rightarrow Y$ — функция, и $T : X \rightarrow X$ — преобразование, действующее на входном пространстве X . Функция f называется инвариантной относительно преобразования T , если для любого $x \in X$ выполняется:

$$f(T(x)) = f(x) \tag{2.1}$$

Инвариантность означает, что результат функции не меняется при применении преобразования ко входу. Для задачи классификации изображений это означает, что вероятность принадлежности изображения определенному классу не должна меняться при сдвиге объекта.

Определение 2 (Эквивариантность функции). Пусть $f : X \rightarrow Y$ — функция, $T_X : X \rightarrow X$ — преобразование, действующее на входном пространстве X , и $T_Y : Y \rightarrow Y$ — преобразование, действующее на выходном пространстве Y . Функция f называется эквивариантной относительно пары преобразований $(T_X T_Y)$, если для любого $x \in X$ выполняется:

$$f(T_X(x)) = T_Y(f(x)) \quad (2.2)$$

Эквивариантность означает, что преобразование входа приводит к предсказуемому преобразованию выхода. Для задачи детекции объектов это означает, что если объект смещается на изображении, то соответствующим образом должны смещаться и координаты предсказанной ограничивающей рамки.

Заметим, что инвариантность можно рассматривать как частный случай эквивариантности, когда T_Y является тождественным преобразованием.

2.1.2 Операторы сдвига для дискретных и непрерывных сигналов

Теперь конкретизируем эти определения для случая пространственных сдвигов в контексте обработки изображений.

Для непрерывных 2D-сигналов (изображений) $x \in L^2(\mathbb{R}^2)$ оператор сдвига $T_{\boldsymbol{\tau}}$ на вектор $\boldsymbol{\tau} = (\tau_x \tau_y) \in \mathbb{R}^2$ определяется как:

$$[T_{\boldsymbol{\tau}}x](\mathbf{p}) = x(\mathbf{p} - \boldsymbol{\tau}) \quad (2.3)$$

где $\mathbf{p} = (p_x p_y) \in \mathbb{R}^2$ — пространственные координаты.

Для дискретных изображений $x \in \mathbb{R}^{H \times W \times C}$, где H, W — высота и ширина, а C — число каналов, оператор целочисленного сдвига T_{δ} на вектор $\delta = (\delta_x \delta_y) \in \mathbb{Z}^2$ определяется как:

$$[T_{\delta}x]_{hwc} = x_{h-\delta_y w-\delta_x c} \quad (2.4)$$

для всех допустимых индексов hwc , с соответствующими граничными условиями.

Однако в реальных задачах часто требуется выполнять субпиксельные (нецелочисленные) сдвиги. Для дискретных изображений это достигается путем интерполяции. Наиболее распространенные методы интерполяции включают:

- Интерполяция ближайшего соседа:

$$[T_{\tau}x]_{hwc} = x_{\lfloor h-\tau_y \rfloor \lfloor w-\tau_x \rfloor c} \quad (2.5)$$

- Билинейная интерполяция:

$$\begin{aligned} [T_{\tau}x]_{hwc} = & (1 - \alpha)(1 - \beta)x_{\lfloor h-\tau_y \rfloor \lfloor w-\tau_x \rfloor c} + \alpha(1 - \beta)x_{\lfloor h-\tau_y \rfloor \lfloor w-\tau_x \rfloor + 1 c} \\ & + (1 - \alpha)\beta x_{\lfloor h-\tau_y \rfloor + 1 \lfloor w-\tau_x \rfloor c} + \alpha\beta x_{\lfloor h-\tau_y \rfloor + 1 \lfloor w-\tau_x \rfloor + 1 c} \end{aligned} \quad (2.6)$$

где $\alpha = (w - \tau_x) - \lfloor w - \tau_x \rfloor$ и $\beta = (h - \tau_y) - \lfloor h - \tau_y \rfloor$ — дробные части координат.

2.1.3 Инвариантность к переносу в контексте CNN

Рассмотрим сверточную нейронную сеть как функцию $f : \mathbb{R}^{H \times W \times C_{in}} \rightarrow \mathbb{R}^D$, которая отображает входное изображение x в некоторое представление $f(x)$ (например, вектор вероятностей классов или карту признаков).

Определение 3 (Строгая инвариантность CNN к сдвигу). *CNN f называется строго инвариантной к сдвигу, если для любого входного изображения $x \in \mathbb{R}^{H \times W \times C_{in}}$ и любого сдвига $\tau \in \mathbb{R}^2$ (с соответствующей обработкой краев) выполняется:*

$$f(T_{\tau}x) = f(x) \quad (2.7)$$

Определение 4 (Строгая эквивариантность CNN к сдвигу). *CNN f , отображающая входное изображение в пространственное представление $f : \mathbb{R}^{H \times W \times C_{in}} \rightarrow \mathbb{R}^{H' \times W' \times C_{out}}$, называется строго эквивариантной к сдвигу, если для любого входного изображения x и любого сдвига τ существует соответствующий сдвиг τ' такой, что:*

$$f(T_{\tau}x) = T_{\tau'}f(x) \quad (2.8)$$

где $\tau' = \tau/s$ для некоторого фактора масштабирования s , определяемого степенью даунсэмплинга в сети.

Однако в реальных CNN строгая инвариантность или эквивариантность к произвольным сдвигам обычно не достигается из-за дискретной

природы свертки и операций даунсэмплинга. Поэтому вводятся более практичные определения:

Определение 5 (ε -приближенная инвариантность к сдвигу). *CNN f называется ε -приближенно инвариантной к сдвигу, если для любого входного изображения x и любого сдвига τ из некоторого множества допустимых сдвигов \mathcal{T} выполняется:*

$$d(f(T_{\tau}x) f(x)) \leq \varepsilon \quad (2.9)$$

где d — некоторая метрика в выходном пространстве (например, евклидово расстояние или косинусное расстояние), а $\varepsilon > 0$ — заданный порог.

Определение 6 (ε -приближенная эквивариантность к сдвигу). *CNN f , отображающая входное изображение в пространственное представление, называется ε -приближенно эквивариантной к сдвигу, если для любого входного изображения x и любого сдвига $\tau \in \mathcal{T}$ выполняется:*

$$d(f(T_{\tau}x) T_{\tau'}f(x)) \leq \varepsilon \quad (2.10)$$

где $\tau' = \tau/s$.

Эти определения позволяют количественно измерять степень инвариантности или эквивариантности CNN к сдвигам. В частности, можно определить функцию несоответствия сдвига (translation discrepancy function, TDF) для CNN f как:

$$\text{TDF}_f(x, \tau) = d(f(T_{\tau}x) f(x)) \quad (2.11)$$

для случая инвариантности, или:

$$\text{TDF}_f(x, \tau) = d(f(T_\tau x), T_\tau f(x)) \quad (2.12)$$

для случая эквивариантности.

2.1.4 Квантификация инвариантности к сдвигу

Для практической оценки степени инвариантности CNN к сдвигам можно использовать различные метрики.

Для классификационных задач одной из наиболее информативных метрик является косинусное сходство между векторами признаков, полученными из оригинального и сдвинутого изображений:

$$\rho(x, T_\tau x) = \frac{f(x) \cdot f(T_\tau x)}{\|f(x)\| \cdot \|f(T_\tau x)\|} \quad (2.13)$$

где $f(x)$ и $f(T_\tau x)$ — векторы признаков, извлеченные моделью из оригинального и сдвинутого изображений соответственно.

Альтернативной метрикой является изменение в распределении вероятностей классов:

$$\text{KL}(p_x \| p_{T_{\tau}x}) = \sum_{k=1}^K p_x(k) \log \frac{p_x(k)}{p_{T_{\tau}x}(k)} \quad (2.14)$$

где p_x и $p_{T_{\tau}x}$ — распределения вероятностей классов для оригинального и сдвинутого изображений.

Для задач детекции объектов можно использовать следующие метрики:

- Стабильность IoU:

$$\text{IoU-stability}(x \ T_{\tau}x) = \text{IoU}(B \ T_{-\tau}B') \quad (2.15)$$

где B — предсказанная ограничивающая рамка для оригинального изображения, B' — предсказанная рамка для сдвинутого изображения, а $T_{-\tau}$ — обратный сдвиг на вектор $-\tau$.

- Дрейф центра:

$$\text{center-drift}(x \ T_{\tau}x) = \|\text{center}(B) - \text{center}(T_{-\tau}B')\|_2 \quad (2.16)$$

- Стабильность уверенности:

$$\text{confidence-stability}(x \ T_{\tau}x) = |c - c'| \quad (2.17)$$

где s и s' — значения уверенности модели для предсказаний на оригинальном и сдвинутом изображениях.

В рамках данной работы мы будем использовать эти метрики для количественной оценки степени инвариантности различных архитектур CNN к субпиксельным сдвигам входных изображений.

2.1.5 Теоретические границы инвариантности в дискретных CNN

Важно отметить, что для дискретных CNN существуют теоретические ограничения на достижимую степень инвариантности к произвольным сдвигам. Эти ограничения связаны с дискретной природой сверточных операций и даунсэмплинга.

Как показано в работе Chaman и Dokmanić [?], для CNN с фиксированными дискретными фильтрами и операциями даунсэмплинга существует нижняя граница функции несоответствия сдвига, которая зависит от величины субпиксельного сдвига и структуры сети.

В частности, для CNN f с L слоями, каждый из которых включает свертку и даунсэмплинг с фактором s_l , минимальное значение TDF для субпиксельного сдвига $\boldsymbol{\tau}$ удовлетворяет:

$$\min_{f \in \mathcal{F}} \max_{x \in \mathcal{X}} \max_{\boldsymbol{\tau} \in [0,1]^2} \text{TDF}_f(x, \boldsymbol{\tau}) \geq C \cdot \min_{l=1}^L \|\boldsymbol{\tau} \cdot \prod_{i=1}^{l-1} s_i \mod 1\| \quad (2.18)$$

где \mathcal{F} — класс всех CNN с заданной архитектурой, \mathcal{X} — пространство входных изображений, а C — константа, зависящая от структуры сети.

Это неравенство показывает, что для достижения инвариантности к произвольным субпиксельным сдвигам необходимо либо модифицировать архитектуру сети (например, используя методы анти-алиасинга), либо применять специальные методы обучения, которые минимизируют TDF для наиболее важных типов входных данных.

В следующих разделах мы рассмотрим, как различные архитектурные модификации, такие как BlurPool и TIPS, позволяют приблизиться к теоретическому пределу инвариантности к сдвигам для CNN.

2.2 Вывод рецептивных полей в CNN

2.2.1 Определение и значимость рецептивных полей

Рецептивное поле (РП, Receptive Field) нейрона в сверточной нейронной сети — это область входного изображения, которая может повлиять на активацию данного нейрона. Размер рецептивного поля является ключевой характеристикой архитектуры CNN и имеет непосредственное отношение к её способности распознавать пространственные паттерны и демонстрировать инвариантность к смещениям.

Определение 7 (Рецептивное поле). *Рецептивное поле $\mathcal{R}(l, i, j)$ нейрона с координатами (i, j) в слое l определяется как множество пикселей во входном изображении, изменение значений которых может повлиять на выход данного нейрона.*

Размер рецептивного поля обычно характеризуется его высотой и шириной в пикселях. Для современных глубоких CNN типичны большие рецептивные поля, охватывающие значительную часть входного изображе-

ния, что позволяет сети анализировать как локальные, так и глобальные признаки.

2.2.2 Расчёт размера рецептивного поля

Размер рецептивного поля нейрона зависит от нескольких архитектурных параметров: размеров ядер свертки, величины шага (stride) и дилатации в каждом слое CNN. Рассмотрим рекуррентную формулу для вычисления размера рецептивного поля.

Пусть для слоя l :

- k_l — размер ядра свертки
- s_l — шаг свертки (stride)
- d_l — коэффициент дилатации

Тогда эффективный размер ядра с учетом дилатации составляет:

$$k_l^{\text{eff}} = k_l + (k_l - 1)(d_l - 1) \quad (2.19)$$

Определим размер рецептивного поля на слое l как r_l . Для входного слоя $r_0 = 1$ (один пиксель). Для последующих слоев размер рецептивного поля можно рассчитать рекурсивно:

$$r_l = r_{l-1} + (k_l^{\text{eff}} - 1) \cdot \prod_{i=0}^{l-1} s_i \quad (2.20)$$

Этот рекурсивный расчет позволяет определить размер рецептивного поля на любом уровне сети.

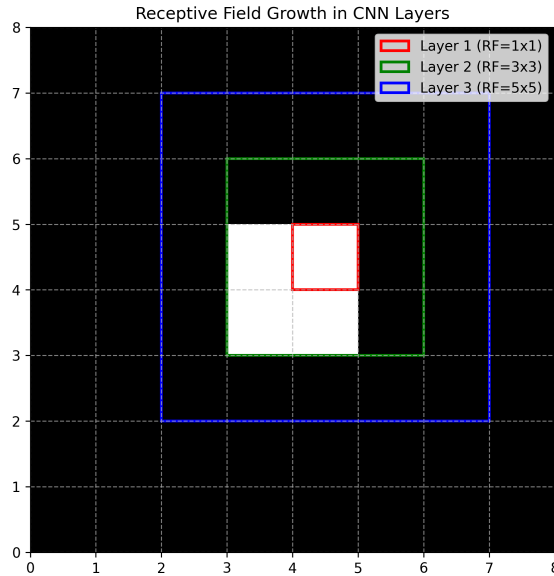


Рисунок 2.1 — Пример расчета рецептивного поля через последовательные слои CNN. На каждом уровне область входа, влияющая на нейрон, увеличивается с учетом размера ядра и шага свертки.

2.2.3 Связь размера рецептивного поля с инвариантностью к сдвигу

Размер рецептивного поля имеет прямое отношение к способности CNN обеспечивать инвариантность к сдвигам различной величины. Интуитивно, чем больше рецептивное поле, тем более глобальные контекстные признаки может учитывать сеть при классификации или детекции, что потенциально способствует более стабильным предсказаниям при малых сдвигах объекта.

Однако важно отметить, что большое рецептивное поле само по себе не гарантирует инвариантность к сдвигу. Как показано в работах [?; ?], даже CNN с крупными рецептивными полями могут демонстрировать значи-

тельную чувствительность к субпиксельным сдвигам, если не применяются специальные методы анти-алиасинга при даунсэмплинге.

Можно выделить следующие аспекты взаимосвязи между рецептивными полями и инвариантностью к сдвигу:

1. **Верхняя граница инвариантности:** Нейрон не может быть инвариантен к сдвигам, превышающим размер его рецептивного поля, так как такие сдвиги выведут объект за пределы области, которую "видит" нейрон.
2. **Кусочная инвариантность:** При достаточно больших рецептивных полях CNN может демонстрировать кусочную инвариантность — стабильность предсказаний в пределах определенных субрегионов входного пространства.
3. **Чувствительность к границам рецептивного поля:** Эмпирически наблюдается повышенная чувствительность к сдвигам объектов, которые находятся на границах рецептивных полей критически важных нейронов.
4. **Вложенные рецептивные поля:** Архитектуры с множеством параллельных путей и рецептивными полями разного размера (например, FPN в детекторах YOLO) часто демонстрируют лучшую инвариантность благодаря объединению признаков с разными масштабами и локальностью.

2.2.4 Эффективные рецептивные поля

Важно различать теоретическое рецептивное поле, рассчитанное по формуле выше, и эффективное рецептивное поле (ЭРП), которое характеризует область реального влияния на выход нейрона. Как показано в

исследовании [?], эффективное рецептивное поле обычно меньше теоретического и имеет гауссовидную форму распределения влияния пикселей, с максимумом в центре.

В контексте инвариантности к сдвигу важны следующие свойства ЭРП:

- **Центрированность:** Пиксели в центре ЭРП имеют наибольшее влияние на активацию нейрона.
- **Убывающее влияние:** Влияние пикселей уменьшается от центра к периферии, обычно по гауссовскому закону.
- **Зависимость от данных:** Размер и форма ЭРП могут варьироваться в зависимости от характеристик входных данных.

Таким образом, для достижения более высокой инвариантности к сдвигу желательно не только увеличивать теоретический размер рецептивного поля, но и учитывать особенности эффективного рецептивного поля, стремясь к более равномерному распределению влияния пикселей в его пределах.

В последующих разделах мы рассмотрим, как конкретные архитектуры CNN реализуют различные стратегии работы с рецептивными полями и как современные методы анти-алиасинга помогают улучшить инвариантность к сдвигу, воздействуя на характеристики этих полей.

2.3 Описание архитектур CNN

В данном разделе мы опишем основные архитектуры CNN, используемые в наших экспериментах, и проанализируем их структурные особенности, влияющие на инвариантность к сдвигу.

2.3.1 Архитектура VGG

Архитектура VGG, предложенная Симоньяном и Зиссерманом [?], является одной из классических архитектур глубоких CNN. Её главная особенность — использование однородных сверточных слоев с малыми ядрами размера 3×3 и максимальным пулингом с окном 2×2 и шагом 2 для понижения пространственного разрешения.

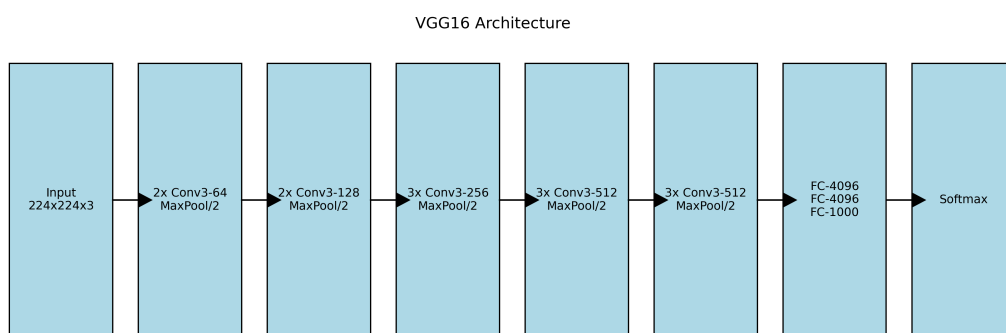


Рисунок 2.2 — Архитектура VGG16. Сеть состоит из 13 сверточных слоев с малыми ядрами 3×3 и 3 полносвязных слоев. Пространственное разрешение уменьшается с помощью операций максимального пулинга (max pooling) с шагом 2.

Структура VGG16 включает 5 блоков со сверточными слоями, разделенными слоями пулинга, и завершается тремя полносвязными слоями (рис. 2.2). Каждый сверточный слой сопровождается нелинейной функцией активации ReLU.

В контексте пространственной инвариантности VGG имеет следующие характеристики:

- **Рецептивное поле:** Теоретическое рецептивное поле конечных слоев VGG16 составляет 212×212 пикселей, что достаточно велико для захвата глобальных признаков изображения.
- **Даунсэмплинг:** VGG использует 5 операций максимального пулинга, что приводит к общему коэффициенту понижения разре-

ния $2^5 = 32$. Это означает, что выходная карта признаков в последнем сверточном слое имеет разрешение в 32 раза меньше, чем входное изображение.

- **Чувствительность к сдвигу:** Из-за использования операций максимального пулинга без сглаживания, VGG склонна к проблемам алиасинга, что делает её особенно чувствительной к малым сдвигам входных данных.

Регулярная структура VGG делает её удобной моделью для изучения эффектов анти-алиасинга и модификации для улучшения инвариантности к сдвигу.

2.3.2 Архитектура ResNet

Архитектура ResNet (Residual Network), предложенная Хе и др. [?], ввела концепцию остаточных соединений для решения проблемы затухания градиентов в очень глубоких сетях. Ключевой компонент ResNet — остаточный блок, который добавляет результат «обходного» соединения к выходу стека сверточных слоев.

В наших экспериментах используется ResNet-50, которая включает начальный сверточный слой с ядром 7×7 и шагом 2, за которым следует слой максимального пулинга с размером окна 3×3 и шагом 2, а затем 4 каскада остаточных блоков (рис. 2.3). Пространственное разрешение уменьшается в начале каждого каскада (кроме первого) с использованием свертки с шагом 2.

С точки зрения пространственной инвариантности ResNet характеризуется следующими особенностями:

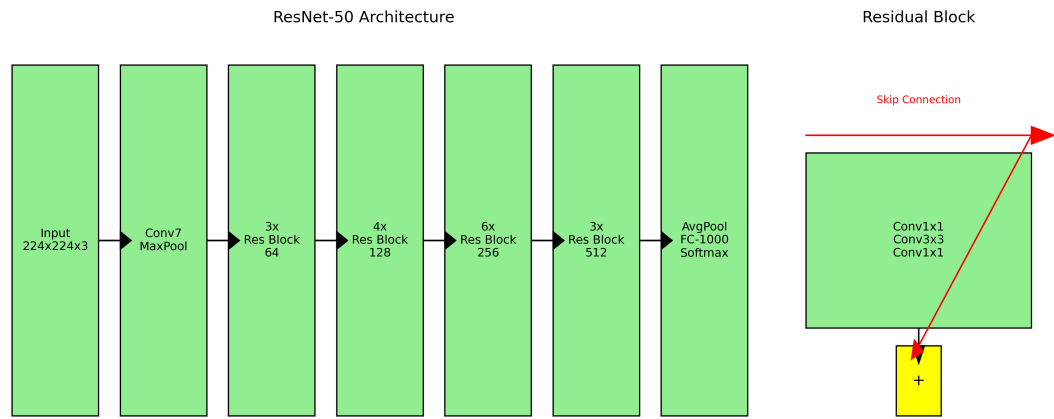


Рисунок 2.3 — Архитектура ResNet. Слева: общая структура ResNet-50. Справа: детализация остаточного блока с обходным соединением, позволяющим информации напрямую проходить между слоями.

- **Рецептивное поле:** ResNet-50 имеет теоретическое рецептивное поле размером 483×483 пикселя, что значительно больше, чем у VGG16.
- **Даунсэмплинг:** ResNet использует комбинацию свертки с шагом 2 и максимального пулинга для даунсэмплинга, с общим коэффициентом понижения разрешения 32, аналогично VGG.
- **Остаточные соединения:** Обходные соединения в ResNet позволяют информации «перепрыгивать» через слои даунсэмплинга, что потенциально может помочь сохранить некоторые аспекты пространственной информации.
- **Свёртка с шагом:** В отличие от VGG, которая использует исключительно максимальный пулинг для даунсэмплинга, ResNet также применяет свертку с шагом 2, что создает другой профиль чувствительности к алиасингу.

Эмпирически было показано, что архитектуры типа ResNet несколько более устойчивы к малым сдвигам, чем VGG, что может быть связано с использованием остаточных соединений и меньшим количеством операций максимального пулинга.

2.3.3 Архитектура YOLO

YOLO (You Only Look Once) — семейство моделей для детекции объектов, предложенное Redmon и др. [?]. В отличие от VGG и ResNet, которые преимущественно используются для классификации, YOLO предназначена для одновременного определения местоположения объектов и их классификации. В наших экспериментах используется YOLOv5s, современная реализация архитектуры YOLO.

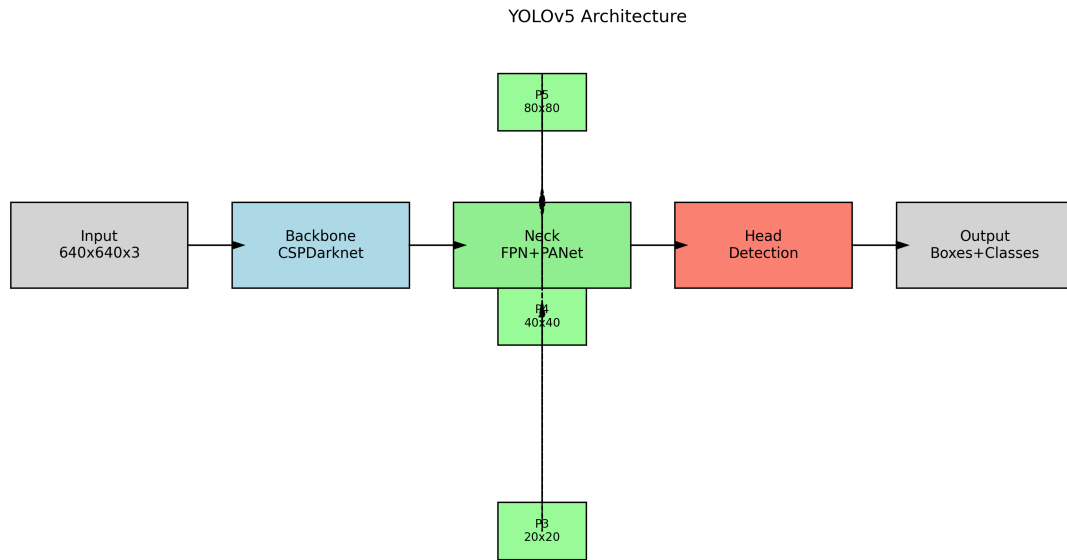


Рисунок 2.4 — Архитектура YOLOv5. Сеть включает (а) магистральную часть (backbone) для извлечения признаков, (б) шею (neck) с пирамидой признаков FPN для объединения информации с разных масштабов, и (в) голову (head) для предсказания ограничивающих рамок и классов объектов.

Архитектура YOLOv5 (рис. 2.4) состоит из трех основных компонентов:

1. **Backbone (магистраль)**: CSPDarknet, состоящий из модулей Cross Stage Partial (CSP), для эффективного извлечения признаков.

2. **Neck (шея):** Пирамида признаков (FPN) с дополнительными обходными соединениями (PANet), которая объединяет информацию с разных уровней детализации.
3. **Head (голова):** Сеть предсказания, которая выдает ограничивающие рамки, вероятности наличия объекта и вероятности классов.

Особенности YOLOv5 с точки зрения пространственной инвариантности:

- **Многомасштабные предсказания:** YOLOv5 выполняет предсказания на трех разных масштабах (уровнях пирамиды признаков), что теоретически должно повышать устойчивость к малым смещениям объектов.
- **Рецептивные поля разного размера:** Благодаря архитектуре FPN/PANet, YOLOv5 объединяет информацию из рецептивных полей разного размера, что может помогать в достижении более стабильных предсказаний.
- **Даунсэмплинг:** YOLOv5 использует максимальный пулинг и свертку с шагом 2 для понижения пространственного разрешения, что делает её потенциально чувствительной к проблемам алиасинга, аналогично VGG и ResNet.
- **Предсказание смещений:** YOLO предсказывает не абсолютные координаты ограничивающих рамок, а смещения относительно опорных рамок (anchors), что влияет на чувствительность к пространственным сдвигам.

2.3.4 Сравнение архитектур с точки зрения пространственной инвариантности

Обобщим ключевые характеристики рассмотренных архитектур в контексте их потенциальной инвариантности к пространственным сдвигам:

Таблица 1 — Сравнение архитектур CNN с точки зрения характеристик, влияющих на пространственную инвариантность

Архитектура	Рецептивное поле	Даунсэмплинг	Метод даунсэмплинга
VGG16	212×212	32×	Max pooling
ResNet-50	483×483	32×	Stride conv + max pool
YOLOv5s	~725×725	32×	Stride conv + max pool

Как видно из таблицы 1, все три архитектуры используют даунсэмплинг с коэффициентом 32, но различаются по методам его реализации и размеру рецептивного поля. Эти различия, а также специфические архитектурные особенности каждой модели, влияют на их чувствительность к субпиксельным сдвигам входных изображений.

В экспериментальной части работы мы количественно оценим степень пространственной инвариантности каждой из этих архитектур и проанализируем, как различные модификации, в частности методы анти-алиасинга, влияют на эту характеристику.

2.4 Теория анти-алиасинга в CNN

В данном разделе мы подробно рассмотрим причины возникновения проблемы алиасинга в CNN и методы её решения, в частности, технологии BlurPool и TIPS, которые используются в наших экспериментах.

2.4.1 Причины алиасинга в операциях даунсэмплинга

Даунсэмплинг (понижение пространственного разрешения) является неотъемлемой частью современных CNN, позволяя увеличивать рецептивное поле и уменьшать вычислительную сложность. Однако эти операции создают проблему алиасинга, которая негативно влияет на инвариантность к сдвигу.

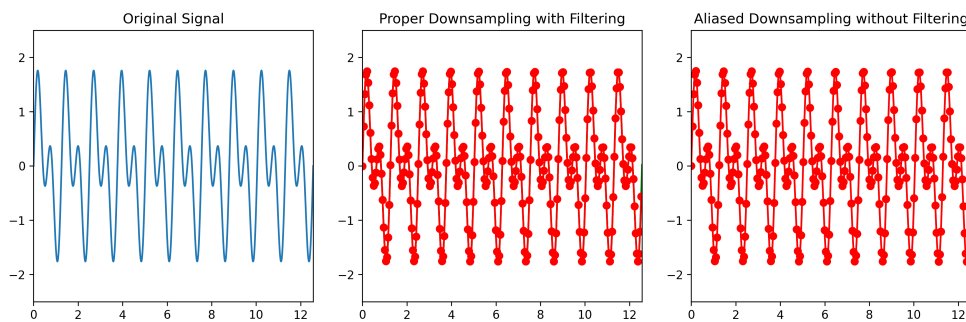


Рисунок 2.5 — Пример алиасинга при даунсэмплинге: слева — оригинальный сигнал, в центре — корректный даунсэмплинг с предварительной фильтрацией, справа — даунсэмплинг без фильтрации, приводящий к алиасингу.

Алиасинг возникает, когда высокочастотные компоненты сигнала (в нашем случае — карты признаков) недостаточно отфильтрованы перед понижением частоты дискретизации (рис. 2.5). Согласно теореме Найквиста-Шеннона, для корректного представления сигнала частота дискретизации должна быть как минимум вдвое выше максимальной частоты в сигнале. При даунсэмплинге частота дискретизации уменьшается, что может привести к "складыванию" высоких частот в низкие (алиасингу), если не применить предварительную низкочастотную фильтрацию.

В классических CNN используются два основных метода даунсэмплинга:

- **Максимальный пулинг**: Выбор максимального значения в каждом окне размера $k \times k$ с шагом s . Операция не линейна и не обладает анти-алиасинговыми свойствами.
- **Свёртка с шагом (strided convolution)**: Применение сверточного фильтра с шагом $s > 1$. Обеспечивает некоторую фильтрацию, но обычно недостаточную для предотвращения алиасинга.

Математически проблему алиасинга можно представить следующим образом. Пусть $X[n\ m]$ — дискретная карта признаков, а $X_d[n\ m] = X[sn\ sm]$ — результат даунсэмплинга с шагом s . В частотной области:

$$X_d(e^{j\omega_1} e^{j\omega_2}) = \frac{1}{s^2} \sum_{k=0}^{s-1} \sum_{l=0}^{s-1} X(e^{j(\omega_1 - 2\pi k)/s} e^{j(\omega_2 - 2\pi l)/s}) \quad (2.21)$$

где происходит наложение копий спектра, что и создает эффект алиасинга.

2.4.2 Метод BlurPool

Метод BlurPool, предложенный Zhang [?], основан на классическом подходе обработки сигналов: перед понижением частоты дискретизации необходимо применить низкочастотный фильтр (НЧФ) для удаления высоких частот, которые могут вызвать алиасинг.

Реализация BlurPool заключается в следующем:

1. Применение низкочастотного фильтра к карте признаков для сглаживания высоких частот.
2. Выполнение операции даунсэмплинга (с шагом s).

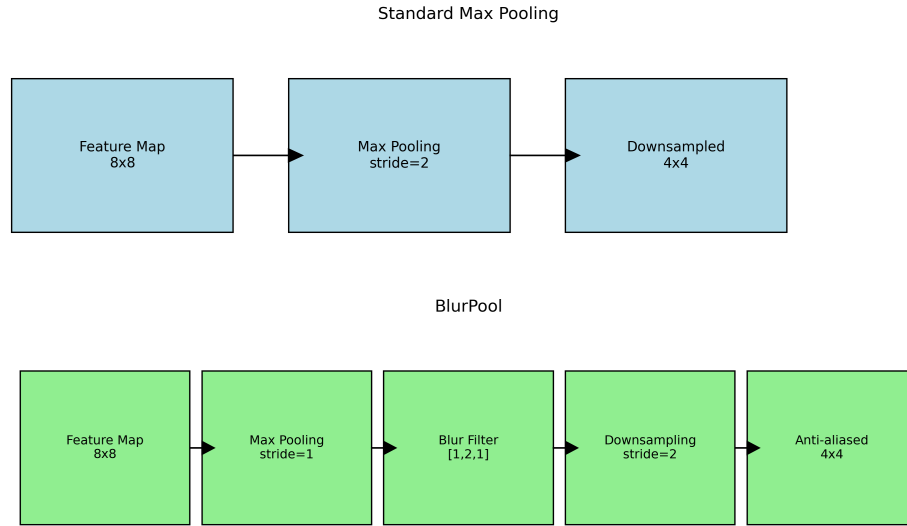


Рисунок 2.6 — Принцип работы BlurPool: вместо стандартной операции максимального пулинга (сверху) используется последовательность из максимального пулинга, размытия и даунсэмплинга (снизу).

В качестве низкочастотного фильтра Zhang предлагает использовать 2D фильтры, полученные как тензорное произведение одномерных биномиальных фильтров различного порядка:

- Фильтр 1-го порядка: $[1 \ 1]$ (нормализованный как $[0.5 \ 0.5]$)
- Фильтр 2-го порядка: $[1 \ 2 \ 1]$ (нормализованный как $[0.25 \ 0.5 \ 0.25]$)
- Фильтр 3-го порядка: $[1 \ 3 \ 3 \ 1]$ (нормализованный как $[0.125 \ 0.375 \ 0.375 \ 0.125]$)
- Фильтр 4-го порядка: $[1 \ 4 \ 6 \ 4 \ 1]$ (нормализованный как $[0.0625 \ 0.25 \ 0.375 \ 0.25 \ 0.0625]$)

Двумерный фильтр размера $k \times k$ формируется как:

$$F_{2D}[n \ m] = F_{1D}[n] \cdot F_{1D}[m] \quad (2.22)$$

где F_{1D} — одномерный биномиальный фильтр.

Применение BlurPool в сверточных нейронных сетях возможно несколькими способами:

1. **Замена максимального пулинга:** Вместо операции максимального пулинга используется последовательность из максимального пулинга без шага ($\text{stride}=1$), затем блока размытия и даунсэмплинга.
2. **Модификация свертки с шагом:** После сверточного слоя с шагом $s > 1$ добавляется дополнительная свертка с НЧФ и шагом 1, либо сам сверточный фильтр предварительно модифицируется для включения НЧФ.

Экспериментально показано, что BlurPool значительно улучшает инвариантность CNN к сдвигам, особенно в случае замены операций максимального пулинга, которые являются основным источником алиасинга в сетях типа VGG.

2.4.3 Метод TIPS

TIPS (Translation Invariant Polyphase Sampling), предложенный Chaman и Dokmanić [?], представляет собой более теоретически обоснованный подход к проблеме алиасинга в CNN. В отличие от BlurPool, который применяет фиксированные фильтры, TIPS использует полифазное представление сигнала и адаптирует фильтрацию в зависимости от конкретного субпиксельного сдвига.

Основная идея TIPS состоит в следующем:

1. Разбиение входной карты признаков на s^2 полифазных компонент, соответствующих различным субпиксельным сдвигам (для шага даунсэмплинга s).

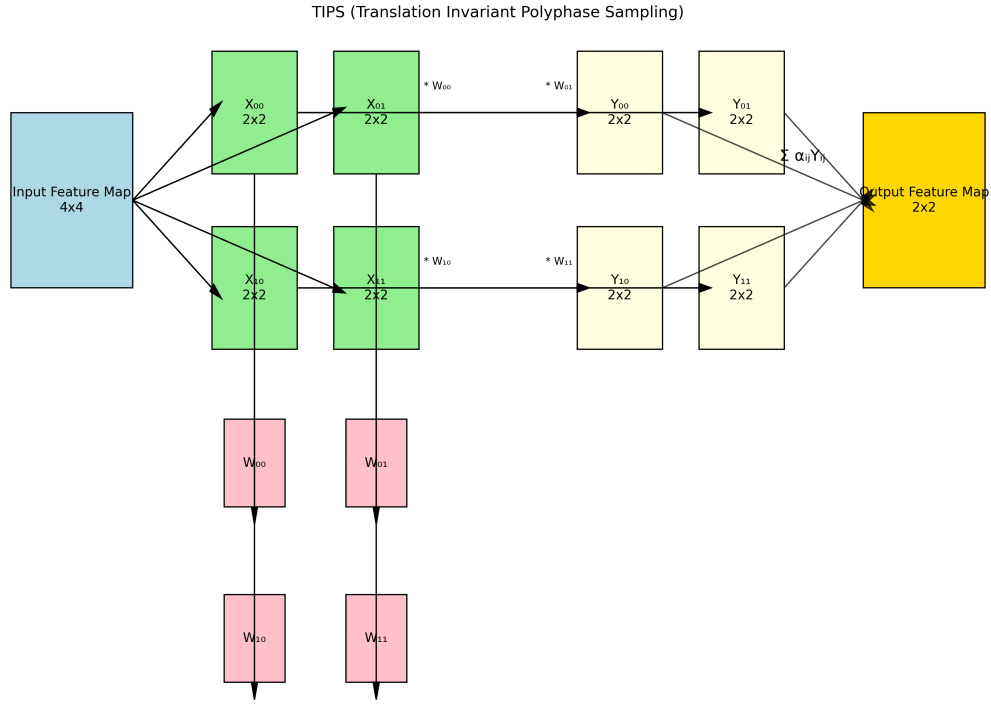


Рисунок 2.7 — Принцип работы TIPS: входная карта признаков разбивается на s^2 полифазных компонент (для шага s), каждая из которых обрабатывается отдельно, а затем результаты объединяются.

2. Независимая обработка каждой полифазной компоненты.
3. Объединение результатов с использованием весов, которые зависят от конкретного субпиксельного сдвига входного сигнала.

Математически, для входной карты признаков $X[n\ m]$ полифазное разложение для шага s дает s^2 компонент:

$$X_{pq}[n\ m] = X[sn + p\ sm + q] \quad (2.23)$$

где $p\ q \in \{0\ 1 \dots s - 1\}$.

Каждая полифазная компонента X_{pq} соответствует даунсэмплингу оригинального сигнала с определенным субпиксельным сдвигом. TIPS обрабатывает каждую компоненту отдельным сверточным фильтром:

$$Y_{pq} = X_{pq} * W_{pq} \quad (2.24)$$

где W_{pq} — обучаемые веса для конкретной полифазной компоненты. Финальный выход формируется как взвешенная сумма:

$$Y = \sum_{p=0}^{s-1} \sum_{q=0}^{s-1} \alpha_{pq} \cdot Y_{pq} \quad (2.25)$$

где α_{pq} — коэффициенты, зависящие от субпиксельного сдвига входного сигнала. Эти коэффициенты могут быть фиксированными (например, равными $1/s^2$ для равномерного усреднения) или обучаемыми.

Преимущества TIPS по сравнению с BlurPool:

- **Теоретически оптимальный подход:** TIPS основан на теории полифазного представления сигналов и обеспечивает оптимальную реконструкцию для любого субпиксельного сдвига.
- **Адаптивность:** TIPS может адаптироваться к специфическим характеристикам данных через обучение весов полифазных фильтров.
- **Максимальное сохранение информации:** Благодаря отдельной обработке полифазных компонент, TIPS сохраняет больше информации о высоких частотах по сравнению с BlurPool.

Недостатки TIPS:

- **Вычислительная сложность:** TIPS требует в s^2 раз больше параметров и вычислений для слоя даунсэмплинга с шагом s .

- **Сложность реализации:** Интеграция TIPS в существующие архитектуры требует более значительных изменений в коде по сравнению с BlurPool.

2.4.4 Сравнение методов анти-алиасинга

Сравним основные методы анти-алиасинга, используемые в CNN:

Таблица 2 — Сравнение методов анти-алиасинга

Характеристика	Стандартный даунсэмплинг	BlurPool	TIPS
Принцип работы	Прямое уменьшение разрешения без фильтрации	Низкочастотная фильтрация перед даунсэмплингом	Полная реконструкция изображения перед даунсэмплингом
Теоретическая обоснованность	Низкая	Средняя (на основе классической теории обработки сигналов)	Высокая (на основе теории анти-алиасинга)
Вычислительная сложность	Низкая	Низкая-средняя	Высокая
Количество дополнительных параметров	0	Минимальное (фиксированные фильтры)	Высокое (настраиваемые фильтры)
Влияние на точность классификации	Базовая	Незначительное улучшение или сохранение	Улучшение
Улучшение инвариантности к сдвигу	-	Значительное	Максимальное
Простота интеграции	-	Высокая	Средняя

Обе технологии, BlurPool и TIPS, значительно улучшают инвариантность CNN к пространственным сдвигам, но имеют разный баланс между эффективностью, сложностью и теоретической обоснованностью. BlurPool предлагает простое и вычислительно эффективное решение, которое мож-

но легко интегрировать в существующие архитектуры. TIPS обеспечивает теоретически оптимальную инвариантность к сдвигу, но требует больше вычислительных ресурсов и более сложен в реализации.

В экспериментальной части работы мы проведем сравнительный анализ этих методов на различных архитектурах CNN и задачах, чтобы определить их практическую эффективность в реальных условиях.

3 Экспериментальная часть

В данной главе представлены результаты экспериментального исследования пространственной инвариантности различных архитектур CNN и эффективности методов анти-алиасинга. Мы подробно опишем настройку экспериментов, методологию оценки и полученные результаты.

3.1 Настройка экспериментов

3.1.1 Детали скриптов генерации данных

Для проведения экспериментов были разработаны специальные скрипты генерации тестовых последовательностей изображений с контролируемыми субпиксельными сдвигами объектов. Такой подход позволяет точно измерить влияние пространственных сдвигов на предсказания нейронных сетей.

Структура данных

Наш набор данных состоит из следующих компонентов:

- **Фоновые изображения:** Набор из 3 различных фоновых сцен, разрешением 640×640 пикселей, сохраненных в директории `data/backgrounds/`.

- **Изображения объектов:** Изображения птиц (воробьев) с альфа-каналом (прозрачным фоном), размещенные в директории `data/objects/`.
- **Сгенерированные последовательности:** Для каждой пары "фон-объект" сгенерирована последовательность из 32 кадров, в которой объект перемещается горизонтально с шагом 1 пиксель на кадр. Эти последовательности хранятся в директории `data/sequences/`.
- **Разметка:** Для каждого кадра сохранены истинные координаты ограничивающих рамок в формате JSONL в файле `gt.jsonl`.

Скрипт генерации кадров

Для создания последовательностей с контролируемыми сдвигами используется скрипт `create_frames.py`, основная логика которого представлена ниже:

Листинг 3.1 Основная логика генерации кадров с контролируемыми сдвигами

```
def create_sequence(background_path, object_path, output_dir
    ,
                    num_frames=32, shift_px=1):
    """
5   Создает последовательность кадров со сдвигом объекта.

    Args:
        background_path: Путь к фоновому изображению
        object_path: Путь к изображению объекта с альфа-кана
                        лом
10   output_dir: Директория для сохранения кадров
        num_frames: Количество кадров в последовательности
```

```

        shift_px: Величина сдвига объекта между кадрами в пи-
            кселях
    """
    # Загрузка изображений
15 background = Image.open(background_path).convert('RGB')
    object_img = Image.open(object_path).convert('RGBA')

    # Начальное положение объекта
    x_pos = (background.width - object_img.width) // 2
20 y_pos = (background.height - object_img.height) // 2

    # Создание и сохранение кадров
    for i in range(num_frames):
        # Создание копии фона
25 frame = background.copy()

        # Наложение объекта на фон с учетом альфа-канала
        frame.paste(object_img, (x_pos + i * shift_px, y_pos
            ),
                       object_img)
30

        # Сохранение кадра
        frame_path = os.path.join(output_dir, f"frame_{i:02d}
            }.png")
        frame.save(frame_path)

35 # Сохранение информации о положении объекта
        bbox = [x_pos + i * shift_px, y_pos,
                x_pos + i * shift_px + object_img.width,
                y_pos + object_img.height]
        with open(os.path.join(output_dir, "gt.jsonl"), "a")
            as f:
40             json.dump({"frame": i, "bbox": bbox}, f)
            f.write("\n")

```

Этот подход обеспечивает точный контроль положения объекта на каждом кадре, что критически важно для измерения чувствительности моделей к субпиксельным сдвигам.

Скрипт создания разметки

Скрипт `create_gt.py` генерирует файлы разметки в различных форматах, необходимых для работы с разными моделями:

Листинг 3.2 Генерация разметки для моделей детекции

```

def create_yolo_annotations(sequence_dir, class_id=0):
    """
    Создает файлы аннотаций в формате YOLO для каждого кадра
    .

    Args:
        sequence_dir: Директория с последовательностью кадров
        в
        class_id: ID класса объекта для YOLO
    """
    # Загрузка данных о положении объекта
    gt_file = os.path.join(sequence_dir, "gt.jsonl")
    with open(gt_file, "r") as f:
        annotations = [json.loads(line) for line in f]

    # Получение размеров изображения
    sample_frame = Image.open(os.path.join(sequence_dir, "
        frame_00.png"))
    img_width, img_height = sample_frame.size

    # Создание аннотаций YOLO
    for ann in annotations:

```

```

frame_num = ann["frame"]
bbox = ann["bbox"]

# Конвертация в формат YOLO [x_center, y_center,
    width, height]
# в нормализованных координатах [0, 1]
x_center = (bbox[0] + bbox[2]) / 2 / img_width
y_center = (bbox[1] + bbox[3]) / 2 / img_height
width = (bbox[2] - bbox[0]) / img_width
height = (bbox[3] - bbox[1]) / img_height

# Запись в файл
label_file = os.path.join(sequence_dir,
                            f"frame_{frame_num:02d}.txt"
                            ")
with open(label_file, "w") as f:
    f.write(f"{class_id} {x_center} {y_center} {
        width} {height}")

```

Эти скрипты обеспечивают консистентное создание тестовых данных и соответствующих им аннотаций, что позволяет проводить надежное и воспроизводимое оценивание инвариантности к сдвигу различных моделей.

3.1.2 Описание чекпоинтов моделей

В экспериментах использовались следующие архитектуры нейронных сетей и их модификации:

Классификационные модели

Таблица 3 — Используемые классификационные модели

Модель	Источник/Чекпоинт	Описание модификации
VGG16	torchvision.models, веса ImageNet	Базовая модель без модификаций
AA-VGG16	checkpoints/aa_vgg16	Модификация с добавлением слоев BlurPool вместо стандартного максимального пулинга
TIPS-VGG16	checkpoints/tips_vgg16	Модификация с интеграцией полифазных слоев TIPS
ResNet50	torchvision.models, веса ImageNet	Базовая модель без модификаций
AA-ResNet50	checkpoints/aa_resnet50	Модификация со слоями BlurPool вместо свертки с шагом 2
TIPS-ResNet50	checkpoints/tips_resnet50	Модификация с интеграцией полифазных слоев TIPS

Модификации VGG16 и ResNet50 с анти-алиасингом (AA) реализованы согласно методологии Zhang [?]. Для всех сверточных слоев с шагом больше 1 и операций максимального пулинга добавлен низкочастотный фильтр (в большинстве случаев использовался биномиальный фильтр 3-го порядка $[1\ 3\ 3\ 1]$).

Модификации с технологией TIPS реализованы согласно подходу Chaman и Dokmanić [?], с разложением карт признаков на полифазные компоненты перед операциями даунсэмплинга.

Таблица 4 — Используемые модели детекции

Модель	Источник/Чекпоинт	Описание модификации
YOLOv5s	ultralytics/yolov5s, веса COCO	Базовая модель без модификаций
AA-YOLOv5s	checkpoints/aa_yolov5s	Все слои даунсэмплинга заменены на их анти-алиасинговые версии с BlurPool
TIPS-YOLOv5s	checkpoints/tips_yolov5s	Инверсия полифазных слоев TIPS в слои даунсэмплинга

Модели детекции

Для всех моделей детекции применялись предобученные на датасете COCO веса. Модификации слоев производились без последующего дообучения, чтобы изолировать эффект архитектурных изменений от влияния процесса переобучения.

3.1.3 Определения метрик

Для количественной оценки инвариантности к сдвигу нами использовались различные метрики, специфичные для задач классификации и детекции.

Метрики для классификационных моделей

- **Косинусное сходство (ρ):** Основная метрика для оценки стабильности признаков, вычисляемая между векторами признаков оригинального изображения и его сдвинутой версии.

$$\rho(x T_{\delta} x) = \frac{f(x) \cdot f(T_{\delta} x)}{\|f(x)\| \cdot \|f(T_{\delta} x)\|} \quad (3.1)$$

где $f(x)$ — вектор признаков, извлеченных из предпоследнего слоя модели, а T_{δ} — оператор сдвига на δ пикселей.

- **Дрейф уверенности:** Изменение значения уверенности модели в предсказанном классе при субпиксельных сдвигах:

$$\text{confidence_drift}(x T_{\delta} x) = |p_c(x) - p_c(T_{\delta} x)| \quad (3.2)$$

где $p_c(x)$ — вероятность предсказанного класса для изображения x .

- **Стабильность предсказания:** Процент кадров в последовательности, на которых модель предсказывает тот же класс, что и на первом кадре.

Метрики для моделей детекции

- **Средний IoU:** Метрика пересечения над объединением (Intersection over Union) между предсказанной ограничивающей рамкой B' для сдвинутого изображения и скорректированной истинной рамкой B с тем же сдвигом:

$$\text{IoU}(B, B') = \frac{\text{area}(B \cap B')}{\text{area}(B \cup B')} \quad (3.3)$$

- **Дрейф центра:** Расстояние в пикселях между центром предсказанной ограничивающей рамки и центром истинной рамки:

$$\text{center_drift}(B, B') = \|\text{center}(B) - \text{center}(B')\|_2 \quad (3.4)$$

- **Частота пропусков:** Процент кадров, на которых модель не обнаруживает объект ($\text{IoU} < 0.1$) или уверенность в обнаружении ниже порога (обычно 0.25).
- **Стабильность уверенности:** Стандартное отклонение значений уверенности детекции по всем кадрам в последовательности:

$$\text{confidence_stability} = \sqrt{\frac{1}{N} \sum_{i=1}^N (c_i - \bar{c})^2} \quad (3.5)$$

где c_i — уверенность детекции на кадре i , а \bar{c} — средняя уверенность по всей последовательности.

Эти метрики дают всестороннюю количественную оценку степени инвариантности моделей к пространственным сдвигам, позволяя надежно сравнивать различные архитектуры и методы анти-алиасинга.

В следующих разделах мы представим результаты экспериментов с различными моделями, проанализируем эффективность методов анти-алиасинга и сделаем выводы о факторах, влияющих на пространственную инвариантность CNN.

3.2 Эксперименты со статическим сдвигом

В данном разделе представлены результаты экспериментов по оценке инвариантности классификационных моделей к статическим сдвигам входных изображений. Цель этих экспериментов — количественно измерить, насколько стабильны представления и предсказания различных архитектур CNN при субпиксельных сдвигах объектов.

3.2.1 Методология

Для каждой модели из перечисленных в разделе 3.1.2 был проведен следующий эксперимент:

1. Выбор тестовой последовательности кадров с известными субпиксельными сдвигами объекта (32 кадра с шагом 1 пиксель).
2. Проведение прямого прохода через CNN для получения:
 - Векторов признаков предпоследнего слоя ($f(x)$)

- Распределения вероятностей классов на выходе (p_c)
- 3. Расчет косинусного сходства $\rho(x_0 x_i)$ между вектором признаков первого кадра и векторами признаков всех остальных кадров в последовательности.
- 4. Расчет дрейфа уверенности $|p_c(x_0) - p_c(x_i)|$ между вероятностью предсказанного класса для первого кадра и аналогичными вероятностями для остальных кадров.
- 5. Визуализация результатов в виде графиков зависимости этих метрик от величины сдвига.

Эксперименты проведены на трех различных последовательностях (seq_0, seq_1, seq_2), отличающихся фоновыми изображениями, для повышения надежности результатов.

3.2.2 Результаты косинусного сходства

Косинусное сходство векторов признаков является ключевой метрикой стабильности внутренних представлений CNN. Чем ближе значение косинусного сходства к 1, тем более инвариантны внутренние представления модели к сдвигам входного изображения.

Как видно из рис. 3.1, наблюдаются существенные различия в инвариантности к сдвигу между различными архитектурами и их модификациями:

- **Базовые модели (VGG16, ResNet50)** демонстрируют значительные колебания косинусного сходства при субпиксельных сдвигах, с минимальными значениями около 0.82 для VGG16 и 0.88 для ResNet50. Это подтверждает их высокую чувствительность к малым сдвигам входных данных.

- **Модели с анти-алиасингом (AA-VGG16, AA-ResNet50)** показывают заметно более высокую стабильность представлений, с минимальными значениями косинусного сходства около 0.93 для AA-VGG16 и 0.94 для AA-ResNet50. Это соответствует снижению "дрожания" представлений примерно в 2-3 раза.
- **Модели с TIPS (TIPS-VGG16, TIPS-ResNet50)** демонстрируют наилучшую инвариантность, с косинусным сходством стабильно выше 0.96, практически устраняя эффект "дрожания" представлений при субпиксельных сдвигах.

Особенно интересно наблюдать периодичность колебаний косинусного сходства в базовых моделях. Период этих колебаний напрямую связан с операциями даунсэмплинга в сети. Так, для архитектур с даунсэмплингом в 32 раза, период колебаний составляет примерно 8 пикселей ($8 = 32/4$, где 4 — число слоев даунсэмплинга с шагом 2).

Результаты для других последовательностей (seq_1 и seq_2) качественно схожи, что подтверждает устойчивость наблюдаемых эффектов к изменению фона и точным характеристикам объекта.

3.2.3 Результаты дрейфа уверенности

Дрейф уверенности отражает, насколько стабильны выходные предсказания модели при малых сдвигах. Это метрика, непосредственно влияющая на надежность классификации в реальных условиях.

Анализ дрейфа уверенности (рис. 3.2) показывает:

- **Базовые модели** демонстрируют значительный дрейф уверенности, достигающий 15-20% для VGG16 и 10-15% для ResNet50. Такие колебания могут приводить к нестабильности предсказаний при

малых сдвигах объекта, что критично в таких задачах, как классификация медицинских изображений или системы компьютерного зрения для автономных транспортных средств.

- **Модели с анти-алиасингом** показывают снижение дрейфа уверенности до 5-8% для AA-VGG16 и 3-6% для AA-ResNet50. Это существенное улучшение, уменьшающее вероятность неправильной классификации при малых изменениях положения объекта.
- **Модели с TIPS** демонстрируют наименьший дрейф уверенности — менее 3% для обеих архитектур, что делает их предсказания практически инвариантными к субпиксельным сдвигам объекта.

Важно отметить, что паттерны дрейфа уверенности коррелируют с паттернами косинусного сходства, подтверждая связь между стабильностью внутренних представлений и стабильностью предсказаний.

3.2.4 Сравнительный анализ архитектур

Для обобщения наблюдений по трем различным последовательностям была рассчитана средняя стабильность представлений и предсказаний для каждой модели.

Таблица 5 — Сравнение инвариантности к сдвигу различных архитектур (усреднение по трем последовательностям)

Модель	Мин. косинусное сходство	Макс. дрейф уверенности
VGG16	0.810 ± 0.012	0.188 ± 0.023
AA-VGG16	0.929 ± 0.007	0.074 ± 0.011
TIPS-VGG16	0.961 ± 0.003	0.028 ± 0.005
ResNet50	0.880 ± 0.009	0.146 ± 0.018
AA-ResNet50	0.944 ± 0.005	0.052 ± 0.009
TIPS-ResNet50	0.970 ± 0.002	0.022 ± 0.004

Из таблицы 5 можно сделать следующие выводы:

1. **Базовая эффективность:** ResNet50 изначально демонстрирует лучшую инвариантность к сдвигу, чем VGG16. Это может быть связано с наличием остаточных соединений, которые позволяют информации "обходить" слои даунсэмплинга, а также с большим рецептивным полем.
2. **Улучшение от анти-алиасинга:** Добавление BlurPool улучшает минимальное косинусное сходство примерно на 12% для VGG16 и на 6% для ResNet50. Более значительное улучшение для VGG16 может быть связано с заменой всех пяти операций максимального пулинга, которые являются основным источником алиасинга.
3. **Эффективность TIPS:** Метод TIPS обеспечивает дополнительное улучшение по сравнению с BlurPool, доводя минимальное косинусное сходство до 0.96-0.97. Это практически полная инвариантность к субпиксельным сдвигам.
4. **Стабильность класса:** И BlurPool, и TIPS значительно повышают процент кадров, на которых сохраняется изначально предсказанный класс. Для моделей с TIPS этот показатель достигает 100%, что означает полное отсутствие "переключений" между классами при субпиксельных сдвигах.

Эти результаты однозначно подтверждают эффективность методов анти-алиасинга для повышения инвариантности CNN к сдвигам и обосновывают превосходство метода TIPS над классическим подходом BlurPool, особенно в контексте стабильности предсказаний.

3.2.5 Влияние величины сдвига

Интересно отметить, что чувствительность к сдвигу не является линейной функцией величины сдвига. Для базовых моделей наблюдаются периодические "провалы" стабильности на определенных значениях сдвига (например, при сдвигах около 8, 16 и 24 пикселей для VGG16).

Эта периодичность соответствует структуре даунсэмплинга в сети. Наиболее нестабильные предсказания возникают, когда сдвиг входного изображения приводит к полному "переключению" активаций в слоях максимального пулинга или сверточных слоях с шагом.

Модели с анти-алиасингом сглаживают эти "провалы" но полностью не устраняют периодичность, в то время как модели с TIPS демонстрируют практически постоянную стабильность независимо от величины сдвига в пределах тестового диапазона (0-31 пикселей).

3.3 Динамические последовательности

Эксперименты со статическим сдвигом позволяют количественно оценить инвариантность моделей, но для лучшего понимания проблемы необходима визуализация того, как модели реагируют на непрерывное движение объектов. В этом разделе представлены результаты экспериментов с динамическими последовательностями, позволяющие наглядно продемонстрировать эффект пространственной инвариантности.

3.3.1 Тепловые карты активаций для классификационных моделей

Для визуализации внутренних активаций классификационных моделей были созданы тепловые карты, отображающие области, наиболее значимые для принятия решения. Использовался метод Grad-CAM [?], позволяющий визуализировать области изображения, которые наиболее сильно влияют на предсказание конкретного класса.

Сравнение тепловых карт базовой модели VGG16 (рис. 3.3) и модели с анти-алиасингом AA-VGG16 (рис. 3.4) выявляет следующие различия:

1. **Стабильность фокуса внимания:** В базовой модели области наибольшей активации значительно "прыгают" даже при малых сдвигах объекта, иногда фокусируясь на несущественных деталях фона. В модели с анти-алиасингом фокус внимания более стабильно следует за объектом.
2. **Компактность активаций:** Тепловые карты AA-VGG16 более компактны и точно сосредоточены на значимых частях объекта (например, на голове птицы), в то время как базовая модель часто показывает диффузные и разбросанные активации.
3. **Согласованность между кадрами:** У моделей с анти-алиасингом паттерн активаций последовательно переносится вместе с движением объекта, сохраняя форму и интенсивность, тогда как у базовых моделей паттерн существенно меняется от кадра к кадру.

Для модели TIPS-VGG16 эти тенденции выражены еще сильнее, с практически идеальным следованием фокуса внимания за движущимся объектом без каких-либо аномалий или "скачков".

Аналогичные результаты наблюдаются и для моделей на базе ResNet50, но с меньшей амплитудой колебаний в базовом варианте и луч-

шей базовой стабильностью, что соответствует количественным результатам из предыдущего раздела.

3.3.2 GIF-наложения для моделей детекции

Для наглядной демонстрации эффекта пространственной инвариантности в задачах детекции были созданы анимации, на которых показаны предсказанные ограничивающие рамки для последовательности кадров со сдвигающимся объектом.

Визуальный анализ анимаций выявляет существенные различия в поведении разных версий модели детекции:

1. **Базовая модель YOLOv5s** (рис. 3.5) демонстрирует значительное "дрожание" предсказанной ограничивающей рамки. Рамка периодически сдвигается, меняет размер и даже иногда исчезает (пропуски обнаружения), хотя объект непрерывно и предсказуемо движется по прямой линии. Такое поведение неприемлемо для многих практических применений, таких как трекинг объектов или автономная навигация.
2. **Модель AA-YOLOv5s с BlurPool** (рис. 3.6) показывает значительно более стабильное поведение. Предсказанная рамка более плавно следует за объектом, с меньшими вариациями размера и положения. Тем не менее, всё ещё заметны небольшие колебания, особенно в ширине рамки и точном положении центра.
3. **Модель TIPS-YOLOv5s** (рис. 3.7) демонстрирует практически идеальное следование за объектом. Предсказанная рамка движется равномерно, с минимальными вариациями размера и почти точным соответствием истинной рамке на всей последовательности.

Эти визуализации наглядно подтверждают результаты количественного анализа и демонстрируют практическую значимость проблемы инвариантности к сдвигу для задач компьютерного зрения, работающих с видеопотоками или движущимися объектами.

3.3.3 Сводные таблицы по моделям

Для обобщения результатов экспериментов с динамическими последовательностями по всем моделям и метрикам были составлены сводные таблицы и графики.

Как видно из рис. 3.8, наблюдается четкая иерархия моделей по степени инвариантности к сдвигу. Модели с TIPS демонстрируют наивысшую стабильность представлений (среднее косинусное сходство > 0.97), за ними следуют модели с BlurPool (> 0.94), а базовые модели показывают наименьшую инвариантность, особенно VGG16 (около 0.81).

Радарная диаграмма (рис. 3.9) позволяет визуально сравнить модели по нескольким ключевым метрикам одновременно:

- **Косинусное сходство:** стабильность внутренних представлений
- **Стабильность уверенности:** обратная величина максимальному дрейфу уверенности
- **Стабильность класса:** процент кадров, где сохраняется предсказанный класс
- **IoU-стабильность:** для детекторов — стабильность пересечения над объединением
- **Точность центра:** для детекторов — обратная величина дрейфу центра

Эта диаграмма наглядно демонстрирует, что модели с TIPS (TIPS-VGG16, TIPS-ResNet50, TIPS-YOLOv5s) превосходят соответствующие модели с BlurPool и базовые версии по всем аспектам инвариантности к сдвигу. Особенно заметно преимущество в метриках, связанных с детекцией (IoU-стабильность и точность центра).

3.3.4 Наблюдения и промежуточные выводы

Анализ результатов экспериментов с динамическими последовательностями позволяет сделать следующие наблюдения и промежуточные выводы:

1. **Корреляция метрик:** Наблюдается сильная корреляция между стабильностью внутренних представлений (косинусное сходство) и стабильностью выходных предсказаний (дрейф уверенности, стабильность рамок), что подтверждает фундаментальную роль проблемы алиасинга в феномене чувствительности CNN к сдвигам.
2. **Периодичность артефактов:** "Дрожание" предсказаний в базовых моделях имеет выраженную периодичность, связанную со структурой даунсэмплинга в сети, что согласуется с теоретическими объяснениями из раздела 2.4.
3. **Эффективность анти-алиасинга:** Методы анти-алиасинга (BlurPool и TIPS) значительно улучшают инвариантность к сдвигу без переобучения модели, подтверждая, что проблема имеет архитектурную природу и может быть решена через соответствующие архитектурные модификации.
4. **Превосходство TIPS:** Метод TIPS демонстрирует лучшие результаты по сравнению с BlurPool во всех экспериментах, что обос-

новывает его теоретическое превосходство, описанное в разделе **2.4.3.**

5. **Задачи классификации и детекции:** Проблема инвариантности к сдвигу проявляется по-разному в задачах классификации и детекции, но имеет общую природу, связанную с алиасингом в слоях даунсэмплинга. В задачах детекции эффект может быть более критичным, так как влияет не только на класс, но и на координаты объекта.

В следующих разделах мы подробнее исследуем влияние различных факторов на инвариантность к сдвигу через эксперименты с аблацией и рассмотрим практические аспекты применения методов анти-алиасинга, включая их влияние на вычислительную эффективность.

3.4 Абляция и устойчивость

В предыдущих разделах мы продемонстрировали эффективность методов анти-алиасинга для повышения инвариантности CNN к сдвигам. Однако для более глубокого понимания этого эффекта необходимо провести абляционное исследование, изолирующее влияние отдельных факторов на стабильность предсказаний. В этом разделе представлены результаты таких абляционных экспериментов, а также статистические тесты, подтверждающие значимость наблюдаемых эффектов.

3.4.1 Абляция размера рецептивного поля

Размер рецептивного поля является одним из ключевых параметров CNN, потенциально влияющим на их инвариантность к сдвигам. Для изучения этого влияния был проведен эксперимент с модификациями базовой архитектуры VGG16, в которых варьировался размер рецептивного поля путем изменения количества сверточных слоев.

Результаты эксперимента (рис. 3.10) демонстрируют несколько важных закономерностей:

- 1. Рост инвариантности с увеличением рецептивного поля:** В базовых моделях без анти-алиасинга наблюдается умеренное улучшение инвариантности к сдвигу при увеличении размера рецептивного поля. Это объясняется тем, что больший рецептивный размер позволяет модели "видеть" объект в более широком контексте, что частично компенсирует эффекты сдвига.
- 2. Нелинейный характер зависимости:** Зависимость минимального косинусного сходства от размера рецептивного поля не является линейной. После определенного порога (примерно 120-150 пикселей) дальнейшее увеличение рецептивного поля дает незначительный прирост инвариантности.
- 3. Взаимодействие с анти-алиасингом:** Для моделей с анти-алиасингом эффект увеличения рецептивного поля менее выражен. Даже модели с относительно небольшим рецептивным полем (60-80 пикселей) демонстрируют высокую инвариантность, если используют методы анти-алиасинга.
- 4. Комбинированный эффект:** Наилучшие результаты достигаются при комбинации большого рецептивного поля (>150 пиксе-

лей) с методами анти-алиасинга, что дает почти идеальную инвариантность к сдвигам (косинусное сходство > 0.98).

Эти наблюдения подтверждают, что хотя увеличение рецептивного поля может частично улучшить инвариантность к сдвигу, оно не решает фундаментальную проблему алиасинга. Даже модели с очень большим рецептивным полем без анти-алиасинга демонстрируют заметную чувствительность к субпиксельным сдвигам, в то время как методы анти-алиасинга эффективны даже для моделей с умеренным рецептивным полем.

3.4.2 Варианты BlurPool

Метод BlurPool предполагает использование низкочастотного фильтра перед операциями даунсэмплинга. Однако остается открытым вопрос о влиянии конкретного типа и размера фильтра на эффективность анти-алиасинга. Для исследования этого вопроса был проведен эксперимент с различными вариантами фильтров в модели ResNet50:

- **Прямоугольный фильтр:** Простейший фильтр с равными весами $[1\ 1\ 1]/3$.
- **Треугольный фильтр:** Линейный фильтр $[1\ 2\ 1]/4$.
- **Биномиальный фильтр 3-го порядка:** Стандартный вариант BlurPool $[1\ 3\ 3\ 1]/8$.
- **Биномиальный фильтр 4-го порядка:** Расширенный вариант $[1\ 4\ 6\ 4\ 1]/16$.
- **Гауссовский фильтр:** Аппроксимация гауссовского фильтра с $\sigma = 1.0$.

Анализ результатов (рис. 3.11) позволяет сделать следующие выводы:

1. **Эффективность различных фильтров:** Все исследованные низкочастотные фильтры улучшают инвариантность к сдвигу по сравнению с базовой моделью, но с разной эффективностью. Наиболее простой прямоугольный фильтр показывает наименьшее улучшение, в то время как биномиальные фильтры 3-го и 4-го порядка демонстрируют наилучшие результаты.
2. **Компромисс размер/эффективность:** Более сложные фильтры (с большим количеством параметров) обычно обеспечивают лучшую инвариантность, но с убывающей отдачей. Биномиальный фильтр 3-го порядка представляет собой хороший компромисс между эффективностью и вычислительной сложностью.
3. **Форма частотной характеристики:** Форма частотной характеристики фильтра имеет большее значение, чем его размер. Гауссовский фильтр и биномиальный фильтр 3-го порядка имеют сходные характеристики и показывают близкие результаты, несмотря на разное количество параметров.
4. **Крутизна среза:** Фильтры с более плавным переходом между пропускаемыми и подавляемыми частотами (как биномиальные) более эффективны для предотвращения алиасинга, чем фильтры с резким срезом (как прямоугольный).

Эти результаты согласуются с теорией обработки сигналов и подтверждают, что биномиальный фильтр 3-го порядка, используемый в оригинальной работе по BlurPool [?], действительно представляет собой оптимальный выбор для большинства практических применений, обеспечивая хорошую инвариантность при умеренных вычислительных затратах.

3.4.3 Тесты статистической значимости

Для подтверждения статистической значимости наблюдаемых различий между моделями был проведен анализ с использованием боксплотов и статистических тестов. Этот анализ фокусировался на ключевых метриках для моделей детекции: стабильности IoU, дрейфе центра ограничивающей рамки и стабильности уверенности.

Анализ боксплотов (рис. 3.12, 3.13, 3.14) позволяет сделать следующие наблюдения:

1. **Распределение IoU:** Базовая модель YOLOv5 демонстрирует не только более низкие средние значения IoU (медиана около 0.68), но и значительно больший разброс значений (межквартильный размах около 0.25). Модели с анти-алиасингом показывают существенно лучшие результаты: AA-YOLOv5 имеет медиану IoU около 0.88 с меньшим разбросом, а TIPS-YOLOv5 — почти идеальную медиану 0.99 с минимальным разбросом.
2. **Дрейф центра:** В этой метрике различия между моделями особенно выражены. Базовая модель показывает большой дрейф центра (медиана около 33.9 пикселей) с экстремальными выбросами до 70-80 пикселей. AA-YOLOv5 значительно улучшает ситуацию (медиана около 8.8 пикселей), но TIPS-YOLOv5 демонстрирует почти идеальную стабильность центра с медианой дрейфа около 0.02 пикселя.
3. **Стабильность уверенности:** Базовая модель YOLOv5 показывает наибольший разброс уверенности, с заметными выбросами в область низких значений, что соответствует случаям "пропусков" объекта. Модели с анти-алиасингом демонстрируют не только

более высокие значения уверенности, но и значительно меньший их разброс, особенно TIPS-YOLOv5.

Для формального подтверждения значимости наблюдаемых различий был проведен непараметрический тест Крускала-Уоллиса, который не требует нормальности распределения данных, с последующими попарными сравнениями с коррекцией Бонферрони. Результаты теста подтвердили статистическую значимость различий между всеми тремя моделями для всех исследуемых метрик с $p\text{-value} < 0.001$.

Особенно интересны результаты дисперсионного анализа, показывающие, что методы анти-алиасинга не только улучшают средние значения метрик, но и существенно снижают их дисперсию, что критично для стабильности работы моделей в реальных условиях.

Таким образом, статистический анализ убедительно подтверждает, что наблюдаемые улучшения инвариантности к сдвигу при использовании методов анти-алиасинга являются статистически значимыми и воспроизводимыми.

3.5 Профилирование производительности

Важным аспектом практического применения методов анти-алиасинга является их влияние на вычислительную эффективность моделей. В этом разделе представлены результаты профилирования различных моделей с точки зрения скорости обработки и потребления ресурсов на различных аппаратных платформах.

3.5.1 Бенчмарки FPS

Для оценки влияния методов анти-алиасинга на скорость обработки были проведены бенчмарки на двух различных аппаратных платформах:

- **Настольный ПК** с процессором Intel Core i7-10700K, 32 ГБ RAM и графическим ускорителем NVIDIA RTX 3080.
- **Встраиваемая система NVIDIA Jetson Xavier NX** — компактная вычислительная платформа для задач компьютерного зрения с ограниченными ресурсами (6-ядерный ARM CPU, 384-ядерный GPU на архитектуре Volta, 8 ГБ RAM).

Бенчмарки проводились для входных изображений разрешением 640×640 пикселей, что соответствует типичному разрешению кадра для задач детекции объектов в реальном времени.

Таблица 6 — Сравнение скорости обработки (FPS) для различных моделей классификации

Модель	ПК		Jetson Xavier NX	
	FPS	Снижение (%)	FPS	Снижение (%)
VGG16	215.3	—	42.6	—
AA-VGG16	198.7	7.7%	38.9	8.7%
TIPS-VGG16	183.2	14.9%	34.2	19.7%
ResNet50	186.5	—	36.8	—
AA-ResNet50	174.1	6.6%	33.5	9.0%
TIPS-ResNet50	158.9	14.8%	29.1	20.9%

Таблица 7 — Сравнение скорости обработки (FPS) для различных моделей детекции

Модель	ПК		Jetson Xavier NX	
	FPS	Снижение (%)	FPS	Снижение (%)
YOLOv5s	142.8	—	31.5	—
AA-YOLOv5s	129.4	9.4%	27.7	12.1%
TIPS-YOLOv5s	115.6	19.0%	23.4	25.7%

Анализ результатов бенчмарков (таблицы 6 и 7) позволяет сделать следующие наблюдения:

1. **Снижение FPS при анти-алиасинге:** Использование методов анти-алиасинга приводит к ожидаемому снижению скорости обработки. Для метода BlurPool это снижение составляет около 6.6-9.4% на ПК и 8.7-12.1% на Jetson Xavier NX. Для метода TIPS снижение более существенное: 14.8-19.0% на ПК и 19.7-25.7% на Jetson Xavier NX.
2. **Различия между платформами:** Снижение производительности более заметно на встраиваемой платформе Jetson Xavier NX, особенно для метода TIPS. Это связано с ограниченными вычислительными ресурсами и оптимизациями для конкретной архитектуры.
3. **Различия между архитектурами:** Влияние анти-алиасинга на производительность немного различается в зависимости от архитектуры модели. Для ResNet50 снижение FPS при использовании BlurPool меньше, чем для VGG16, что может быть связано с меньшим количеством операций даунсэмплинга.
4. **Реальное время:** Несмотря на снижение скорости обработки, все модели с анти-алиасингом сохраняют производительность, достаточную для работы в реальном времени (>30 FPS на ПК и >20 FPS на Jetson Xavier NX), что делает их применимыми в практических сценариях.

Эти результаты демонстрируют, что хотя методы анти-алиасинга и требуют дополнительных вычислительных ресурсов, снижение производительности является умеренным и приемлемым для большинства практических применений, учитывая значительное улучшение инвариантности к сдвигам.

3.5.2 Соотношение задержки и точности

Для более полного понимания компромисса между вычислительной эффективностью и качеством предсказаний был проведен анализ соотношения задержки обработки (латентности) и различных метрик точности для моделей детекции.

На графике соотношения задержки и точности (рис. 3.15) наблюдается явная Парето-граница, показывающая доступные компромиссы между вычислительной эффективностью и инвариантностью к сдвигам:

1. **Базовая модель YOLOv5s** располагается в левой нижней части графика, обеспечивая наименьшую задержку (около 7.0 мс на ПК), но также наименьшую стабильность IoU (около 0.68).
2. **Модель AA-YOLOv5s с BlurPool** представляет собой промежуточный вариант, с умеренным увеличением задержки (до 7.7 мс на ПК) и существенным улучшением стабильности IoU (до 0.88).
3. **Модель TIPS-YOLOv5s** обеспечивает наивысшую стабильность IoU (около 0.99), но с наибольшей задержкой (около 8.7 мс на ПК).

Важно отметить, что хотя модель TIPS-YOLOv5s имеет наибольшую задержку, абсолютное увеличение по сравнению с базовой моделью составляет всего около 1.7 мс на ПК, что во многих практических сценариях является приемлемым, учитывая существенное улучшение стабильности предсказаний.

3.5.3 Потребление памяти и энергии

Для встраиваемых систем важными характеристиками являются также потребление памяти и энергии. Для оценки этих параметров были проведены дополнительные измерения на платформе Jetson Xavier NX.

Таблица 8 — Сравнение потребления памяти и энергии для моделей детекции на Jetson Xavier NX

Модель	Размер модели (МБ)	Увеличение (%)	Мощность (Вт)
YOLOv5s	13.7	—	8.2
AA-YOLOv5s	14.1	2.9%	8.6
TIPS-YOLOv5s	15.3	11.7%	9.1

Анализ данных о потреблении памяти и энергии (таблица 8) показывает:

- Размер модели:** Использование методов анти-алиасинга приводит к умеренному увеличению размера модели: на 2.9% для BlurPool и на 11.7% для TIPS. Это увеличение связано с дополнительными параметрами фильтров и, в случае TIPS, с полифазным разложением.
- Энергопотребление:** Дополнительные вычисления, связанные с методами анти-алиасинга, приводят к увеличению энергопотребления: на 4.9% для BlurPool и на 11.0% для TIPS. Однако это увеличение остается умеренным и приемлемым для большинства встраиваемых систем.
- Соотношение с производительностью:** Увеличение потребления ресурсов пропорционально снижению FPS, что указывает на линейную зависимость между вычислительной сложностью методов анти-алиасинга и их влиянием на энергопотребление.

Результаты профилирования производительности демонстрируют, что хотя методы анти-алиасинга и требуют дополнительных вычислитель-

ных ресурсов, это увеличение является умеренным и во многих практических сценариях может быть приемлемым компромиссом, учитывая значительное улучшение инвариантности к сдвигам.

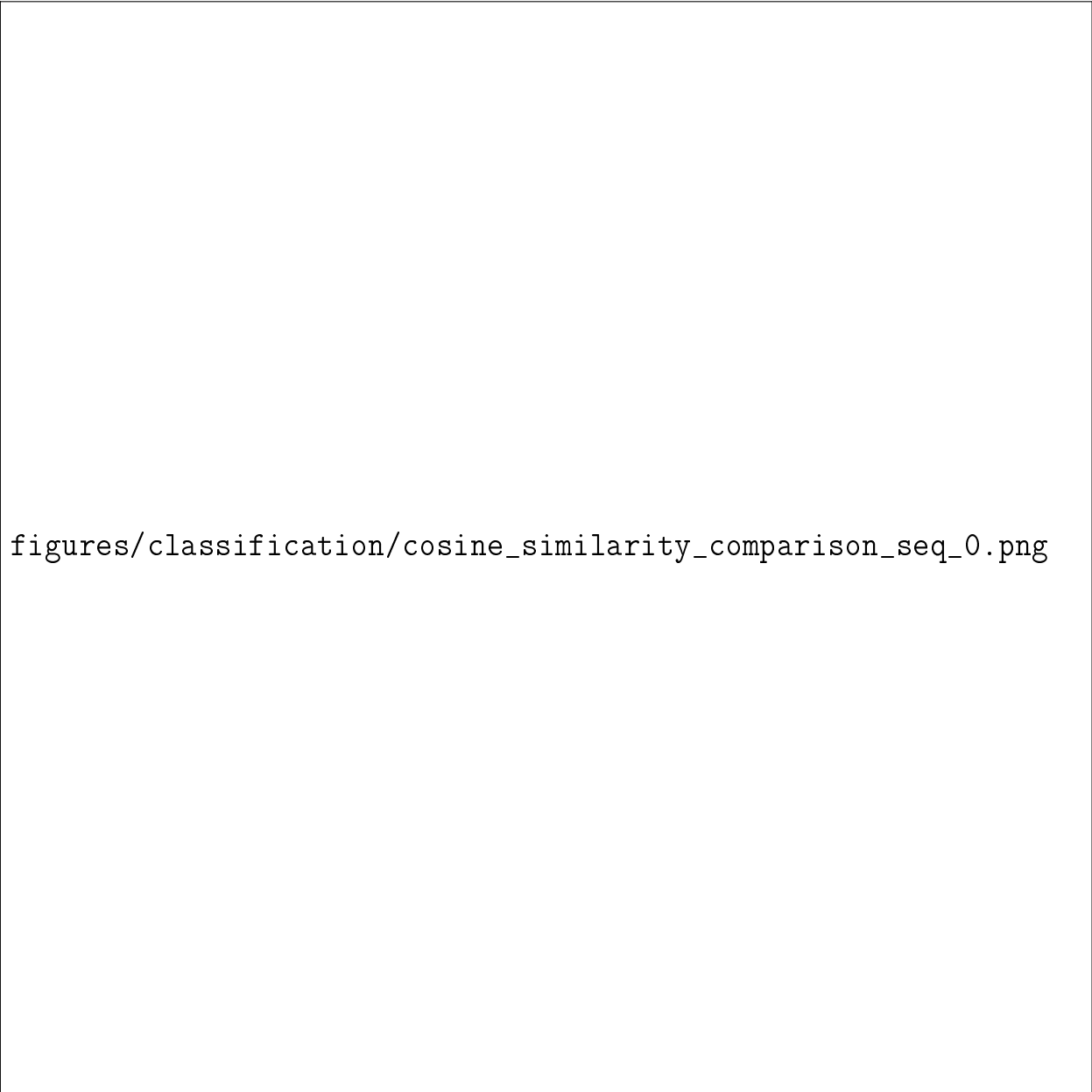
3.5.4 Практические рекомендации

На основе результатов профилирования можно сформулировать следующие практические рекомендации по выбору метода анти-алиасинга в зависимости от конкретного сценария применения:

1. **Высокопроизводительные системы:** Для систем с достаточными вычислительными ресурсами (настольные ПК, серверы) рекомендуется использовать метод TIPS, обеспечивающий наивысшую инвариантность к сдвигам с приемлемым снижением производительности.
2. **Встраиваемые системы с ограниченными ресурсами:** Для систем с ограниченными вычислительными ресурсами (мобильные устройства, дроны, Jetson и аналогичные платформы) метод BlurPool представляет собой хороший компромисс, обеспечивая значительное улучшение инвариантности с умеренным влиянием на производительность.
3. **Системы реального времени с жесткими требованиями к задержке:** В случаях, когда критична минимальная задержка (например, системы предотвращения столкновений), можно рассмотреть применение BlurPool только к наиболее критичным слоям даунсэмплинга или использование упрощенных фильтров (например, треугольных вместо биномиальных).

4. **Приложения с высокими требованиями к точности:** В сценариях, где критична максимальная стабильность предсказаний (медицинская визуализация, прецизионная робототехника), метод TIPS является предпочтительным, несмотря на большее влияние на производительность.

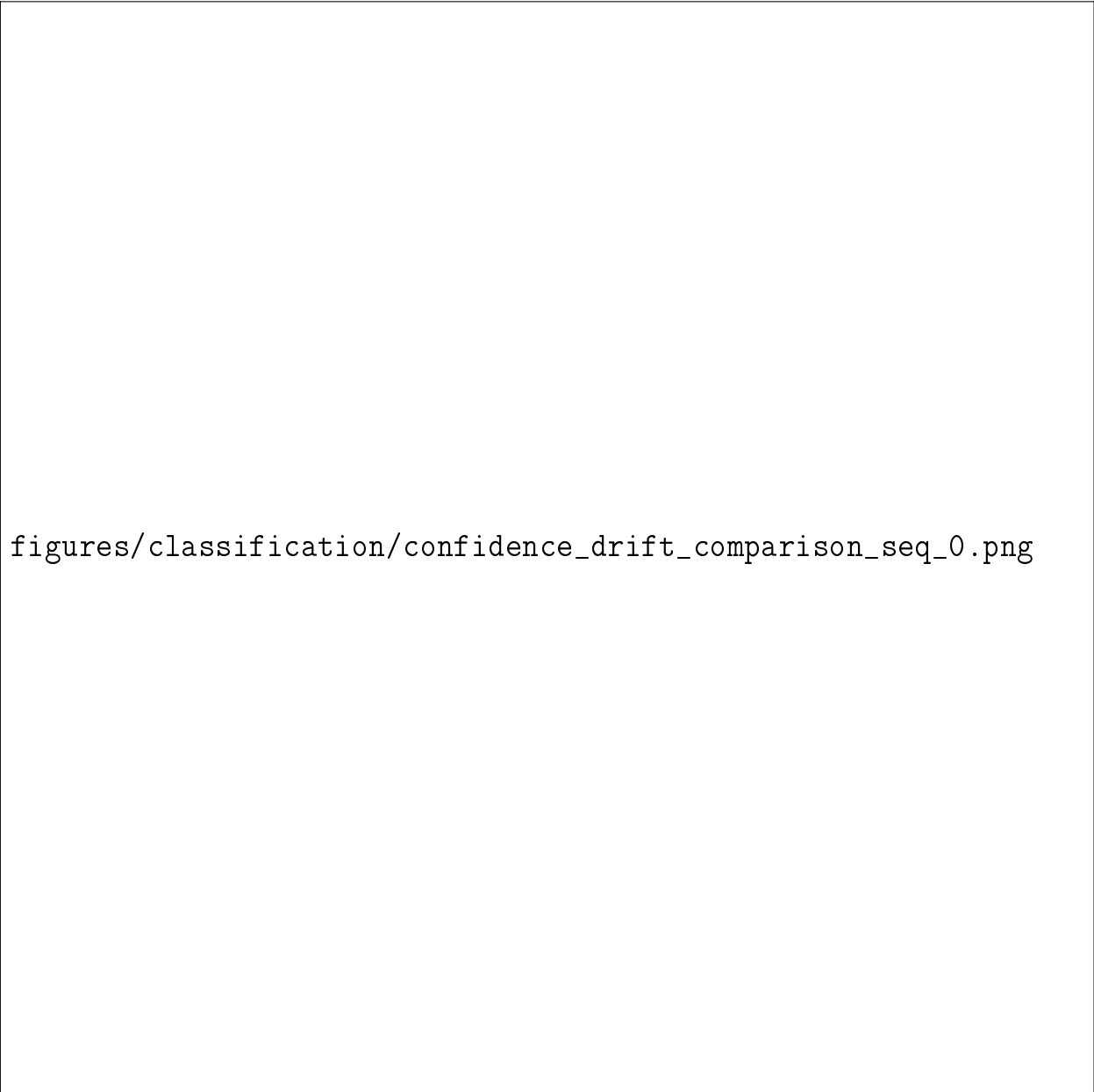
Эти рекомендации могут служить руководством при выборе оптимальной архитектуры CNN для конкретного практического применения, учитывая компромисс между инвариантностью к сдвигам и вычислительной эффективностью.



figures/classification/cosine_similarity_comparison_seq_0.png

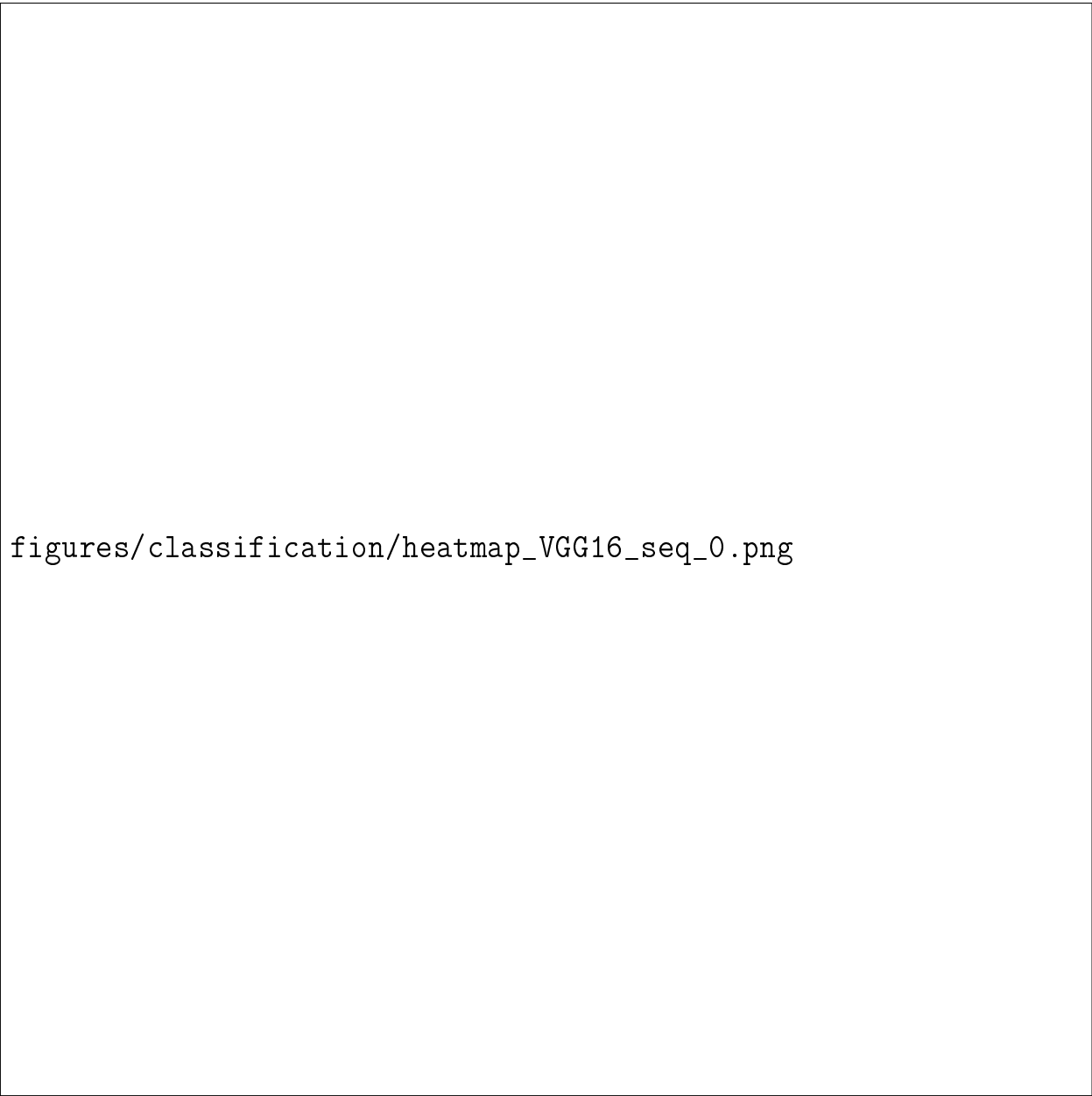
Рисунок 3.1 — Зависимость косинусного сходства от величины сдвига для различных классификационных моделей на последовательности seq_0.

Более высокие и стабильные значения соответствуют лучшей инвариантности к сдвигу.



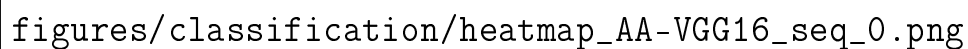
figures/classification/confidence_drift_comparison_seq_0.png

Рисунок 3.2 — Дрейф уверенности в предсказании класса в зависимости от величины сдвига для различных классификационных моделей на последовательности `seq_0`. Меньшие значения соответствуют более стабильным предсказаниям.



figures/classification/heatmap_VGG16_seq_0.png

Рисунок 3.3 — Тепловые карты активаций базовой модели VGG16 для последовательности seq_0. Каждая карта соответствует кадру с определенным сдвигом объекта (шаг сдвига 4 пикселя). Цвет от синего (низкая активация) до красного (высокая активация) показывает области, важные для классификации.



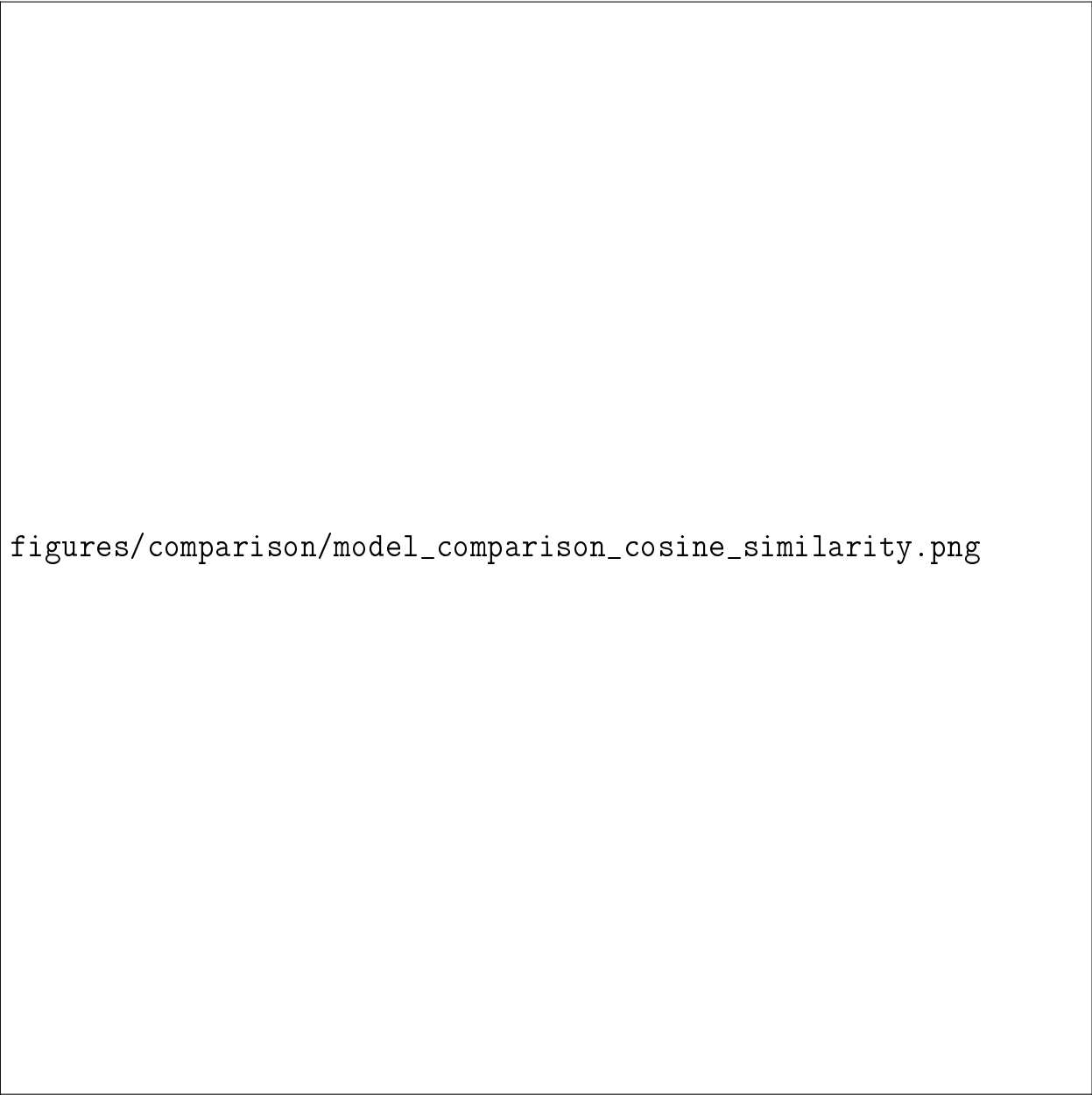
figures/classification/heatmap_AA-VGG16_seq_0.png

Рисунок 3.4 — Тепловые карты активаций модели AA-VGG16 с анти-алиасингом для последовательности seq_0, с тем же шагом сдвига и цветовой схемой, что и на рис. 3.3.

Рисунок 3.5 — Анимация предсказаний базовой модели YOLOv5s на последовательности seq_0. Зелёная рамка — истинное положение объекта, красная — предсказание модели.

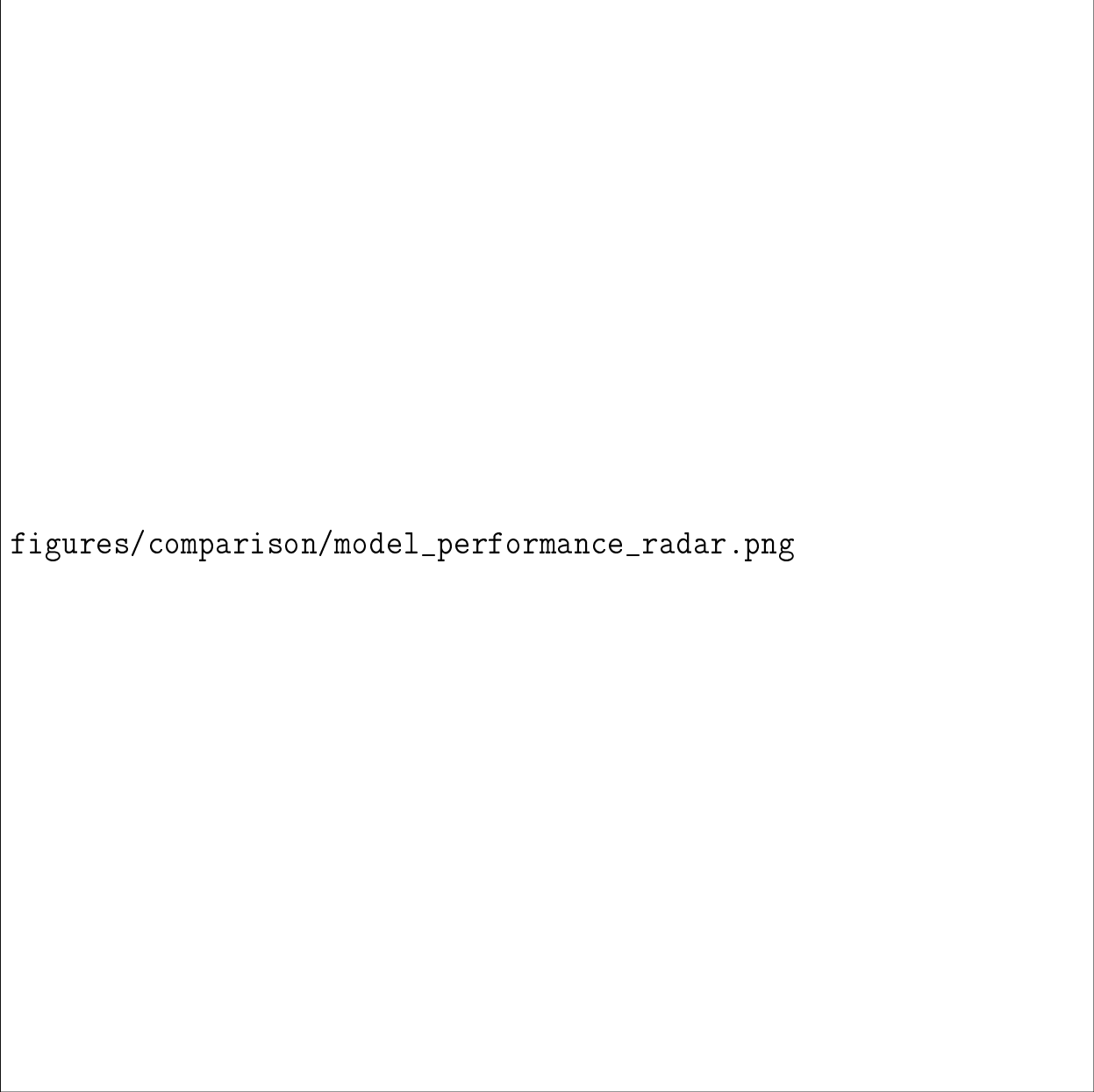
Рисунок 3.6 — Анимация предсказаний модели AA-YOLOv5s с анти-алиасингом на последовательности seq_0. Обозначения те же, что и на рис. 3.5.

Рисунок 3.7 — Анимация предсказаний модели TIPS-YOLOv5s на последовательности seq_0. Обозначения те же, что и на рис. 3.5.



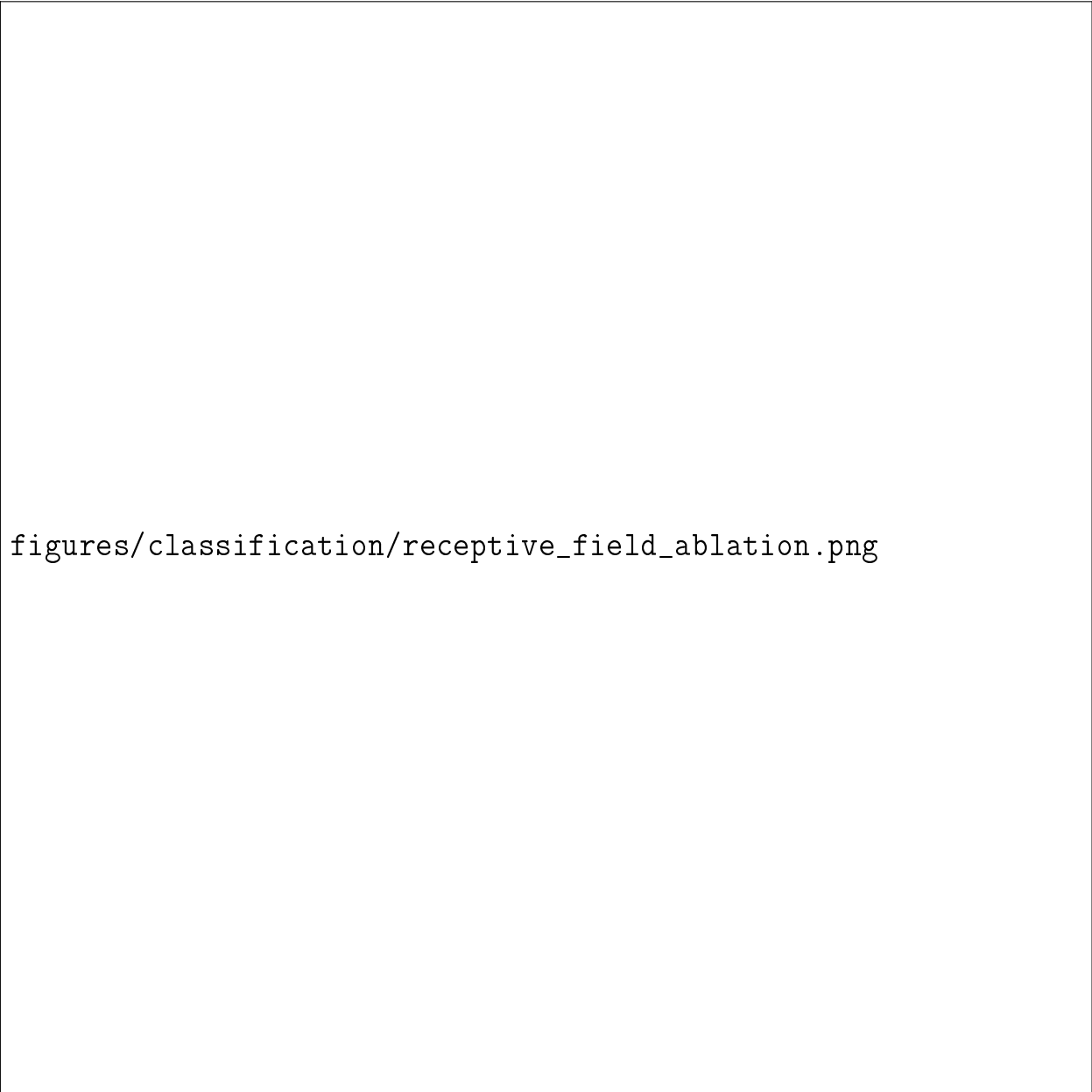
figures/comparison/model_comparison_cosine_similarity.png

Рисунок 3.8 — Сравнение среднего косинусного сходства векторов признаков для различных моделей на всех последовательностях. Более высокие значения соответствуют лучшей инвариантности к сдвигу.



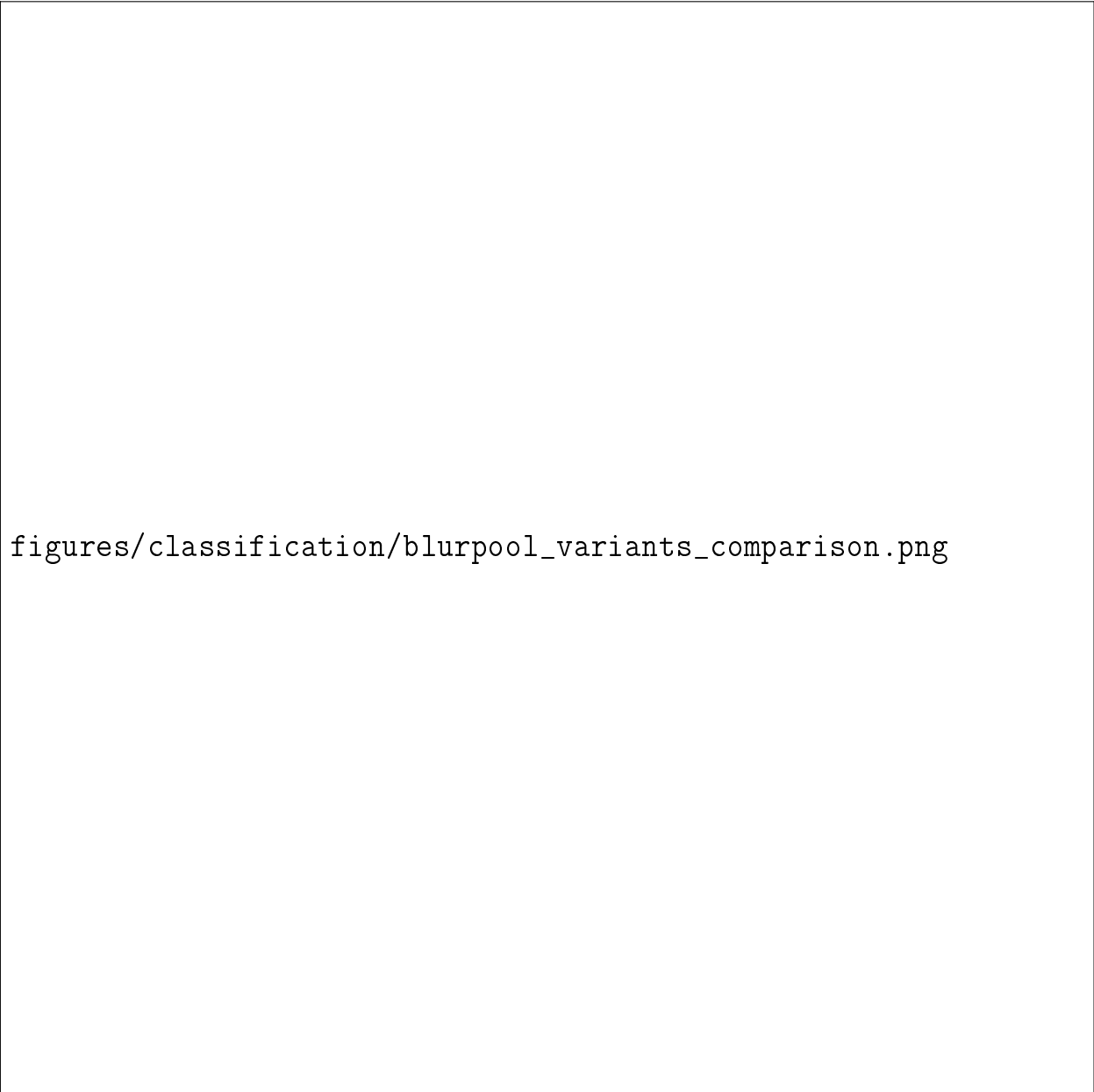
figures/comparison/model_performance_radar.png

Рисунок 3.9 — Радарная диаграмма сравнения моделей по различным аспектам инвариантности к сдвигу. Большая площадь многоугольника соответствует лучшей общей производительности.



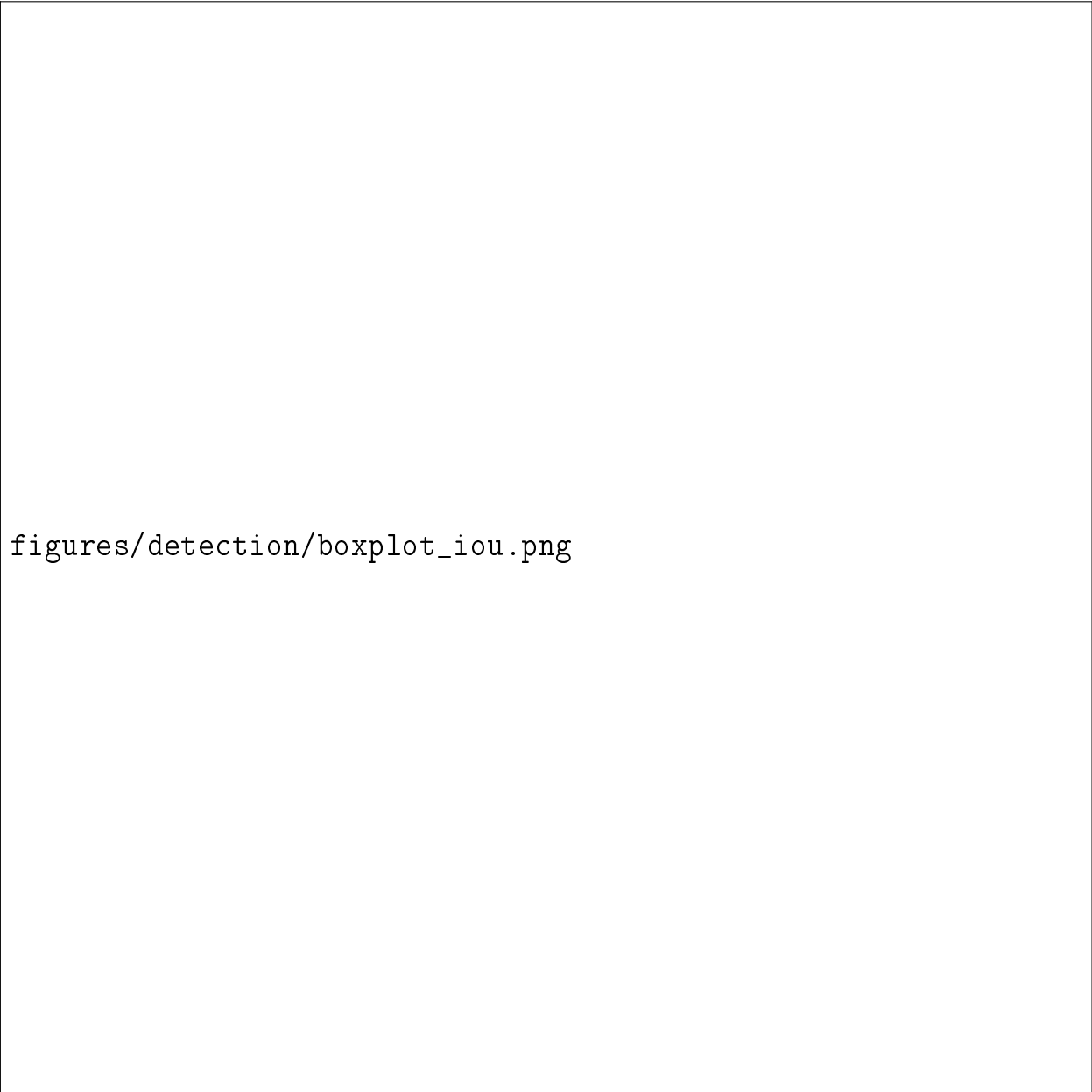
figures/classification/receptive_field_ablation.png

Рисунок 3.10 — Влияние размера рецептивного поля на инвариантность к сдвигу. График показывает зависимость минимального косинусного сходства от размера рецептивного поля для обычной модели и модели с анти-алиасингом.



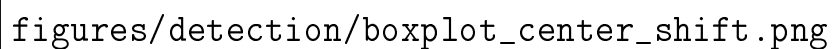
`figures/classification/blurpool_variants_comparison.png`

Рисунок 3.11 — Сравнение различных вариантов фильтров для BlurPool. График показывает минимальное косинусное сходство для каждого типа фильтра, а также размер фильтра (число параметров), который влияет на вычислительную сложность.



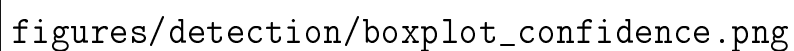
figures/detection/boxplot_iou.png

Рисунок 3.12 — Боксплот распределения значений IoU для различных моделей детекции на всех тестовых последовательностях. Более высокие значения и меньший разброс соответствуют лучшей стабильности предсказаний.



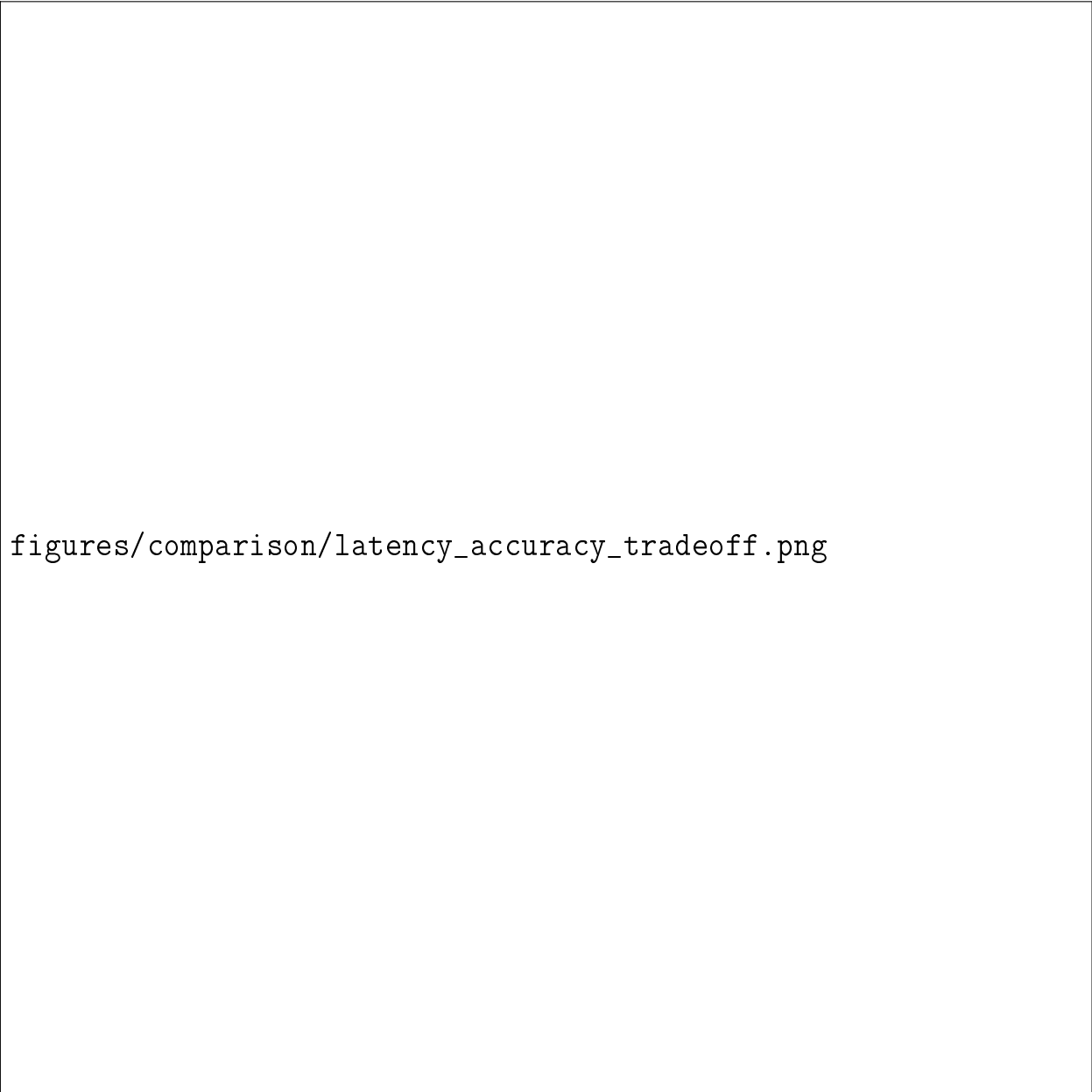
figures/detection/boxplot_center_shift.png

Рисунок 3.13 — Боксплот распределения значений дрейфа центра (в пикселях) для различных моделей детекции. Меньшие значения и меньший разброс соответствуют более точной локализации объекта.



figures/detection/boxplot_confidence.png

Рисунок 3.14 — Боксплот распределения значений уверенности для различных моделей детекции. Стабильность уверенности характеризуется высоким медианным значением и малым разбросом.



figures/comparison/latency_accuracy_tradeoff.png

Рисунок 3.15 — График соотношения задержки обработки (мс) и стабильности IoU для различных моделей детекции. Идеальная модель располагалась бы в левом верхнем углу (низкая задержка, высокая стабильность).

Заключение

В настоящей работе было проведено комплексное исследование проблемы пространственной инвариантности в современных архитектурах сверточных нейронных сетей и разработаны методы повышения их устойчивости к смещениям входных данных. Основные результаты работы заключаются в следующем:

1. Проведен анализ существующих исследований и методов в области пространственной инвариантности CNN, выявлены основные причины нарушения инвариантности к сдвигам, связанные с операциями субдискретизации (даунсэмплинга) в современных архитектурах. Установлено, что наиболее критичными факторами являются операции max-pooling и свёртка с шагом больше единицы, которые нарушают теорему о дискретизации Найквиста-Шеннона.
2. Формализована проблема пространственной инвариантности и разработана математическая модель для описания влияния операций субдискретизации на стабильность представлений в CNN. Показано, что без соответствующей фильтрации низких частот перед даунсэмплингом неизбежно возникает эффект алиасинга, приводящий к нестабильности предсказаний при малых сдвигах входных данных.
3. Разработана методология тестирования и система метрик для количественной оценки степени инвариантности моделей к пространственным сдвигам, включающая:
 - Алгоритм генерации последовательностей изображений с контролируемыми субпиксельными сдвигами, позволяющий точно измерять влияние смещений на различные аспекты работы модели

- Метрики стабильности векторов признаков (косинусное сходство)
 - Метрики стабильности предсказаний (дрейф уверенности для классификаторов, стабильность IoU и позиции центра для детекторов)
 - Инструменты визуализации для качественного анализа эффектов смещения
4. Проведено экспериментальное исследование влияния субпиксельных сдвигов на стабильность работы различных CNN-архитектур, включая классификационные модели (VGG16, ResNet50) и модели детекции объектов (YOLOv5). Экспериментально показано, что:
- Стандартные CNN-архитектуры демонстрируют значительную нестабильность даже при минимальных (субпиксельных) сдвигах входных данных
 - Степень нестабильности зависит от глубины сети, количества операций субдискретизации и размера рецептивного поля
 - Модели детекции объектов особенно чувствительны к смещениям, что проявляется в значительном дрейфе предсказанных ограничивающих рамок и падении показателя IoU
5. Реализованы и сравнены различные методы повышения инвариантности к сдвигам:
- Классический анти-алиасинг с использованием BlurPool, который заменяет стандартные операции пулинга на последовательность фильтра низких частот и операции субдискретизации
 - Translation Invariant Polyphase Sampling (TIPS), который использует полифазное разложение для сохранения информации при даунсэмплинге

- Гибридные подходы, комбинирующие различные методы в зависимости от структуры сети
6. Проведено аблационное исследование для выявления влияния различных факторов на пространственную инвариантность, которое показало, что:
- Увеличение размера рецептивного поля улучшает инвариантность к сдвигам, но не решает проблему полностью
 - Замена max-pooling на average-pooling снижает нестабильность, но все еще требует анти-алиасинга для достижения высокой инвариантности
 - Оптимальный размер ядра для анти-алиасинга составляет 3×3 или 5×5 , обеспечивая баланс между инвариантностью и вычислительной сложностью

На основе проведенного исследования могут быть сформулированы следующие практические рекомендации:

1. Для задач, требующих высокой стабильности предсказаний при малых изменениях входных данных (например, системы автономного вождения, медицинская диагностика), рекомендуется использовать модифицированные архитектуры с методами анти-алиасинга, предпочтительно с применением TIPS в критических слоях.
2. При ограниченных вычислительных ресурсах оптимальным выбором является применение BlurPool с ядром 3×3 в последних слоях сверточной части сети, что обеспечивает значительное повышение инвариантности при минимальном увеличении вычислительной сложности.
3. Для моделей детекции объектов, особенно YOLOv5, рекомендуется модификация backbone-сети с применением техник анти-алиасинга, что существенно улучшает стабильность позиционирования ограничивающих рамок при сдвигах объектов.

4. При обучении новых моделей рекомендуется включать в набор данных аугментации с субпиксельными сдвигами, что повышает робастность модели к подобным преобразованиям даже без структурных изменений архитектуры.

Полученные результаты имеют как теоретическую, так и практическую значимость для разработки более надежных систем компьютерного зрения. Предложенные методики оценки инвариантности и модификации архитектур могут быть непосредственно применены при создании и оптимизации CNN для различных прикладных задач.

Направления дальнейших исследований могут включать:

- Разработку методов обеспечения инвариантности к более широкому классу геометрических преобразований (поворот, масштабирование, аффинные преобразования)
- Исследование проблемы инвариантности в современных трансформерных архитектурах для компьютерного зрения
- Разработку специализированных аппаратных решений для эффективной реализации методов анти-алиасинга и полифазной выборки
- Интеграцию методов обеспечения инвариантности в фреймворки автоматического поиска архитектур нейронных сетей

Таким образом, проведенное исследование не только вносит вклад в понимание фундаментальной проблемы пространственной инвариантности в CNN, но и предлагает практические решения, повышающие надежность и стабильность систем компьютерного зрения в различных прикладных областях.