

Проблемы искусственного интеллекта. 2026. N 1 (40). С. 75-88

Problems of Artificial Intelligence. 2026;1(40):75-88.

Искусственный интеллект и машинное обучение

Научная статья

УДК 004.932.2

doi: 10.24412/2413-7383-2026-1-40-75-88

К.И. Морев

АО «Научно-конструкторское бюро вычислительных систем»

ул. 1-я Линия, 144-а, г. Таганрог, Ростовская область, 347936, Российская Федерация.

УНИВЕРСАЛЬНЫЙ ГЕНЕРАТОР ПРИЗНАКОВ ДЛЯ РЕШЕНИЯ ЗАДАЧ КОМПЬЮТЕРНОГО ЗРЕНИЯ

K. I. Morev

JSC «Scientific design bureau of computing systems»

1st Line street, 144-a, Taganrog, Rostov region, 347936, Russian Federation.

UNIVERSAL BACKBONE FOR SOLVING COMPUTER VISION TASKS

В работе исследуется возможность использования единого генератора признаков для решения нескольких задач компьютерного зрения – обнаружения, классификации объектов и извлечения локальных дескрипторов особых точек. Исследуются архитектуры YOLOv5s, MobileNetV2 и ResNet-50 в качестве моделей, предоставляющих генератор признаков для решения всех упомянутых задач. Эксперименты показывают, что применение общего генератора признаков позволяет существенно ускорить обучение и эксплуатацию алгоритмов компьютерного зрения, сохраняя качество решения.

Ключевые слова: компьютерное зрение, универсальный генератор признаков, архитектуры нейронных сетей, особые точки изображений, сопоставление особых точек.

The paper explores the possibility of using a single backbone to solve several computer vision tasks – detecting and classifying objects, extracting descriptors of image keypoints. The YOLOv5s, MobileNetV2, and ResNet-50 architectures are being investigated as models providing a backbone for solving all the mentioned tasks. Experiments show that the use of a common feature generator can significantly speed up the learning and operation of computer vision algorithms, while maintaining the quality of tasks solving.

Keywords: computer vision, universal backbone, neural network architectures, image keypoints, keypoints matching.

Введение

На сегодняшний день системы видеоаналитики эволюционировали по применяемым методам анализа наблюдаемой сцены от классических алгоритмов обработки изображений к современным подходам, основанным на технологиях искусственного интеллекта [1]. Вместе с тем, все методы компьютерного зрения основаны на построении решающих правил в некотором признаковом пространстве, которое формируется ядрами свертки (классические подходы) или генераторами признаков (современные нейросетевые решения). Следовательно, обобщенная структура любого алгоритма компьютерного зрения может быть разделена на две основные части: извлечение признаков и формирование решения на основе этих признаков. Современные методы анализа видеосцены активно используют нейросетевые архитектуры различные по сложности и назначению. Однако, в ряде приложений необходимо решать несколько задач одновременно, например, обнаружение объектов, их классификация, семантическая сегментация, определение собственного местоположения и т.д. [2]. Использование отдельных генераторов признаков для каждой задачи требует значительных вычислительных ресурсов и времени как на этапе разработки алгоритмов, так и на этапе их применения. Отсюда возникает потребность в использовании общего генератора признаков для решения различных задач компьютерного зрения.

Цель данной работы – провести исследование возможности использования единого генератора признаков для решения задач обнаружения объектов с последующей классификацией и формирование дескрипторов для особых точек на изображении. Исследуется применение архитектур YOLOv5s [3], ResNet-50 [4] и MobileNetV2 [5] в качестве общего генератора признаков: один и тот же набор признаков используется для формирования решений как для обнаружения и классификации объектов, так и для вычисления описаний особых точек и их сопоставления. Для достижения поставленной цели решены следующие задачи:

1. Изучение архитектуры и свойств исследуемых архитектур нейронных сетей;
2. Выделение основной сверточной части как генератора признаков для каждой из исследуемых архитектур;
3. Формирование способа формирования вектора-дескриптора для особой точки на основе сформированных общим генератором признаков;
4. Проведение экспериментов и формирование выводов.

Исследуемые задачи и метрики качества

В рамках данного исследования требуется решать несколько задач компьютерного зрения на основе одного общего набора признаков, сформированного по одному входному изображению. А именно требуется решать задачу обнаружения и классификации объектов, а также сопоставления дескрипторов особых точек соседних в видеоряде изображений.

Задача обнаружения и классификации объектов

Задача обнаружения и классификации заключается в указании положения и размера объекта на изображении, выдачи решения о принадлежности обнаруженного объекта к одному из классов заранее известного списка [1], [6]. Для каждого входного изображения от системы обнаружения требуется сформировать множество ограничивающих прямоугольников и соответствующих им вероятностей принадлежности к известным классам. Ниже подробнее описывается методика оценки качества решения задачи обнаружения и классификации.

Используемая метрика качества агрегирует два показателя: ошибку обнаружения и ошибку классификации.

Ошибка обнаружения определяется с помощью коэффициента Жаккара [7], который в контексте цифровой обработки изображений используется для оценки величины сходства истинного и вычисленного системой положения объекта в кадре (положение определяется обрамляющим прямоугольником). Формула для вычисления ошибки обнаружения выглядит следующим образом:

$$IoU = \frac{Intersection}{Union} , \tag{1}$$

где *Intersection* - площадь объединения обрамляющих прямоугольников, а *Union* - площадь объединения этих прямоугольников.

Ошибка классификации определяется на основе матрицы сопоставления возможных результатов классификатора и истинной метки класса. Эта таблица содержит четыре ячейки:

TruePositive (*TP*)– количество истинно позитивных срабатываний;

FalsePositive (*FP*)– количество ложно позитивных срабатываний;

FalseNegative (*FN*)– количество ложно негативных срабатываний;

TrueNegative(*TN*)–количество истинно негативных срабатываний.

На основании значений в таблице сопоставления, формируются показатели ошибок классификации: точность (*Precision*) и полнота (*Recall*). Формула для вычисления *Precision*:

$$Precision = \frac{TP}{TP + FP} . \tag{2}$$

И формула для вычисления *Recall*:

$$Recall = \frac{TP}{TP + FN} . \tag{3}$$

Далее для оценки качества решения задачи классификации строится *Precision-Recall* функция, отображающая зависимость величин при различных порогах уверенности алгоритма классификации. И подсчет площади под этой прямой позволяет вычислить критерий *Average precision*, характеризующий качество решения задачи классификации для одного класса. А для определения качества решения задачи классификации на множестве классов критерий *Average Precision* масштабируется до *Mean Average Precision* [8].

Таким образом, в рамках данного исследования, качестве основной метрики качества решения задачи обнаружения и классификации является интегральная характеристика *Mean Average Precision*, объединяющая (1), (2) и (3):

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} . \tag{4}$$

Задача сопоставления особых точек

Задача сопоставления особых точек заключается в назначении соответствий между особыми точками пары изображений [9]. Сопоставление проводится с целью вычисления линейного оператора *H* (матрицы гомографии), связывающего плоскости изображений и обеспечивающего преобразование одного изображения в другое.

Используемые для сопоставления дескрипторы особых точек должны позволять сопоставлять локальные точки между двумя изображениями при условии значительных изменений масштаба, поворота, смещения и прочих видов преобразований плоскости изображения. Качество сопоставления оценивается с помощью метрики точности сопоставления *Match Accuracy*:

$$MA = \frac{N_{correct}}{N_{all}} \tag{5}$$

где $N_{correct}$ – число пар особых точек, в которых найденная сопоставлением по дескрипторам особая точка является той же точкой 3-Д пространства, а $N_{incorrect}$ – общее число сопоставленных особых точек. В рамках исследования, невозможно гарантировать точного попадания в ту же точку 3-Д пространства. Поэтому если по результатам сопоставления выбранная точка соседнего изображения попадает в заданную окрестность относительно истинного положения в соответствии с известной матрицей гомографии, сопоставление считается истинным. Радиус окрестности принимается $R=3$ пикселя.

Исследуемые архитектуры и их генераторы признаков

В качестве общих генераторов признаков в работе используются три архитектуры: YOLOv5s, MobileNetV2 и ResNet-50. Они различаются как по вычислительной сложности, так и по дискриминативной способности формируемых признаков, что позволяет искать компромисс между сложностью получения признаков и их эффективностью [10, 11].

Архитектура YOLOv5s

Архитектура YOLOv5s традиционно включает три компонента: «позвоночник» или генератор признаков, «шея» для подготовки извлеченных признаков и «голова» для формирования решения. В качестве генератора признаков используется модифицированная сеть CSPDarknet [12], основанная на извлечении признаков различных масштабов с последующим их слиянием, такой подход уменьшает общее количество вычислительных операций и позволяет сохранить дискриминативную способность извлекаемых признаков.

В контексте исследования универсальности выделяемых признаков, сверточная часть YOLOv5s (три выхода различных масштабов генератора признаков оригинальной архитектуры) рассматривается как общий генератор признаков, будем называть его генератор признаков YOLOv5s-gen. Разно-масштабная природа формируемых признаков YOLOv5s-gen делает их удобными для решения различных задач, в которых требуется одновременно анализировать как глобальный контекст сцены, так и локальные детали, что важно как для обнаружения объектов и их классификации, так и для извлечения локальных дескрипторов [13], [14].

На рисунке 1 ниже приведена схема части архитектуры YOLOv5s, назначаемая общим генератором признаков YOLOv5s-gen.

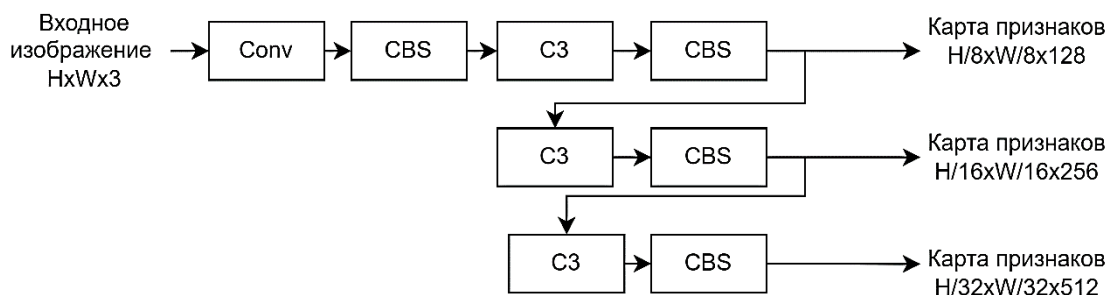


Рисунок 1 – Схема генератора YOLOv5s-gen

Ниже на рисунке 2 приведены более подробные описания используемых в рисунке 1 блоков.

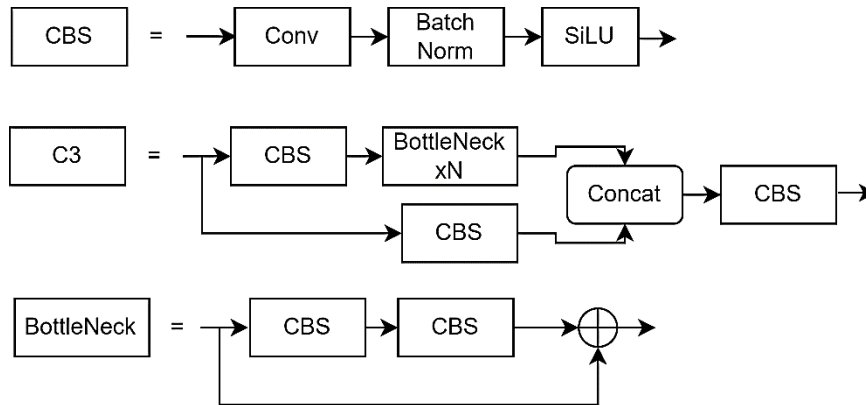


Рисунок 2 – Схемы блоков, используемых в генераторе YOLOv5s-gen

Архитектура MobileNetV2

MobileNetV2 представляет собой лёгкую сверточную архитектуру, ориентированную на использование в составе систем с ограниченными вычислительными ресурсами. Например, мобильные устройства или встраиваемые системы [15]. Основной элемент архитектуры – инвертированный остаточный блок (inverted residual block) в комбинации с bottleneck-блоком. Внутри такого блока входное узкое представление сначала расширяется в пространство повышенной размерности с помощью $1 \times 1 \times 1$ -свёртки, затем к нему применяется depthwise-свёртка (поканально, а не в глубину), после чего выполняется проекция обратно в пространство меньшей размерности линейной $1 \times 1 \times 1$ -свёрткой.

Также архитектура MobileNetV2 использует сепарабельные фильтры, что позволяет существенно снизить число операций по сравнению с классическими сверточными архитектурами при сопоставимом качестве решения ряда задач компьютерного зрения [5]. MobileNetV2 широко используется как генератор признаков для классификации, детекции и сегментации, особенно в системах с жесткими требованиями по времени и условиям работы.

Аналогично генератору признаков YOLOv5s-gen, из оригинальной архитектуры MobileNetV2 была выделена часть, отвечающая за генерацию признаков. Будем называть ее генератор признаков MobileNetV2-gen. Ниже на рисунке 3 приведена схема архитектуры MobileNetV2-gen которая используется в качестве универсального генератора признаков.

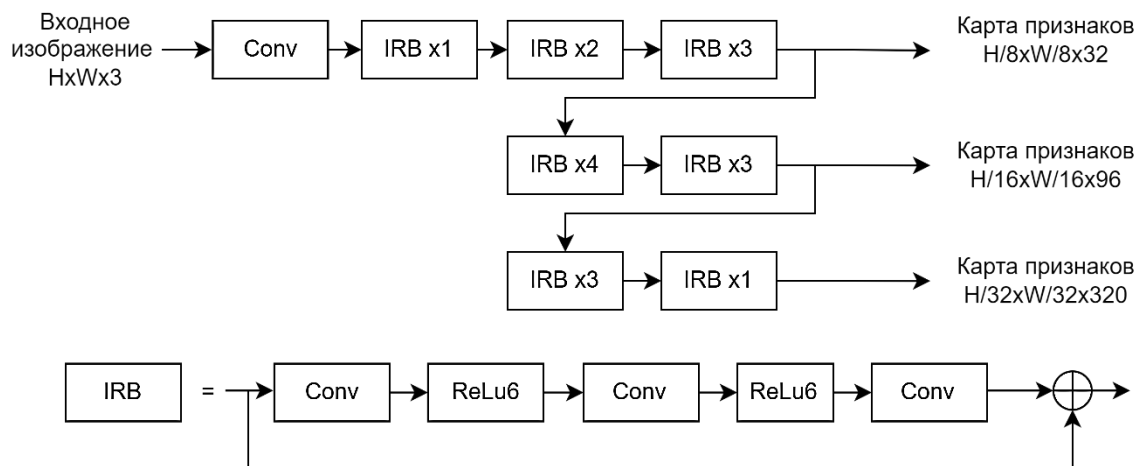


Рисунок 3 – Схема генератора MobileNetV2-gen

Архитектура ResNet-50

ResNet-50 – ставшая классической архитектура сверточной нейронной сети с остаточными связями (residual connections). Семейство сетей ResNet первыми решило проблему затухания градиента при обучении глубоких сетей [4]. Сеть содержит 50 слоёв, организованных в 4 группы bottleneck-блоков (3, 4, 6, 3 блока в группе соответственно).

Блоки остаточной связи (residual connection block) позволяют обучать глубокие нейронные сети за счет передачи сигнала не только через сверточные слои, но и напрямую между слоями. Передача напрямую между слоями позволяет сохранить значения градиентов при вычислении обратного распространения ошибки даже для глубоких и многослойных сетей, где более глубокие сверточные слои формируют признаки всё более высокого уровня абстракции: от контуров и градиентов до сложных структурных паттернов. Архитектура ResNet-50 широко используется как генератор признаков в различных задачах компьютерного зрения [16].

Изначально архитектура ResNet была разработана для решения задачи классификации изображений, но ее основная часть – генератор признаков выделена для экспериментов. Будем называть эту часть ResNet-50-gen. Ниже на рисунке 4 приведена структурная схема ResNet-50-gen.

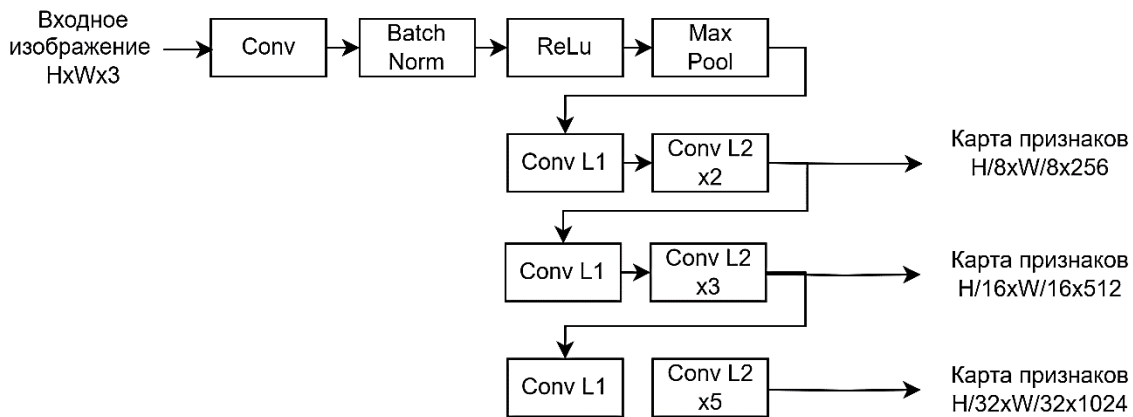


Рисунок 4 – Схема генератора ResNet-50-gen

На рисунке 5 приведены схемы подробно описывающие структурные элементы на рисунке 4.

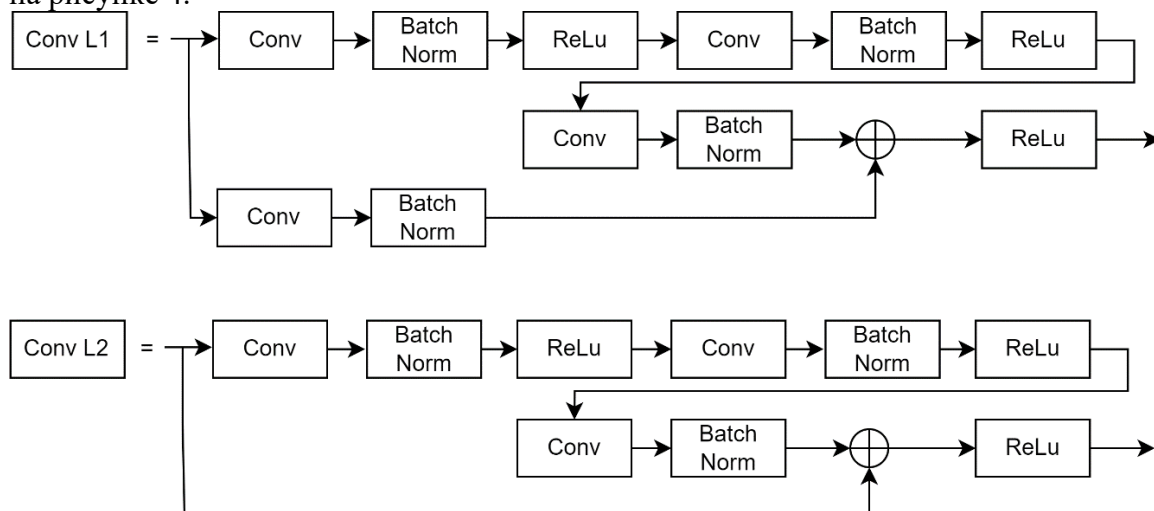


Рисунок 5 – Схемы блоков, используемых в генераторе ResNet-50-gen

Структура экспериментов

Наборы данных для экспериментов

В рамках эксперимента, с целью оценки точности решения задачи обнаружения и классификации используется открытый набор данных COCO 2017 [17]. COCO 2017 – это большой набор данных, содержащий более 200000 размеченных изображений. В нем присутствуют объекты 80 различных классов в том числе автомобили, самолеты, домашние животные и т.д. С целью обучения и проверки результатов обучения COCO 2017 набор данных был разделен на две части – обучающая и тестовая выборки. Где тестовая составляет 5% от общего объема данных [18]. Таким образом, обучение всех исследуемых архитектур происходит на обучающей выборке из COCO 2017, а тестирование и получение оценок качества производится на тестовой выборке [19]. На тестовой выборке производится оценка качества как для обнаружения и классификации, так и для сопоставления особых точек изображений.

Для оценки качества решения задачи сопоставления особых точек, для каждого изображения из тестовой выборки синтетически формируется второе изображение путём применения заданной матрицы гомографии H , что позволяет заранее определить истинные соответствия между точками двух изображений. Для точки x на первом изображении её истинное положение x' на втором изображении вычисляется как

$$x' = H * x. \quad (6)$$

Положения особых точек на исходном изображении определяется детектором особых точек SIFT [20]. А особые точки на изображении-паре не выделяются детектором, но их положения вычисляются из известной матрицы гомографии, в соответствии с выражением (6).

Важно отметить, что информация о значениях в применяемой матрице гомографии предоставляет априорную информацию о положении особых точек, выделенных на оригинальном, неискаженном изображении. Использование этой априорной информации для вычисления дескрипторов особых точек на синтезированном изображении позволяет исключить ошибку обнаружения и локализации особых точек на изображении.

Основная экспериментальная архитектура

В рамках исследования и построения комплексной цепочки вычислений для решения сразу нескольких задач компьютерного зрения, был разработан общий вид модульной архитектуры. На рисунке 6 приведена схема общей структуры базового алгоритма для проведения экспериментов.

В рамках данной работы в качестве генераторов признаков исследуются упомянутые выше YOLOv5s-gen, MobileNetV2-gen и ResNet-50-gen. Таким образом, на рисунке 6 блок «оригинальная архитектура» поочередно заменяется одной из трех соответствующих генератору признаков архитектурой.

Для решения назначенных задач используются независимые модули, формирующие решения на основе общего набора признаков. Решение задачи обнаружения и классификации объектов выполняется в классическом для компьютерного зрения ключе: исследуемые архитектуры нейронных сетей обучаются задаче обнаружения и классификации объектов [21]. В качестве обучающей выборки используется описанная выше выборка из COCO 2017 набора данных.

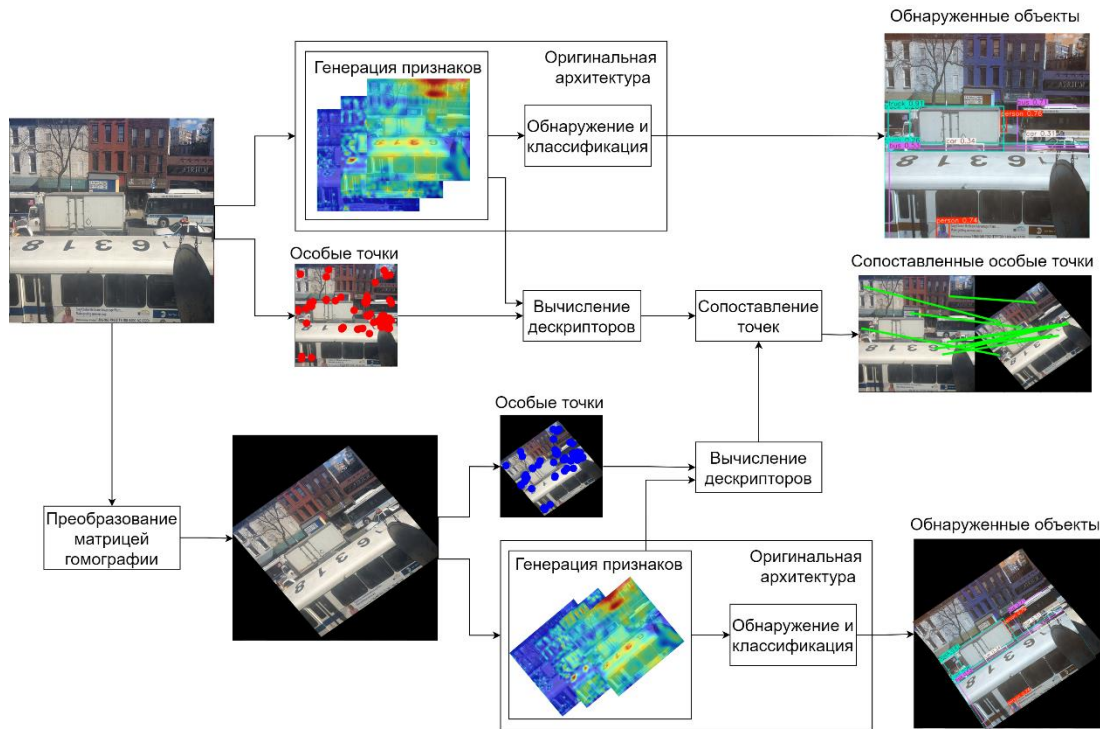


Рисунок 6 – Схема общей структуры базового алгоритма

Вычисление локальных дескрипторов выполняется описанным ниже способом. Каждый из генераторов формирует три представления входного изображения, соответствующие деталям определенного масштаба изображения. А именно формируются три тензора в 8, 16 и 32 раза меньше в ширину и высоту чем входное изображение, но глубина тензоров значительно больше глубины тензора входного изображения. Далее, для каждой обнаруженной особой точки с координатами (i, j) на входном изображении определяется масштабный тензор наиболее соответствующий точке по масштабу признаков. Так для особых точек, извлеченных с 1 слоя пирамиды масштабов выбирается тензор в 8 раз меньший входного изображения, для особых точек с 2 и 3 слоя тензор в 16 раз меньший и для остальных тензор в 32 раза меньше входного изображения. В выбранном тензоре вычисляется положение особой точки (i_n, j_n) путем применения масштаба тензора. И в вычисленных координатах ведется вычисление дескриптора методом билинейной интерполяции по 4-м ближайшим соседям вокруг вычисленных вещественных координат (i_n, j_n) .

Таким образом, для формирования комплексной архитектуры, решающей несколько задач, сначала ведется обучение классической архитектуры нейронной сети для обнаружения и классификации объектов. Далее, полученные в ходе обучения весовые коэффициенты фиксируются и сформированный в ходе обучения генератор признаков используется не только для классификации, но и для вычисления дескрипторов особых точек. Особые точки выделяются алгоритмом SIFT и для вычисленных точек формируются дескрипторы путем выделения вектора-строки из выходного тензора соответствующего особой точке масштаба. Данный сценарий позволяет оценить полученные непосредственно для обнаружения и классификации генераторы признаков в качестве генераторов дескрипторов особых точек. Также подобная комбинация вычислений позволяет ощутимо снизить количество операций на вычисление дескрипторов особых точек на прямую. Например, для вычисления классического

дескриптора SIFT требуется около 36864 операций в числах с плавающей точкой, а для формирования дескриптора из карты признаков используемой нейронной сети занимает около 4606 аналогичных операций.

Результаты экспериментов и выводы

При проведении экспериментов, основным набором данных являлся COCO 2017. Для оценки качества решения задач обнаружения и классификации использовалась описанная выше тестовая выборка из набора данных COCO 2017. Примерный объем обучающей и тестовой выборки: 118000 изображений в обучающей и 5000 изображений в тестовой. Набор данных содержит объекты 80 различных классов.

С целью оценки качества решения задачи сопоставления особых точек, использовались синтезированные на базе COCO 2017 изображения в соответствии с описанным в разделе «Наборы данных для экспериментов» методом. Все входные изображения в тестовой выборке нормализованы по яркости и приведены к размеру 640×640.

Также с целью оценки эффективности измерялись вычислительная сложность (GFLOPs) одного прямого прохода нейронной сети, предоставляющей генератор признаков.

Ниже, на рисунке 7 представлены результаты экспериментов и отображена зависимость полученных в ходе эксперимента оценок качества (в соответствии с выражениями (4) и (5), только характеристика mAP для наглядности представления нормирована в диапазоне [0;1]) от вычислительной сложности нейронной сети, предоставляющей признаки для решения задач. На рисунке 7 звездам соответствуют результаты обнаружения и классификации, ромбам – результаты сопоставления особых точек. Синей горизонтальной линией на рисунке 7 отображен уровень качества решения задачи сопоставления особых точек с использованием дескрипторов SIFT.

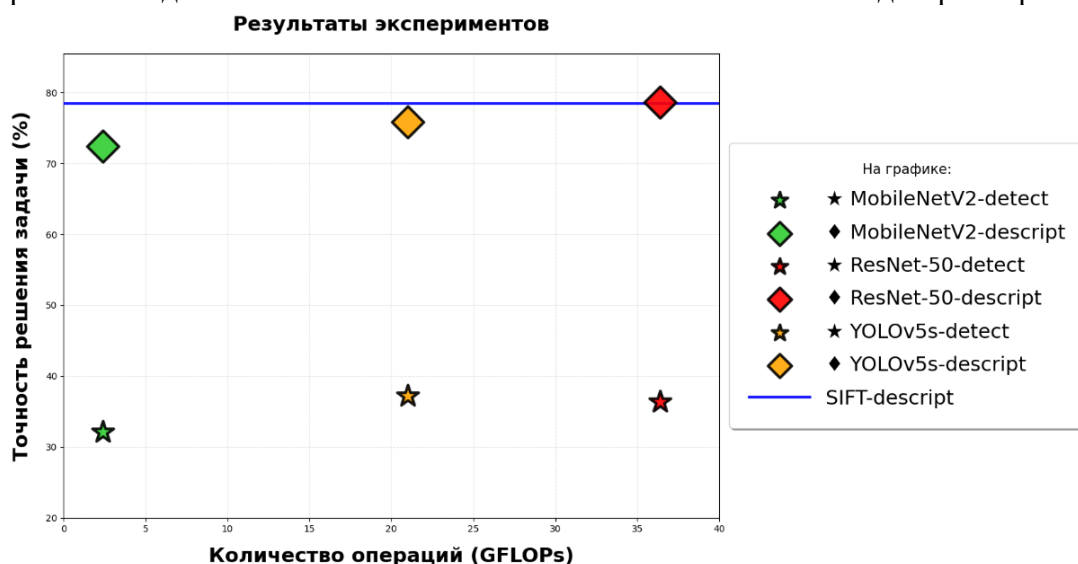


Рисунок 7 – Результаты экспериментов

Полученные результаты подтверждают возможность использования единого генератора признаков для решения разнородных задач компьютерного зрения – обнаружения, классификации объектов и сопоставления особых точек изображений.

В ходе экспериментов архитектура YOLOv5s показала наилучший показатель $mAP(4)$ 37.2%, что ожидаемо, поскольку модель была опубликована позднее остальных

и включает в себя больше технологических приемов. Однако точность сопоставления локальных точек (75.8%) оказывается ниже, чем у ResNet-50, что может быть связано с чрезмерной специализацией признаков на объектно-ориентированные паттерны и взаимосвязи, менее подходящие для точного сопоставления особых точек на уровне пикселей.

Архитектура MobileNetV2 уступает конкурентам в метриках качества (mAP 32.1%, точность $MA(5)$ 72.4%), но значительно выигрывает по вычислительной эффективности. Лёгкая архитектура с свёртками в глубину позволяет эффективно выделять признаки, отражающие как глобальный контекст (обнаружение, классификация), так и детали сцены (сопоставление точек).

Модель ResNet-50 показала наивысшую точность сопоставления особых точек (mAP 35.8%, MA 78.6%), однако составив наибольшую вычислительную сложность. Карты признаков, формируемые генератором ResNet-50-gen, содержат больше слоев (глубина) чем прочие исследуемые архитектуры. Вероятно, глубина тензора признаков определяет информативность этих признаков.

Продемонстрированные на рисунке 7 оценки также приведены ниже в таблице 1.

Таблица 1 - Результаты проведенных экспериментов

Архитектура	Вычислительная сложность ($GFLOPs$)	Обнаружение и классификация, mAP	Сопоставление, MA
Yolo5s	21	37.2	75.8
MobileNet	2.4	32.1	72.4
ResNet-50	36.4	36.4	78.6
SIFT	-	-	78.4

Заключение

В работе продемонстрирована возможность эффективного использования единого генератора признаков для решения разнородных задач компьютерного зрения – обнаружения с классификацией объектов и извлечения локальных дескрипторов для сопоставления особых точек.

Рассмотренные генераторы признаков (YOLOv5s-gen, MobileNetV2-gen, ResNet-50-gen) обеспечивают сопоставимое качество решения задачи сопоставления особых точек по сравнению с классическим, признанным сообществом исследователей в области компьютерного зрения, методом SIFT. Кроме того, использование одного генератора признаков для различных задач позволяет ощутимо снизить вычислительные затраты.

Стоит отметить, что подход с использованием единого генератора признаков применим и к другим задачам компьютерного зрения (сегментация, реидентификация) и позволяет существенно сократить ресурсы при создании мультизадачных систем. Перспективы дальнейших исследований включают совместное обучение агрегаторов признаков для решения различных задач, адаптивный механизм обратного распространения ошибки учитывающий вклад различных агрегаторов на комплексную точность работы системы.

Список литературы

1. ГОСТ 59385—2021. Информационные технологии. Искусственный интеллект. Ситуационная видеоаналитика. Термины и определения. Москва: Стандартинформ, 2021. 16 с.
2. Уланов С.И., Криводубский О.А., Никитина А.А. Анализ возможностей применения летательных аппаратов в контроле состояния местности // Проблемы искусственного интеллекта. 2024. № 1(32). С. 94-102. URL: [http://paijournal.guiaidn.ru/ru/2024/1\(32\)-7.html](http://paijournal.guiaidn.ru/ru/2024/1(32)-7.html) (дата обращения: 12.12.2025).
3. Jocher G. YOLOv5 by Ultralytics [Электронный ресурс] // GitHub. URL: <https://github.com/ultralytics/yolov5> (дата обращения: 04.01.2026).
4. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770-778.
5. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 4510-4520.
6. Зуев В.М. Сравнение обнаружения объектов средствами искусственного интеллекта в сравнении с классическими методами // Проблемы искусственного интеллекта. 2024. № 3(34). С. 30-35. URL: [http://paijournal.guiaidn.ru/ru/2024/3\(34\)-3.html](http://paijournal.guiaidn.ru/ru/2024/3(34)-3.html) (дата обращения: 16.01.2026).
7. Баженов А.Н., Тельнова А.Ю. Обобщение коэффициента Жаккара для анализа данных с интервальной неопределённостью // Измерительная техника. 2022. № 12. С. 15-22.
8. Радиолокационные системы: учебник / В.П. Бердышев, Е.Н. Гарин, А.Н. Фомин [и др.]; под общ. ред. В.П. Бердышева; разработ.: Центр обучающих систем ИнТК СФУ. Версия 1.0. URL: https://vii.sfu-kras.ru/images/libs/Radiolokacionnye_sistemy_SFU_elektronnyu_resurs.pdf (дата обращения: 15.01.2026).
9. Морев К.И., Боженюк А.В. Сопоставление изображений по особым точкам различных категорий // Известия ЮФУ. Технические науки. 2020. № 3. С. 192-201.
10. Наумов М.А., Карповский А.Ю. Методы повышения точности классификации объектов при построении системы зрения для исследования ограниченных пространств // Проблемы искусственного интеллекта. 2025. № 2(37). С. 4-12. DOI: 10.24412/2413-7383-2025-2-37-4-12.
11. Исследование применения моделей Mask R-CNN и Segment Anything Model (SAM) для инстанс-сегментации мезенхимных стволовых клеток на микрофотографиях / М.В. Солопов, Е.С. Чечехина, А.Г. Попандопуло [и др.] // Проблемы искусственного интеллекта. 2025. № 2(37). С. 21-29. DOI: 10.24412/2413-7383-2025-2-37-21-29.
12. Wang C.-Y., Liao H.-Y.M., Yeh I.-H., Wu Y.-H., Chen P.-Y., Hsieh J.-W. CSPNet: A New Backbone that can Enhance Learning Capability of CNN // arXiv. 2019. arXiv:1911.11929. URL: <https://arxiv.org/abs/1911.11929> (дата обращения: 30.12.2025).
13. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 779-788.
14. Wang C.-Y., Bochkovskiy A., Liao H.-Y.M. Scaled-YOLOv4: Scaling Up Common Objects Detection // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021. P. 117-125.
15. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications / A.G. Howard, M. Zhu, B. Chen [и др.] // arXiv. 2017. arXiv:1704.04861.
16. Nagpal P., Bhinge S.A., Shitole A. A Comparative Analysis of ResNet Architectures // 2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON). Bangalore, India, 2022. P. 1-8. DOI: 10.1109/SMARTGENCON56628.2022.10083966.
17. Microsoft COCO 2017: Common Objects in Context / T.-Y. Lin, M. Maire, S. Belongie [и др.] // European Conference on Computer Vision. Springer, 2014. P. 740-755.
18. Хакимов Р.С., Нижникова О.Л., Близна М.В. К вопросу о разработке системы аннотирования данных для задач компьютерного зрения // Проблемы искусственного интеллекта. 2024. № 3(34). С. 70-79. URL: [http://paijournal.guiaidn.ru/ru/2024/3\(34\)-8.html](http://paijournal.guiaidn.ru/ru/2024/3(34)-8.html) (дата обращения: 18.01.2026).
19. Устенко В.Ю., Бондаренко В.И. Разработка программного комплекса аннотирования данных для задач компьютерного зрения: объектно-ориентированный подход на основе winforms // Проблемы искусственного интеллекта. 2024. № 4(35). С. 151-163. URL: [http://paijournal.guiaidn.ru/ru/2024/4\(35\)-13.html](http://paijournal.guiaidn.ru/ru/2024/4(35)-13.html) (дата обращения: 20.01.2026).
20. Lowe D. Distinctive image features from scale invariant keypoints // International Journal of Computer Vision. 2004. Vol. 60, № 2. P. 91-110.
21. Bozhenyuk A., Morev K. Influence of Membership Function Type on the Accuracy of Video Tracking Algorithm // Communications in Computer and Information Science. 2021. Vol. 1395. P. 444-453. DOI: 10.1007/978-981-16-1480-4_40.

References

1. GOST 59385-2021. Information technology. Artificial intelligence. Situational video analytics. Terms and definitions. Moscow: Standartinform, 2021. 16 p.
2. Ulanov S.I., Krivodubskiy O.A., Nikitina A.A. Analysis of the possibilities of using unmanned aerial vehicles for terrain condition monitoring // *Problems of Artificial Intelligence*. 2024. No. 1(32). P. 94-102. URL: [http://paijournal.guiaidn.ru/ru/2024/1\(32\)-7.html](http://paijournal.guiaidn.ru/ru/2024/1(32)-7.html) (accessed: 12.12.2025).
3. Jocher G. YOLOv5 by Ultralytics. URL: <https://github.com/ultralytics/yolov5> (accessed: 04.01.2026).
4. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 770-778.
5. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. P. 4510-4520.
6. Zuev V.M. Comparison of object detection using artificial intelligence methods versus classical methods // *Problems of Artificial Intelligence*. 2024. No. 3(34). P. 30-35. URL: [http://paijournal.guiaidn.ru/ru/2024/3\(34\)-3.html](http://paijournal.guiaidn.ru/ru/2024/3(34)-3.html) (accessed: 16.01.2026).
7. Bazhenov A.N., Telnova A.Yu. Generalization of the Jaccard coefficient for analyzing data with interval uncertainty // *Measurement Techniques*. 2022. No. 12. P. 15-22.
8. Radar Systems: textbook / V.P. Berdyshev, E.N. Garin, A.N. Fomin et al.; ed. by V.P. Berdyshev; developers: Center for Training Systems InTK SFU. Version 1.0. URL: https://vii.sfu-kras.ru/images/libs/Radiolokacionnye_sistemy_SFU_elektronnyy_resurs.pdf (accessed: 15.01.2026).
9. Morev K.I., Bozhnyuk A.V. Image matching by feature points of various categories // *Izvestiya YuFU. Technical Sciences*. 2020. No. 3. P. 192-201.
10. Naumov M.A., Karpovskiy A.Yu. Methods for improving object classification accuracy in vision systems for confined space exploration // *Problems of Artificial Intelligence*. 2025. No. 2(37). P. 4-12. DOI: 10.24412/2413-7383-2025-2-37-4-12.
11. Investigation of Mask R-CNN and Segment Anything Model (SAM) for instance segmentation of mesenchymal stem cells in micrographs / M.V. Solopov, E.S. Chechekhin, A.G. Popandopulo et al. // *Problems of Artificial Intelligence*. 2025. No. 2(37). P. 21-29. DOI: 10.24412/2413-7383-2025-2-37-21-29.
12. Wang C.-Y., Liao H.-Y.M., Yeh I.-H., Wu Y.-H., Chen P.-Y., Hsieh J.-W. CSPNet: A New Backbone that can Enhance Learning Capability of CNN // *arXiv*. 2019. arXiv:1911.11929. URL: <https://arxiv.org/abs/1911.11929> (accessed: 30.12.2025).
13. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 779-788.
14. Wang C.-Y., Bochkovskiy A., Liao H.-Y.M. Scaled-YOLOv4: Scaling Up Common Objects Detection // *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2021. P. 117-125.
15. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications / A.G. Howard, M. Zhu, B. Chen et al. // *arXiv*. 2017. arXiv:1704.04861.
16. Nagpal P., Bhinge S.A., Shitole A. A Comparative Analysis of ResNet Architectures // *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. Bangalore, India, 2022. P. 1-8. DOI: 10.1109/SMARTGENCON56628.2022.10083966.
17. Microsoft COCO 2017: Common Objects in Context / T.-Y. Lin, M. Maire, S. Belongie et al. // *European Conference on Computer Vision*. Springer, 2014. P. 740-755.
18. Khakimov R.S., Nizhnikova O.L., Blizno M.V. On the development of data annotation system for computer vision tasks // *Problems of Artificial Intelligence*. 2024. No. 3(34). P. 70-79. URL: [http://paijournal.guiaidn.ru/ru/2024/3\(34\)-8.html](http://paijournal.guiaidn.ru/ru/2024/3(34)-8.html) (accessed: 18.01.2026).
19. Ustenko V.Yu., Bondarenko V.I. Development of data annotation software complex for computer vision tasks: object-oriented approach based on WinForms // *Problems of Artificial Intelligence*. 2024. No. 4(35). P. 151-163. URL: [http://paijournal.guiaidn.ru/ru/2024/4\(35\)-13.html](http://paijournal.guiaidn.ru/ru/2024/4(35)-13.html) (accessed: 20.01.2026).
20. Lowe D. Distinctive image features from scale invariant keypoints // *International Journal of Computer Vision*. 2004. Vol. 60, No. 2. P. 91-110.
21. Bozhnyuk A., Morev K. Influence of Membership Function Type on the Accuracy of Video Tracking Algorithm // *Communications in Computer and Information Science*. 2021. Vol. 1395. P. 444-453. DOI: 10.1007/978-981-16-1480-4_40.

RESUME

K.I. Morev

Universal backbone for solving computer vision tasks

Background: the current level of development of computer vision technologies, in particular, is due to the rapid growth of computing capabilities. Today, researchers in the field of computer vision have the widest opportunities to implement a variety of ideas for combining and aggregating existing approaches and solutions, and creating their own technologies. However, in practice, vision complexes are equipped with limited computing resources: it is required to ensure operation at extreme temperatures, ensure heat dissipation, and withstand vibration and shock loads. This determines the relevance of the main goal of this study: the development of a universal backbone that makes it possible to generate features in one full cycle to solve several computer vision problems.

Materials and methods: various scientific research methods were used during the research: experiment, modeling, measurement, comparison, analysis. For the purposes of the study, the COCO2017 open dataset was used, as well as an additional dataset synthesized on its basis. To conduct experiments and simulations, a software package was developed in the Python3.11 language. Calculations were performed on a stationary computer of class G3 (GOST R 71224-2025).

Results: using a feature map of a convolutional neural network generator to solve problems unrelated to its training procedure turns out to be a very real prospect. Even without using complex aggregators, it was possible to solve the problem of image matching based on descriptors from the neural network feature map. This approach saves more than 30,000 operations on a single singular point in the image. And the number of points processed in real systems reaches 5,000.

Conclusion: the conducted research should be considered successful, and the results successful. There are opportunities to optimize algorithms for multitasking systems of various purposes, including unmanned vehicles, aircraft, situational video analytics systems. In the future, it makes sense to explore the possibilities of multitasking and unsupervised learning techniques.

РЕЗЮМЕ

К.И. Морев

Универсальный генератор признаков для решения задач компьютерного зрения

Справочная информация: современный уровень развития технологий компьютерного зрения, в частности, обусловлен бурным ростом вычислительных возможностей. Сегодня перед исследователями в области компьютерного зрения открываются широчайшие возможности по реализации разнообразных идей по комбинированию и агрегации имеющихся подходов и решений, созданию собственных технологий. Однако, на практике комплексы технического зрения оснащены ограниченными вычислительными ресурсами: требуется гарантировать работу в условиях экстремальных температур, обеспечивать отвод тепла, выдерживать требования к вибрационным и ударным нагрузкам. Этим обуславливается актуальность основной цели данного исследования: разработки универсального генератора признаков, позволяющего за один полный цикл сформировать признаки для решения сразу нескольких задач компьютерного зрения.

Материалы и методы: при проведении исследования использовались различные методы научного исследования: эксперимент, моделирование, измерение, сравнение, анализ. В целях исследования использовался открытый набор данных COCO2017,

а также синтезированный на его основе дополнительный набор данных. Для проведения экспериментов и моделирования был разработан программный комплекс на языке Python3.11. Вычисления производились на стационарной ЭВМ класса ГЗ (ГОСТ Р 71224 – 2025).

Результаты: использование карты признаков генератора сверточной нейронной сети для решения задач, не связанных с процедурой ее обучения, оказывается вполне реальной перспективой. Даже без использования сложных агрегаторов и комплексирования, удалось решить задачу сопоставления изображений на основе дескрипторов из карты признаков нейронной сети. Такой подход позволяет экономить более 30000 операций на одной особой точке изображения. А количество обрабатываемых в реальных системах точек доходит до 5000.

Вывод: проведенное исследование следует считать удачным, а результаты успешными. Открываются возможности по оптимизации алгоритмов для многозадачных систем разнообразного назначения, в том числе беспилотные автомобили, летательные аппараты, системы ситуационной видеоаналитики. В будущем имеет смысл исследовать возможности многозадачного обучения и техники обучения без учителя.

Морев Кирилл Иванович – программист АО «Научно-конструкторское бюро вычислительных систем», 347936, Ростовская область, г. Таганрог, ул. 1-я Линия, 144-а, тел. +7 (8634) 68-25-60.

E-mail: morev-ki@ya.ru. ORCID – orcid.org/0000-0003-2622-358X.

Научные интересы: цифровая обработка изображений, машинное обучение, нечеткая логика, комплексные системы интеллектуальной видеоаналитики.

Morev Kirill Ivanovich – programmer JSC «Scientific design bureau of computing systems», 347936, Rostov region, Taganrog, 1st Line street, 144-a. +7 (8634) 68-25-60.

E-mail: morev-ki@ya.ru. ORCID – orcid.org/0000-0003-2622-358X.

Scientific interests: digital image processing, machine learning, fuzzy logic, complex intellectual video analytics systems.

Статья поступила в редакцию 11.02.2026