УДК 004.93
doi:10.31799/1684-8853-2022-3-9-19

Научные статьи
Articles

Метод создания синтетических наборов данных для обучения нейросетевых моделей распознаванию объектов

С. Ю. Пчелинцев[а], *аспирант, orcid.org/0000-0001-9195-8318, veselyrojer@mail.ru*
М. А. Ляшков[а], *аспирант, orcid.org/0000-0002-7793-7024*
О. А. Ковалева[а,б], *доктор техн. наук, профессор, orcid.org/0000-0003-0735-6205*
[а]*Тамбовский государственный университет им. Г. Р. Державина, Интернациональная ул., 33, Тамбов, 392000, РФ*
[б]*Тамбовский государственный технический университет, Советская ул., 106, Тамбов, 392000, РФ*

Введение: *недостаток обучающих данных приводит к низкой точности распознавания визуальных образов. Одним из способов решения данной проблемы является использование реальных данных в сочетании с синтетическими.* ***Цель:*** *повышение эффективности распознавания образов системами компьютерного зрения путем использования для обучения смешанных (реальных и синтетических) данных; снижение временных затрат на подготовку данных обучающей выборки.* ***Результаты:*** *на базе предложенного метода генерации синтетических изображений построена интеллектуальная информационная система, позволяющая генерировать репрезентативные выборки большого объема, содержащие изображения, предназначенные для обучения нейронных сетей распознаванию образов. Разработано программно-алгоритмическое обеспечение генератора синтетических изображений для обучения нейросетей. Разработанный генератор имеет модульную архитектуру, что позволяет легко модифицировать, удалять или добавлять отдельные этапы в конвейер генерирования синтетических изображений. Отдельные параметры (как освещение или размытие) для генерируемых изображений можно настраивать. Идея эксперимента заключалась в сравнении точности распознавания образов для нейронной сети, обученной на различных обучающих выборках. Комбинация реальных и синтетических данных при обучении модели показала наилучшую эффективность распознавания. Искусственные обучающие выборки, в которых масштаб фоновых объектов примерно равен масштабу объекта интереса, а количество объектов интереса в кадре выше, оказались эффективнее других искусственных обучающих выборок. Изменение фокусного расстояния камеры в сцене генерации синтетических изображений не оказало влияния на эффективность распознавания.* ***Практическая значимость:*** *предложенный метод генерирования изображений позволяет создать большой набор искусственных данных для обучения нейронных сетей распознаванию образов за меньшее время, чем заняло бы создание такого же набора реальных данных.*

Ключевые слова *— нейронные сети, искусственный интеллект, машинное обучение, синтетические наборы данных, генерирование изображений.*

Для цитирования: Пчелинцев С. Ю., Ляшков М. А., Ковалева О. А. Метод создания синтетических наборов данных для обучения нейросетевых моделей распознаванию объектов. *Информационно-управляющие системы*, 2022, № 3, с. 9–19. doi:10.31799/1684-8853-2022-3-9-19

For citation: Pchelintsev S. Y., Liashkov M. A., Kovaleva O. A. Method for creating synthetic data sets for training neural network models for object recognition. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 3, pp. 9–19 (In Russian). doi:10.31799/1684-8853-2022-3-9-19

Введение

Способность обнаруживать объекты в сложных условиях является фундаментальной для многих задач машинного зрения и робототехники. Современные архитектуры сверточных нейронных сетей, такие как Faster R-CNN, SSD, R-FCN, Yolo9000, YoloV5 и RetinaNet, достигли очень впечатляющих результатов в области распознавания образов. Однако обучение таких моделей с миллионами параметров требует огромного количества маркированных обучающих данных для достижения конкурентных результатов. Очевидно, что создание таких массивных наборов данных стало одним из основных ограничений этих подходов: они требуют участия человека и много времени, очень дороги и подвержены ошибкам.

Обучение с использованием искусственных данных снижает нагрузку, затрачиваемую на сбор данных и их аннотацию [1]. Кроме того, оно решает некоторые проблемы формирования обучающей выборки [2]. Теоретически можно генерировать бесконечное количество обучающих изображений с большими вариациями, где разметка осуществляется автоматически. Процесс генерирования данных называют аугментацией [3]. Обучение с искусственными образцами позволяет точно контролировать рендеринг изображений и, следовательно, различные свойства набора данных [4].

Существует понятие области, или домена (domain), — характеристики способа сбора данных для обучения моделей искусственного интеллекта. Так, в частности, изображения для

обучения распознаванию образов могут быть получены с камеры либо программно генерироваться, как в предлагаемом методе, и относиться, таким образом, к разным областям.

Очевидно, что изображения с камеры и сгенерированные изображения могут и, по существу, должны отличаться. Поэтому модели, обученные на данных, собранных в одной области, обычно имеют низкую точность в других областях. Такое явление называется доменным сдвигом (domain shift), или доменным разрывом (domain gap). Для решения этой проблемы можно повышать реалистичность обучающих данных, смешивать искусственные и реальные данные, использовать архитектуры с предварительно обученными экстракторами признаков или применять трансферное обучение [5].

Предлагаемое в данной работе решение использует рандомизацию доменов (domain randomization) [6]. Суть данного подхода заключается в том, что генерируются заведомо нереалистичные данные, и, таким образом, реальные данные можно рассматривать как частный случай сгенерированных искусственных данных. Совсем недавно эта концепция была расширена за счет добавления реальных фоновых изображений, смешанных со случайными сценами, и дополнительно улучшена за счет фотореалистичного рендеринга [7]. Несмотря на то, что такой подход дал впечатляющие результаты, основным его недостатком по-прежнему остается зависимость от реальных данных. Распространенным подходом к повышению эффективности обнаружения также является расширение реального обучающего набора данных путем добавления искусственных данных. И хотя эти методы демонстрируют значительное улучшение по сравнению с использованием только реальных данных, они по-прежнему требуют как минимум реальных фоновых изображений для конкретной предметной области.

Существует также подход композиции изображения для создания искусственных изображений путем комбинирования вырезанных объектов из разных изображений [8]. Преимущество заключается в использовании данных из одной и той же области, поскольку вырезанные объекты являются копиями реальных изображений и близко соответствуют характеристикам реального мира. Основное ограничение этих подходов состоит в том, что они требуют выполнения громоздкого процесса захвата изображений объектов со всех возможных точек обзора и их маскировки. В частности, эти методы не позволят создавать изображения из разных ракурсов или разных условий освещения, если набор для обучения объекта фиксирован. Это явное ограничение.

Другие направления работы используют фотореалистичный рендеринг и реалистичные композиции сцены для преодоления разрыва в предметной области путем синтеза изображений, максимально приближенных к реальному миру [9]. Хотя эти методы показали многообещающие результаты, они сталкиваются с множеством проблем. Во-первых, создание фотореалистичных обучающих изображений требует сложных механизмов рендеринга, а также значительных вычислительных ресурсов. Во-вторых, реалистичная композиция сцены сама по себе является нетривиальной задачей. В-третьих, современные движки рендеринга, применяемые для создания искусственных сцен, в значительной степени используют преимущества системы человеческого восприятия, чтобы обмануть человеческий глаз. Однако эти уловки не обязательно работают в нейронных сетях, и, следовательно, необходимы дополнительные усилия для преодоления доменного разрыва.

Существуют исследования, в которых были использованы генеративные состязательные сети для дальнейшего преодоления доменного разрыва [10, 11]. Однако такие подходы значительно усложняют работу, поскольку их сложно разработать и обучить. Насколько нам известно, они еще не применялись для задач обнаружения.

Другое направление работ использует адаптацию предметной области или переносное обучение, чтобы преодолеть разрыв между искусственной и реальной предметной областью [12, 13]. Это может быть достигнуто путем объединения двух предикторов, по одному для каждого домена, или путем объединения данных из двух доменов. Адаптация предметной области и переносное (transferred) обучение имеют применения, выходящие далеко за рамки переноса искусственных данных в реальные. Тем не менее они требуют значительного количества реальных данных.

Концепция рандомизации доменов для преодоления доменных разрывов предполагает использование нереалистичных текстур для рендеринга искусственных сцен, чтобы обучить детектор объектов, который обобщается на реальный мир. Другое направление работ [14] объединяет рандомизацию домена и рендеринг фотореалистичного изображения. В нем генерируют два типа данных: во-первых, искусственные изображения со случайными отвлекающими факторами и вариациями, которые кажутся неестественными с реальными фотографиями в качестве фона, и, во-вторых, фотореалистичные визуализации случайно сгенерированных сцен с использованием физического движка для обеспечения физической правдоподобности. Комбинация этих двух типов данных дает значительное улучшение по сравнению с одним источником данных и расши-

ряет области применения сети. Также возможно использование структурированной рандомизации домена, в которой сеть может учитывать контекст. В контексте структурированных сред, таких как уличные сцены, это дает самые современные результаты, но неприменимо к таким сценариям, как выбор предмета из коробки, где нет четких пространственных отношений между расположением различных объектов.

Целью исследования является повышение эффективности распознавания образов системами компьютерного зрения путем использования для обучения смешанных (реальных и синтетических) данных, а также снижение временных затрат на подготовку данных обучающей выборки. Задачами исследования являются:

— разработка метода генерирования синтетических изображений, предназначенных для обучения нейросетевой модели распознаванию образов;

— сравнение эффективности обучения нейронных сетей с применением реальных, синтетических и смешанных данных.

Реализация предлагаемого метода генерирования данных велась в среде Unity 3D с использованием пакета Unity Perception.

Предлагаемый метод генерации искусственных обучающих данных

Метод генерации фона разработан в соответствии с тремя принципами: максимизировать фоновый беспорядок, минимизировать риск отображения в сети одного и того же фонового изображения дважды и создать фоновые изображения из элементов, подобных по масштабу объектам переднего плана. Проведенные эксперименты показывают, что эти принципы помогают создавать обучающие данные, которые позволяют сетям запоминать форму и внешний вид объектов, сводя к минимуму шансы научиться отличать искусственные объекты переднего плана от фоновых объектов просто по различным свойствам, таким как, например, различные размеры объекта или распределение шума.

Суть предлагаемого метода заключается в создании трехмерной сцены в виртуальной среде. На сцену случайным образом добавляются различного рода трехмерные объекты, а также источники освещения. Их параметры изменяются случайным образом в заданных интервалах. Кроме того, добавляются шум и размытие, а также могут меняться внутренние параметры камеры. После всех этих манипуляций осуществляются захват кадра и его сохранение. Потом происходит очистка сцены, и процесс генерирования кадра начинается заново, пока не будет достигнуто целевое количество кадров. На рис. 1 представлен предлагаемый алгоритм генерирования искусственных обучающих данных.

Перед запуском алгоритма программы требуется подготовить трехмерные модели объектов интереса, а также объектов фонового слоя и слоя помех. В качестве 3D-моделей объектов интереса используются модели объектов, поиск которых будет осуществляться системой распознавания образов, обученной на генерируемом наборе данных. Так, для системы, распознающей дорожные знаки, необходимо подготовить 3D-модели распознаваемых знаков. В слоях фоновом и окклюзии (помех) используются одни и те же модели, но масштаб может различаться. Это множество моделей не пересекается со множеством моделей объектов интереса и, по существу, может содержать модели объектов и из других предметных областей (это допустимо в соответствии с используемой концепцией рандомизации домена). Более того, в качестве таких моделей могут выступать

■ ***Рис. 1.*** Алгоритм генерирования изображений
■ ***Fig. 1.*** The algorithm of images generation

даже геометрические примитивы (кубы, шары и т. п.), но все они должны быть текстурированы, наложение текстур на эти объекты осуществляется на этапах создания соответствующих слоев. Но текстуры объектов интереса не изменяются.

Каждая обучающая выборка создается путем смешивания трех слоев изображения: искусственного фонового слоя, слоя объектов переднего плана, построенного в соответствии со стратегией учебной программы, и, наконец, последнего слоя, содержащего преграды.

Фоновый слой создается из текстурированных 3D-моделей M_{bg}, которые не пересекаются с набором объектов переднего плана M_{fg}:

$$M_{bg} \cap M_{fg} = \varnothing. \quad (1)$$

Все трехмерные фоновые модели изначально уменьшены и масштабированы таким образом, чтобы они вписывались в единичную сферу. При создании фона происходит последовательный выбор области на заднем плане, где не был визуализирован другой объект, и визуализация случайного фонового объекта в этой области. Каждый фоновый объект визуализируется в произвольной позе, и процесс повторяется до тех пор, пока весь фон не будет покрыт искусственными фоновыми объектами. Ключом к созданию фона является размер проецируемых фоновых объектов, который определяется по размеру объекта переднего плана. Поэтому мы генерируем рандомизированное изотропное масштабирование S, которое применяем к нашим унифицированным 3D-моделям перед их рендерингом. Мы используем масштабирование для создания объектов таким образом, чтобы размер их проекций на плоскость изображения соответствовал размеру среднего объекта переднего плана. Конкретнее, мы вычисляем диапазон масштабирования $S = [S_{\min}, S_{\max}]$, представляющий масштабы, которые могут применяться к объектам так, что они появляются в пределах [0,9; 1,5] размера, соответствующего среднему размеру объекта переднего плана. Затем для каждого фонового изображения мы создаем случайное подмножество $S_{bg} \subset S$, чтобы гарантировать, что мы создаем не только фоновые изображения с объектами, равномерно распределенными по всем размерам, но также и изображения, в основном, с большими или маленькими объектами. Значение S_{bg} изотропного масштабирования теперь выбирается случайным образом из S, так что размеры фоновых объектов в изображении распределяются равномерно.

Для каждой фоновой сцены дополнительно конвертируем текстуру каждого объекта в пространство HSV. Значение цветового тона H вычисляется по формуле

$$H = \begin{cases} 0^\circ, \Delta = 0 \\ 60^\circ \times \left\{ \dfrac{G-B}{\Delta} \bmod 6 \right\}, C_{\max} = R \\ 60^\circ \times \left\{ \dfrac{B-R}{\Delta} + 2 \right\}, C_{\max} = G \\ 60^\circ \times \left\{ \dfrac{R-G}{\Delta} + 4 \right\}, C_{\max} = B \end{cases}, \quad (2)$$

где R, G, $B \in [0, 1]$ — насыщенности красного, желтого, синего цветов;

$$C_{\max} = \max(R, G, B); \quad (3)$$

$$C_{\min} = \min(R, G, B); \quad (4)$$

$$\Delta = C_{\max} - C_{\min}. \quad (5)$$

Значение насыщенности S вычисляется по формуле

$$S = \begin{cases} 0, C_{\max} = 0 \\ \dfrac{\Delta}{C_{\max}}, C_{\max} \neq 0 \end{cases}. \quad (6)$$

Значение яркости V вычисляется по формуле

$$V = C_{\max}. \quad (7)$$

Затем мы случайным образом изменяем значение оттенка и, наконец, конвертируем оттенок обратно в RGB, чтобы разнообразить фон и обеспечить хорошее распределение цветов фона:

$$(R', G', B') = \begin{cases} (C+m, X+m, m), 0^\circ \le H < 60^\circ \\ (X+m, C+m, m), 60^\circ \le H < 120^\circ \\ (m, C+m, X+m), 120^\circ \le H < 180^\circ \\ (m, X+m, C+m), 180^\circ \le H < 240^\circ \\ (X+m, m, C+m), 240^\circ \le H < 300^\circ \\ (C+m, m, X+m), 300^\circ \le H < 360^\circ \end{cases}, \quad (8)$$

где R', G', $B' \in [0, 1]$ — новые значения насыщенностей красного, желтого, синего;

$$C = V \times S; \quad (9)$$

$$X = C \times \left(1 - \left| \frac{H}{60^\circ} \bmod 2 - 1 \right| \right); \quad (10)$$

$$m = B - C. \quad (11)$$

Этап генерирования объектов переднего плана (рис. 2) является ключевым в работе данного алгоритма.

Для успешного распознавания требуется, чтобы каждый объект интереса присутствовал в обучающей выборке достаточное количество раз и в различных положениях. Для равномерного рас-

■ ***Рис. 2.*** Алгоритм генерирования объектов переднего плана

■ ***Fig. 2.*** The algorithm of foreground objects generation

пределения поз объекта была придумана стратегия, именуемая обучающим планом. Для каждого объекта рекурсивно генерируются 20 вращений таким образом, чтобы множество всех вращений объекта представляло одну из 20 граней выпуклого правильного икосаэдра. Таким образом, каждая вершина представляет собой отдельный вид объекта, определяемый двумя вращениями вне плоскости. Кроме того, мы выбираем расстояние, на котором визуализируем объект переднего плана обратно пропорционально его проецируемому размеру, чтобы гарантировать приблизительное линейное изменение пиксельного покрытия объекта между последовательными уровнями масштабирования. Начинаем с ближайшего к камере расстояния и постепенно переходим к самому дальнему. В результате каждый объект изначально кажется самым большим на изображении, поэтому его легче изучить для сети. В последующих итерациях, удаляясь от камеры, объекты становятся меньше и сложнее для распознавания. Для каждого масштаба перебираем все возможные вращения вне плоскости, а для каждого вращения вне плоскости перебираем все повороты в плоскости. Когда у нас есть масштаб, вращение вне плоскости и в плоскости, перебираем все объекты и визуализируем каждый из них с заданной позой в случайном месте с использованием равномерного распределения. После обработки всех объектов, всех вращений в плоскости и вне плоскости переходим к следующему уровню масштабирования.

Для рендеринга мы разрешаем обрезку объектов переднего плана по границам изображения до 50 %. Кроме того, допускаем перекрытие между каждой парой объектов переднего плана до 30 %. Для каждого объекта случайным образом пытаемся разместить его $n = 100$ раз на сцене переднего плана. Если он не может быть помещен в сцену из-за нарушения ограничений обрезки или перекрытия, прекращаем обработку текущей сцены переднего плана и начинаем со следующей. Для последующей сцены переднего плана начинаем с того места, где остановились в последней сцене.

Также создается слой окклюзии, где случайным объектам из множества объектов, используемых в фоновом слое, разрешается перекрывать объекты переднего плана. Это делается путем определения ограничивающего прямоугольника каждого визуализированного объекта переднего плана и визуализации случайно выбранного закрывающего объекта в однородном случайном месте внутри этого ограничивающего прямоугольника. Затеняющий объект масштабируется случайным образом так, что его проекция покрывает определенный процент соответствующего объекта переднего плана (в диапазоне от 10 до 30 % объекта переднего плана). Поза и цвет закрывающего объекта рандомизируются так же, как и для фоновых объектов.

Имея фон, передний план и слой окклюзии, объединяем все три слоя в одно комбинированное изображение: слой окклюзии визуализируется поверх слоя переднего плана, а результат визуализируется поверх фонового слоя.

Далее добавляем случайные света со случайными искажениями оттенка света, а также с изменяемым направлением источников света.

Наконец, добавляем белый шум и размываем изображение с помощью размытия Гаусса, в котором случайным образом выбираются размер ядра r и стандартное отклонение σ:

$$G(r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-r^2/(2\sigma^2)}, \qquad (12)$$

где $G(r)$ — функция Гаусса.

Таким образом, фон, передний план и закрывающие части имеют одни и те же свойства изображения, что противоречит подходам, в которых смешиваются реальные изображения и искусственные визуализации. Это делает невозможным для сети отличать передний план от фона только по атрибутам, специфичным для их домена.

Поскольку конечная цель использования сгенерированных данных — поиск экземпляров объекта, требуется обеспечить геометрическую корректность рендеринга наших объектов. Для этого настраиваются внутренние параметры ка-

меры — фокусное расстояние и главная точка. Допускаются небольшие случайные изменения этих параметров.

Эксперименты

В ходе экспериментов фокусируемся на обнаружении и распознавании дорожных знаков, соответствующих ГОСТу [15, 16].

Современные модели обнаружения объектов состоят из экстрактора признаков, который нацелен на проецирование изображений из необработанного пиксельного пространства в многоканальное пространство признаков, и нескольких весов, которые решают различные аспекты проблем обнаружения, такие как сужение ограничительной рамки и классификация. В настоящей работе мы используем популярную архитектуру Faster R-CNN с экстрактором функций InceptionResNet [17]. Веса экстрактора признаков были предварительно обучены на наборе данных ImageNet [18]. Используется общедоступная реализация GoogleFaster R-CNN с открытым исходным кодом [19].

В качестве успешного распознавания в ходе экспериментов засчитывалось 50 %-е пересечение площади рамок прогнозируемого положения объекта с его истинным положением. В качестве метрик для оценки эффективности распознавания в наших экспериментах используются усредненная средняя точность (mean average precision, mAP) и усредненный средний отзыв (mean average recall, mAR).

$$mAP = \frac{1}{K}\sum_{i=1}^{K} AP_i. \qquad (13)$$

Здесь K — количество классов распознаваемых объектов;

$$AP = \int_0^1 p(r)\mathrm{d}r, \qquad (14)$$

где r — это значение отзыва, процента истинно положительных результатов, обнаруженных среди всех истинных фактов; $p(r)$ — соответствующая точность, процент правильных положительных прогнозов.

$$mAR = \frac{1}{K}\sum_{i=1}^{K} AR_i. \qquad (15)$$

Здесь

$$AR = 2\int_{0,5}^1 R(o)\mathrm{d}o, \qquad (16)$$

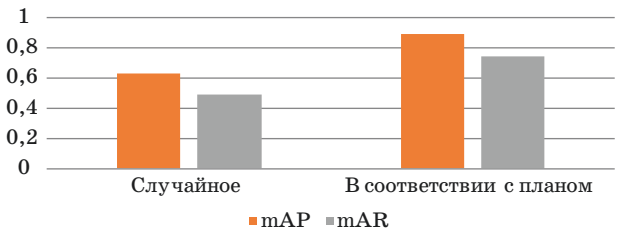
где o — это IoU, уровень перекрытия между реальным и предсказанным системой положением ограничивающей рамки объекта; $R(o)$ — соответствующий отзыв.

В следующих экспериментах мы подчеркиваем преимущества нашей стратегии обучения по учебному плану и исследуем влияние относительного масштаба фоновых объектов по отношению к объектам переднего плана, влияние количества объектов переднего плана, визуализируемых на изображение, влияние композиции фона и, наконец, эффекты случайных цветов и размытия. Модели обучаются с использованием распределенного асинхронного стохастического градиентного спуска.

Данные генерируются в соответствии с учебным планом, который гарантирует, что все модели представлены одинаково в плане позы и в условиях с возрастающей сложностью. В этом эксперименте мы сравниваем две модели Faster R-CNN, инициализированные с одинаковыми весами, первая из которых обучается с использованием полной выборки случайных поз, а другая — в соответствии с нашей стратегией учебного плана. Очевидные преимущества нашего подхода по сравнению со стратегией простой случайной выборки демонстрирует рис. 3.

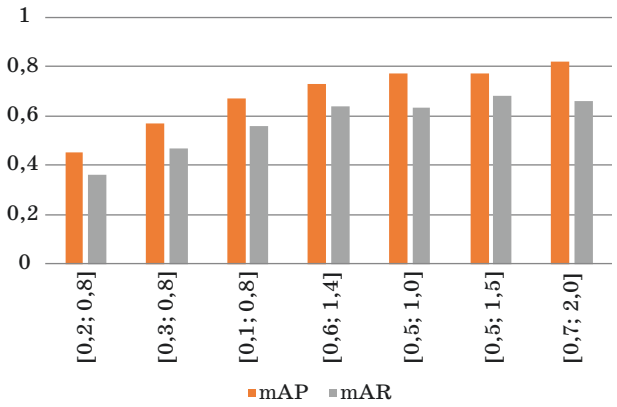
В следующих экспериментах мы анализируем влияние изменения относительного диапазона масштабов фоновых объектов по отношению к объектам переднего плана. На рис. 4 показано, что наилучшие результаты могут быть получены для диапазона, в котором фоновые объекты имеют такой же или больший размер, чем объекты переднего плана. Использование меньших диапазонов масштабирования дает фоновые изображения, которые больше похожи на текстуры, что упрощает сети распознавание объектов переднего плана.

В следующем эксперименте мы изучаем влияние количества объектов переднего плана, отображаемых в обучающих изображениях. Видно (рис. 5), что большее количество объектов переднего плана дает лучшую производительность. Мы устанавливаем только верхний предел количества объектов переднего плана, нарисованных на одном изображении, поэтому среднее количе-



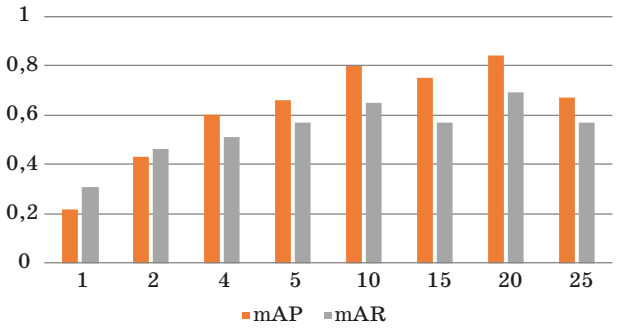
■ ***Рис. 3.*** Сравнение стратегий плана обучения и случайных поз

■ ***Fig. 3.*** Comparison of training plan and random pose strategies



■ ***Рис. 4.*** Сравнение распознаваний с разными масштабами фоновых объектов

■ ***Fig. 4.*** Comparison of recognitions with different background objects scale



■ ***Рис. 5.*** Сравнение распознаваний с разным количеством объектов интереса в кадре

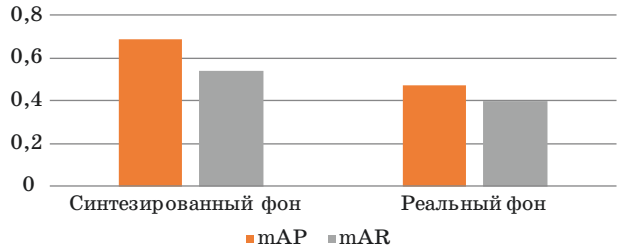
■ ***Fig. 5.*** Comparison of recognitions with different number of objects of interest in one frame

ство объектов обычно ниже. В частности, на начальных этапах изучения учебного плана можно уместить в среднем только 8–9 объектов на одном изображении.

В следующем эксперименте сравнивалось обучение на синтетических данных, в которых фоновый слой был сгенерирован путем добавления множества мелких текстурированных объектов (как и предполагает предложенный алгоритм), с обучением на данных, в которых фон представлял собой реальное изображение, растянутое на весь кадр. Обучение на наборе данных, в котором фоновые слои сгенерированы, показывает лучшие результаты (рис. 6).

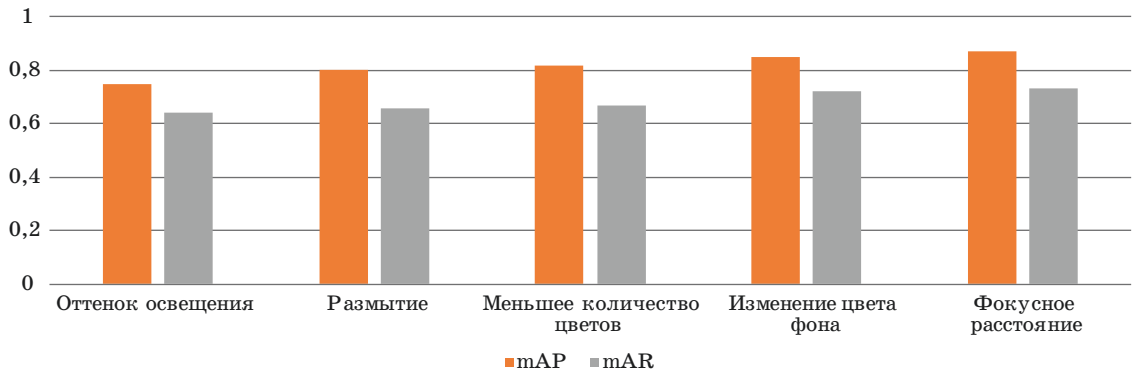
В серии экспериментов (рис. 7) мы исследовали влияние отдельных шагов в конвейере генерации изображений. Было обнаружено, что наибольшее влияние оказывают размытие и случайный оттенок источника света. Наименее важным оказалось изменение фокусного расстояния камеры.

В следующей серии экспериментов мы сравниваем временные затраты на создание реальных и синтетических наборов данных. Весь сбор реальных данных осуществлялся с помощью камеры смартфона. Было отобрано 1200 снимков, вошедших в итоговый набор для обучения распознаванию на реальных данных. Разрешение каждого снимка составляет 1280 × 720 пикселей. На всех этих изображениях содержатся случайные подмножества объектов, подлежащих распознаванию. Фон разный на всех изображениях, освеще-



■ ***Рис. 6.*** Сравнение обучения с полностью искусственным фоном и обучения с реальным фоном на данных в обучающей выборке

■ ***Fig. 6.*** Comparison of training with a completely artificial background and training with a real background on the data in the training sample



■ ***Рис. 7.*** Влияние характеристик изображения на распознавание

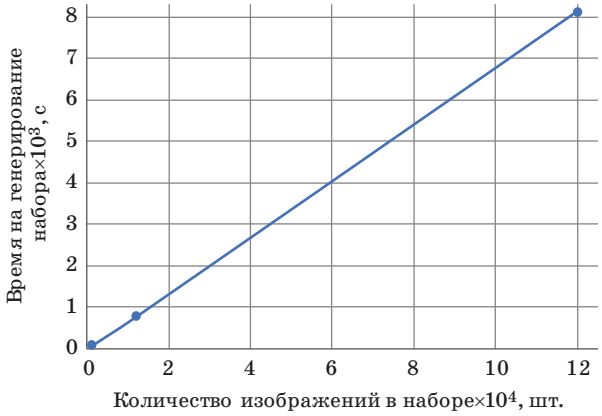
■ ***Fig. 7.*** Influence of image characteristics on recognition

ние тоже отличается, сами фотографии сделаны с различных ракурсов. Это нужно, чтобы соблюсти равномерность данных в тестовой выборке для лучших результатов распознавания.

Разметка данных на изображениях велась вручную с использованием программы VGG Image Annotator. Результат разметки корректировался сторонним наблюдателем, что позволило исправить возникшие в ходе разметки ошибки. Количество времени, затраченного на получение реальных изображений, составило около 10 ч, для маркировки обучающего набора потребовалось примерно 170 ч, а еще 5 ч было потрачено на исправления. Стоит отметить, что для добавления дополнительных реальных данных в набор всегда требуются дополнительные действия по их сбору и разметке, кроме того, будет необходимо создать снимки, сочетающие новые и старые объекты в одном кадре.

Подготовка синтетических данных заняла в общей сложности 5 ч. За это время созданы путем сканирования 3D-модели дорожных знаков — объектов интереса, частично представленных на рис. 8, а также были загружены из открытых источников 3D-модели и текстуры объектов для фонового и помехового слоев. Добавление дополнительных моделей — процесс единоразовый: от пользователя требуется просто добавить модель в проект и запустить генерирование новых данных.

В рамках следующей серии экспериментов измерялось время, затрачиваемое на генерирование набора данных определенного размера. Соотношение количества сгенерированных изображений в наборе ко времени их генерирования представлено на рис. 9. Проводить дополнительную разметку для сгенерированных данных не требуется, поскольку она осуществлялась автоматически на этапе генерирования средствами пакета Unity Perception. Таким образом, с учетом времени на подготовку данных получение набора

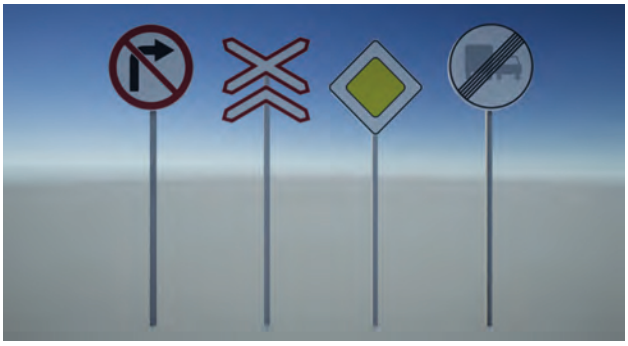


■ ***Рис. 9.*** Зависимость времени генерирования от размера набора данных

■ ***Fig. 9.*** Dependence of generation time on data set size

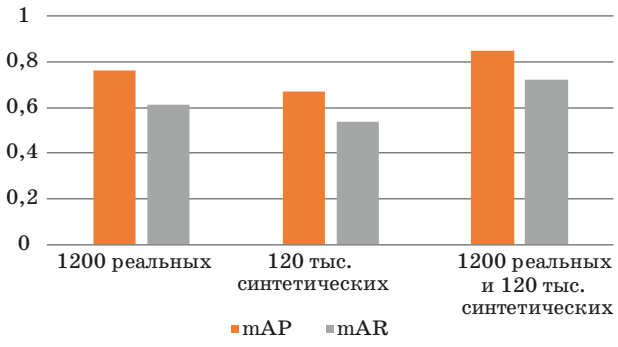
из 1200 реальных изображений заняло 185 ч, что в 37 раз больше, чем время на получение синтетического набора данных аналогичного размера. Российский набор дорожных знаков содержит более 100 тыс. изображений [16]. Генерирование схожего количества изображений, как видно из рис. 9, заняло 2 ч 15 мин (без учета времени, затраченного на подготовку к генерированию). Если и этих данных окажется недостаточно для обучения модели, то можно сгенерировать дополнительные.

В следующем эксперименте мы сравниваем эффективность распознавания при обучении на реальных, синтетических и смешанных данных. Были заняты все 1200 собранных реальных изображений, а также все 120 тыс. изображений, сгенерированных с использованием нашего алгоритма. Как видно по рис. 10, обучение на полностью синтетических данных позволяет распознавать объекты, однако распознавание даже на в стократ меньшем наборе реальных данных может быть эффективнее. Вместе с тем объеди-



■ ***Рис. 8.*** Часть используемых моделей дорожных знаков на одной сцене

■ ***Fig. 8.*** Some of the used road signs models on the same scene



■ ***Рис. 10.*** Сравнение подходов с обучением на реальных данных, на синтетических и на смешанных данных

■ ***Fig. 10.*** Comparison of approaches with training on real data, on synthetic data and on mixed data

■ Характеристики наборов данных
■ Characteristics of datasets

Данные	Количество изображений, шт.	mAP	mAR	Время на формирование
Реальные	1200	0,76	0,61	185 ч
Синтетические	120 000	0,67	0,54	7 ч 15 мин
Смешанные	132 000	0,85	0,72	192 ч 15 мин

нение реальных и синтетических данных в одну обучающую выборку повышает эффективность распознавания.

Данные по времени генерирования наборов данных, их размер и эффективность распознавания с точки зрения характеристик mAP и mAR приведены в итоговой таблице.

Заключение

В работе представлен собственный алгоритм создания искусственных данных для обучения нейронных сетей распознаванию образов. Был использован большой набор трехмерных фоновых моделей, которые плотно визуализированы в частично рандомизированном режиме для создания фоновых изображений. Это позволило создавать локально реалистичные искажения фона, которые делают обученные модели устойчивыми к изменениям окружающей среды. Поверх этих фоновых изображений были визуализированы трехмерные модели интересующих нас объектов. Во время обучения процесс генерации данных следует стратегии обучения, которая гарантирует, что все объекты переднего плана представлены в сгенерированной выборке в равной степени в случайном порядке во всех возможных позах с возрастающей сложностью распознавания, с учетом добавления случайного освещения, размытия и шума. Разработанный подход не требует сложных композиций сцены, создания сложных фотореалистичных изображений или реальных фоновых изображений для обеспечения необходимого фонового беспорядка и хорошо масштабируется для больших наборов исходных данных.

По результатам исследования была разработана «Интеллектуальная система генерирования изображений, предназначенных для обучения нейросетевых моделей распознавания визуальных образов» [20].

Экспериментально доказаны преимущества разработанной стратегии аргументации по сравнению со случайной генерацией поз. В ходе экспериментов установлено, что сгенерированные изображения в идеале должны состоять только из искусственного контента и что все фоновое изображение должно быть заполнено фоновым беспорядком. Проведенные эксперименты также позволили выделить влияние различных факторов, таких как шум или масштаб объектов, на эффективность распознавания.

Выполнено сравнение обучения распознаванию на реальном, на сгенерированном с помощью предложенного решения, а также на смешанном наборах данных. Наилучшую эффективность распознавания показал смешанный набор данных. Предлагаемый метод генерации позволяет создавать большие объемы синтезированных изображений с автоматической разметкой в существенно меньшие сроки, чем заняло бы создание такого же набора из реальных изображений. При этом реальные данные все так же остаются более предпочтительными для обучения моделей. Однако, поскольку их сбор и обработка в достаточном количестве занимают большое количество времени, комбинация синтетических и реальных данных позволяет повысить эффективность обучения модели распознаванию объектов и получить выигрыш по времени формирования обучающих данных.

Литература

1. **Беляева О. В., Перминов А. И., Козлов И. С.** Использование синтетических данных для тонкой настройки моделей сегментации документов. *Тр. ИСП РАН*, 2020, т. 32, № 4, с. 189–202. doi:10.15514/ISPRAS-2020-32(4)-14
2. **Парасич А. В., Парасич В. А., Парасич И. В.** Формирование обучающей выборки в задачах машинного обучения. Обзор. *Информационно-управляющие системы*, 2021, № 4, с. 61–70. doi:10.31799/1684-8853-2021-4-61-70
3. **Konushin A. S., Faizov B. V., Shakhuro V. I.** Road images augmentation with synthetic traffic signs using neural networks. *Computer Optics*, 2021, no. 5, pp. 736–748. doi:10.18287/2412-6179-CO-859
4. **Пчелинцев С. Ю., Ковалева О. А., Суслин А. А.** Использование синтетических данных для обучения нейронных сетей. *Наука. Технология. Производство — 2021: материалы Всерос. науч.-техн. конф.*, Салават, 19–23 апреля 2021 г., Уфа, 2021, с. 8–10.
5. **Каляшов Е. В., Савельева А. А., Тарлыков А. В.** Сегментация реальных объектов с использованием нейронной сети, обученной на синтетических данных. *Актуальные проблемы инфотелекоммуникаций в науке и образовании: VIII Междунар. науч.-техн. и науч.-метод. конф.; сб. науч. ст. в 4 т.*,

Санкт-Петербург, 27–28 февраля 2019 г., СПб., 2019, с. 472–476.

6. **Tobin J., Fong R., Ray A., Schneider J., Zaremba W., Abbeel P.** Domain randomization for transferring deep neural networks from simulation to the real world. *IEEE/RSJ Intern. Conf. on Intelligent Robots and Systems (IROS)*, Vancouver, 2017, pp. 23–30. doi:10.1109/IROS37595.2017
7. **Prakash A., Boochoon S., Brophy M., Acuna D., Cameracci E., State G., Shapira O., Birchfield S.** Structured domain randomization: Bridging the reality gap by context aware synthetic data. *Intern. Conf. on Robotics and Automation (ICRA)*, Montreal, 2019, pp. 7249–7255. doi:10.1109/ICRA39644.2019
8. **Dwibedi D., Misra I., Hebert M.** Cut, paste and learn: surprisingly easy synthesis for instance detection. *IEEE Intern. Conf. on Computer Vision (ICCV)*, Venice, 2017, pp. 1310–1319. doi:10.1109/ICCV.2017.146
9. **Richter S. R., Vineet V., Roth S., Koltun V.** Playing for data: Ground truth from computer games. *European Conf. on Computer Vision*, Amsterdam, 2016, pp. 102–118. doi:10.1007/978-3-319-46475-6_7
10. **Bousmalis K., Silberman N., Dohan D., Erhan D., Krishnan D.** Playing for data: Unsupervised pixel-level domain adaptation with generative adversarial networks. *Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 2017, pp. 95–104. doi:10.1109/CVPR.2017.18
11. **Chen B.-C., Kae A.** Playing for data: Toward realistic image compositing with adversarial learning. *Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 2019, pp. 8407–8416. doi:10.1109/CVPR.2019.00861
12. **Inoue T., Chaudhury S., De Magistris G., Dasgupta S.** Transfer learning from synthetic to real images using variational autoencoders for precise position detection. *Intern. Conf. on Image Processing (ICIP)*, Athens, 2018, pp. 2725–2729. doi:10.1109/ICIP.2018.8451064
13. **Yao T., Pan Y., Ngo C.-W., Li H., Mei T.** Semi-supervised domain adaptation with subspace learning for visual recognition. *Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015, pp. 2142–2150. doi:10.1109/CVPR31182.2015
14. **Tremblay J., To T., Sundaralingam B., Xiang Y., Fox D., Birchfield S.** Deep object pose estimation for semantic robotic grasping of household objects. *Conf. on Robot Learning (CoRL)*, Zurich, 2018, pp. 306–316.
15. ГОСТ Р 52289-2019. *Технические средства организации дорожного движения. Правила применения дорожных знаков, разметки, светофоров, дорожных ограждений и направляющих устройств*. М., Стандартинформ, 2020. 134 с.
16. **Шахуро В. И., Конушин А. С.** Российская база изображений автодорожных знаков. *Компьютерная оптика*, 2016, т. 40, № 2, с. 294–300. doi:10.18287/2412-6179-2016-40-2-294-300
17. **Ren S., He K., Girshick R., Sun J.** Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 39, no. 6, pp. 1137–1149. doi:10.1109/TPAMI.2016.2577031
18. **Krizhevsky A., Sutskever I., Hinton G.** ImageNet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 2012, no. 25, pp. 1097–1105. doi:10.1145/3065386
19. *ImageNet*. https://www.image-net.org (дата обращения: 05.01.2022).
20. Свид. о рег. прогр. для ЭВМ RU 2021666818. *Интеллектуальная система генерирования изображений, предназначенных для обучения нейросетевых моделей распознавания визуальных образов*, С. Ю. Пчелинцев, О. А. Ковалева, С. В. Ковалев. № 2021666314; заявл. 20.10.21; опубл. 20.10.21, Бюл. № 10.

UDC 004.93
doi:10.31799/1684-8853-2022-3-9-19

Method for creating synthetic data sets for training neural network models for object recognition

S. Y. Pchelintsev[a], Post-Graduate Student, orcid.org/0000-0001-9195-8318, veselyrojer@mail.ru
M. A. Liashkov[a], Post-Graduate Student, orcid.org/0000-0002-7793-7024
O. A. Kovaleva[a,b], Dr. Sc., Tech., Professor, orcid.org/0000-0003-0735-6205
[a]Derzhavin Tambov State University, 33, Internatsionalnaya St., 392000, Tambov, Russian Federation
[b]Tambov State Technical University, 106, Sovetskaya St., 392000, Tambov, Russian Federation

Introduction: The lack of training data leads to low accuracy of visual pattern recognition. One way to solve this problem is to use real data in combination with synthetic data. **Purpose:** To improve the performance of pattern recognition systems in computer vision by mixing real and synthetic data for training, and to reduce the time needed for preparing training data. **Results:** We have built an intelligent information system on the basis of the proposed method which allows the generation of synthetic images. The system allows to generate large and representative samples of images for pattern recognition neural network training. We have also developed software for the synthetic image generator for neural network training. The generator has a modular architecture, which makes it easy to modify, remove or add individual stages to the synthetic image generation pipeline. One can adjust individual parameters (like lighting or blurring) for generated images. The experiment was aimed to compare the accuracy of pattern recognition for a neural network trained on different training samples. The combination of real and synthetic data in model training showed the best recognition performance. Artificially generated training samples, in which the scale of background objects is approximately equal to

the scale of the object of interest, and the number of objects of interest in the frame is higher, turned out to be more efficient than other artificially constructed training samples. Changing focal length of the camera in the synthetic image generation scene had no effect on the recognition performance. **Practical relevance:** The proposed image generation method allows to create a large set of artificially constructed data for training neural networks in pattern recognition in less time than it would take to create the same set of real data.

Keywords — neural networks, artificial intelligence, machine learning, synthetic data sets, image generation.

For citation: Pchelintsev S. Y., Liashkov M. A., Kovaleva O. A. Method for creating synthetic data sets for training neural network models for object recognition. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 3, pp. 9–19 (In Russian). doi:10.31799/1684-8853-2022-3-9-19

References

1. Belyaeva O. V., Perminov A. I., Kozlov I. S. Synthetic data usage for document segmentation models fine-tuning. *Proc. of ISP RAS*, 2020, vol. 32, no. 4, pp. 189–202 (In Russian). doi:10.15514/ISPRAS-2020-32(4)-14
2. Parasich A. V., Parasich V. A., Parasich I. V. Training set formation in machine learning tasks. Survey. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 61–70 (In Russian). doi:10.31799/1684-8853-2021-4-61-70
3. Konushin A. S., Faizov B. V., Shakhuro V. I. Road images augmentation with synthetic traffic signs using neural networks. *Computer Optics*, 2021, no. 5, pp. 736–748. doi:10.18287/2412-6179-CO-859
4. Pchelintsev S. Y., Kovaleva O. A., Suslin A. A. Using synthetic data for training neural networks. *Materialy Vseros. nauch.-tekhn. konf. "Nauka. Tekhnologiya. Proizvodstvo — 2021"* [Proc. Vseros. Sci.-Tech. Conf. "Science. Technology. Production — 2021"]. Ufa, 2021, pp. 8–10 (In Russian).
5. Kalyashov E. V., Savel'eva A. A., Tarlykov A. V. Segmentation of real objects with using neural network trained on real data. *Sbornik statej VIII Mezhdunarodnoj nauchno-tekhnicheskoj i nauchno-metodicheskoj konferencii "Aktual'nye problemy infotelekommunikacij v nauke i obrazovanii"* [Proc. VIII Intern. Scien.-Tech. and Scien.-Method. Conf. "Actual problems of infotelecommunication in science and education"]. Saint-Petersburg, 2019, pp. 472–476 (In Russian).
6. Tobin J., Fong R., Ray A., Schneider J., Zaremba W., Abbeel P. Domain randomization for transferring deep neural networks from simulation to the real world. *IEEE/RSJ Intern. Conf. on Intelligent Robots and Systems (IROS)*, Vancouver, 2017, pp. 23–30. doi:10.1109/IROS37595.2017
7. Prakash A., Boochoon S., Brophy M., Acuna D., Cameracci E., State G., Shapira O., Birchfield S. Structured domain randomization: Bridging the reality gap by context aware synthetic data. *Intern. Conf. on Robotics and Automation (ICRA)*, Montreal, 2019, pp. 7249–7255. doi:10.1109/ICRA39644.2019
8. Dwibedi D., Misra I., Hebert M. Cut, paste and learn: surprisingly easy synthesis for instance detection. *IEEE Intern. Conf. on Computer Vision (ICCV)*, Venice, 2017, pp. 1310–1319. doi:10.1109/ICCV.2017.146
9. Richter S. R., Vineet V., Roth S., Koltun V. Playing for data: Ground truth from computer games. *European Conf. on Computer Vision*, Amsterdam, 2016, pp. 102–118. doi:10.1007/978-3-319-46475-6_7
10. Bousmalis K., Silberman N., Dohan D., Erhan D., Krishnan D. Playing for data: Unsupervised pixel-level domain adaptation with generative adversarial networks. *Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 2017, pp. 95–104. doi:10.1109/CVPR.2017.18
11. Chen B.-C., Kae A. Playing for data: Toward realistic image compositing with adversarial learning. *Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 2019, pp. 8407–8416. doi:10.1109/CVPR.2019.00861
12. Inoue T., Chaudhury S., De Magistris G., Dasgupta S. Transfer learning from synthetic to real images using variational autoencoders for precise position detection. *Intern. Conf. on Image Processing (ICIP)*, Athens, 2018, pp. 2725–2729. doi:10.1109/ICIP.2018.8451064
13. Yao T., Pan Y., Ngo C.-W., Li H., Mei T. Semi-supervised domain adaptation with subspace learning for visual recognition. *Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015, pp. 2142–2150. doi:10.1109/CVPR31182.2015
14. Tremblay J., To T., Sundaralingam B., Xiang Y., Fox D., Birchfield S. Deep object pose estimation for semantic robotic grasping of household objects. *Conf. on Robot Learning (CoRL)*, Zurich, 2018, pp. 306–316.
15. State Standard 52289-2019. *Traffic control devices. Rules of application of traffic signs, markings, traffic lights, guardrails and delineators*. Moscow, Standartinform Publ., 2020. 134 p. (In Russian).
16. Shakhuro V. I., Konushin A. S. Russian traffic sign images dataset. *Computer Optics*, 2016, no. 2, pp. 294–300 (In Russian). doi:10.18287/2412-6179-2016-40-2-294-300
17. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 39, no. 6, pp. 1137–1149. doi:10.1109/TPAMI.2016.2577031
18. Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 2012, no. 25, pp. 1097–1105. doi:10.1145/3065386
19. *ImageNet*. Available at: https://www.image-net.org (accessed 5 January 2022).
20. Pchelintsev S. Y., et al. *Intellektual'naya sistema generirovaniya izobrazhenij, prednaznachennyh dlya obucheniya nejrosetevyh modelej raspoznavaniya vizual'nyh obrazov* [Intellectual system of generation images aimed at training neural network models for visual images recognition]. Computer program registration certificate Russia, no. 2021666818, 2021.