ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 58777— 2019

Воздушный транспорт

АЭРОПОРТЫ. ТЕХНИЧЕСКИЕ СРЕДСТВА ДОСМОТРА

Методика определения показателей качества распознавания незаконных вложений по теневым рентгеновским изображениям

Издание официальное

Москва
Стандартинформ
2020

Предисловие

1 РАЗРАБОТАН Федеральным государственным автономным образовательным учреждением высшего образования «Санкт-Петербургский политехнический университет Петра Великого»

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 164 «Искусственный интеллект»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 20 декабря 2019 г. № 1430-ст

4 ВВЕДЕН ВПЕРВЫЕ

Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячно издаваемых информационных указателях «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (www.gost.ru)

© Стандартинформ, оформление, 2020

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 1
4 Требования к условиям работы системы 2
5 Тестовая выборка 3
6 Сценарные испытания 4
7 Показатели качества 4
8 Верификация экспертных оценок 5
Приложение А (обязательное) Вычисление неопределенностей, необходимых размеров выборок и объема испытаний 7
Приложение Б (справочное) Оценка качества локализации 9
Приложение В (справочное) Интегральные показатели качества обнаружения 10
Приложение Г (справочное) Рекомендуемый список типов опасных и условно-опасных вложений для проведения испытаний 11
Библиография 12

Введение

Интроскоп является одним из основных средств досмотра, необходимых для обеспечения безопасности на территории аэропорта. С использованием интроскопов выявляются незаконные вложения в сумках багажа и ручной клади. Обычно выявление незаконных вложений происходит путем ручного дешифрирования теневых снимков, полученных с использованием интроскопа.

В настоящее время существует множество алгоритмов и систем технического зрения, предназначенных для автоматического обнаружения и распознавания опасных вложений.

Настоящий стандарт определяет методику определения показателей качества распознавания незаконных вложений по теневым рентгеновским изображениям, предназначенную для тестирования таких систем и алгоритмов.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Воздушный транспорт

АЭРОПОРТЫ. ТЕХНИЧЕСКИЕ СРЕДСТВА ДОСМОТРА

Методика определения показателей качества распознавания незаконных вложений по теневым рентгеновским изображениям

Air transport.
Airports. Technical means of inspection. Methodology for determination of quality indicators of recognition of illegal attachments by shadow X-ray images

Дата введения — 2020—09—01

1 Область применения

Настоящий стандарт устанавливает методику определения показателей качества распознавания незаконных вложений по теневым рентгеновским изображениям.

Требования настоящего стандарта распространяются:

- на разработку программ и методик испытаний систем автоматического анализа теневых снимков;
- методы проведения испытаний алгоритмов анализа теневых снимков.

2 Нормативные ссылки

В настоящем стандарте использована нормативная ссылка на следующий стандарт:

ГОСТ Р 55249—2012 Воздушный транспорт. Аэропорты. Технические средства досмотра. Общие технические требования

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины и определения

В настоящем стандарте применены следующие термины с соответствующими определениями:

3.1 **рентгенотелевизионный стационарный интроскоп;** интроскоп:

Устройство для контроля ручной клади, багажа, почты и груза по теневому изображению их внутреннего содержания на экране телемонитора. [ГОСТ Р 55249—2012, пункт 3.1.3]

3.2 **сумка:** Ручная кладь, багаж либо другая емкость, пригодная для провоза либо проноса незаконных вложений, предъявляемые к анализу.

Издание официальное

3.3 **теневое рентгеновское изображение;** изображение: Рентгеновское изображение либо набор изображений, формируемый интроскопом при анализе сумки.

3.4 **незаконное (опасное) вложение; опасные предметы:** Предмет либо вещество, запрещенные к провозу либо проносу на территорию аэропорта.

3.5 **условно-опасное вложение; условно-опасные предметы:** Предмет либо вещество, сами по себе не являющиеся опасными, однако в определенном сочетании предметов внутри сумки являющиеся таковыми.

Примечание — Примером условно-опасных предметов являются поражающие элементы самодельных взрывных устройств (болты, гайки и т. д.), часы, рации и прочие элементы, часто использующиеся при создании самодельных взрывных устройств. Рекомендуемый список типов опасных и условно-опасных вложений для проведения испытаний представлен в приложении Г.

3.6 **опасная сумка:** Сумка, содержащая опасные предметы, либо определенный набор условно-опасных предметов, достаточный для признания содержимого сумки опасным.

3.7 **безопасная сумка:** Сумка, не являющаяся опасной.

3.8 **результат распознавания:** Результат анализа алгоритмом либо прототипом сумки, содержащий список опасных предметов, находящихся в сумке.

3.9 **технологическое испытание:** Испытание одного или более алгоритмов распознавания опасных вложений с использованием существовавшей ранее или специально собранной базы данных образцов.

3.10 **натурное (сценарное) испытание:** Испытание, при проведении которого эксплуатационные характеристики системы определяются с помощью прототипа системы технического зрения.

3.11 **тестовая выборка:** Набор теневых рентгеновских изображений, формируемый для проведения тестирования качества работы алгоритмов анализа теневых рентгеновских изображений.

3.12 **верификационная выборка:** Набор теневых рентгеновских изображений, передаваемых разработчикам алгоритма для задания требований к условиям работы алгоритма.

3.13 **СК снимка:** Координатная система, заданная на плоскости снимка. Две ее оси параллельны сторонам снимка.

Примечание — Рекомендуемое направление осей: ось OX направлена слева направо, OY – снизу вверх. Начало координат соответствует нулевому пикселю изображения.

3.14 **описывающий прямоугольник объекта:** Минимальный по площади прямоугольник, полностью содержащий изображение объекта.

Примечание — Координаты минимального описывающего прямоугольника задаются в СК снимка.

3.15 **результат обнаружения:** Результат анализа алгоритмом либо прототипом сумки, содержащий набор описывающих прямоугольников, найденных алгоритмом обнаружения объектов, а также соответствующие им метки класса опасного предмета.

3.16 **сигнал тревоги:** Сигнал, формируемый алгоритмом либо прототипом, указывающий наличие в исследуемой сумке опасных предметов.

Примечание — на практике данный сигнал является сигналом к досмотру сумки представителями правоохранительных органов.

4 Требования к условиям работы системы

4.1 Перед проведением тестирования должны быть четко сформулированы требования к условиям работы алгоритма либо прототипа. Требования к условиям работы (входным данным) могут формулироваться:

- набором примеров входных данных (валидационной выборкой);
- границами однозначно вычислимых численных параметров;
- качественными словесными описаниями, достаточно подробными для однозначной оценки экспертами.

4.2 При задании требований путем предоставления валидационной выборки должны быть выполнены следующие условия:

- для выборки должен быть составлен список условий (требований), задающих условия работы алгоритма;

- данные, входящие в выборку, должны быть размечены метками, соответствующими тем или иным условиям;

- для требований, задаваемых в виде диапазонов, рекомендуется определить интервал (шкалу), интерпретируемый экспертами, а также предоставить достаточное число примеров для различных значений введенной шкалы для однозначной оценки текущего значения. Требования в таком случае задаются для некоторых уровней введенной шкалы.

4.3 При задании требований границами однозначно вычислимых численных параметров (характеристик) необходимо предоставить методики вычислений данных характеристик (однозначные и вычислимые).

4.4 При задании требований качественным описанием оно должно содержать достаточно информации для воспроизведения условий при проведении технологического испытания.

4.5 При задании требований качественным описанием оно должно содержать достаточно информации для однозначного ответа эксперта (либо группы экспертов) по каждому элементу тестовой выборки.

4.6 При проведении технологических испытаний требования к условиям работы, задаваемые путем предоставления валидационной выборки, проверяются путем экспертной оценки.

5 Тестовая выборка

5.1 Тестовая выборка формируется путем съемки с использованием интроскопа (получения изображений) сумок, содержащих и не содержащих опасные вложения.

5.2 Модель интроскопа, используемого для формирования тестовой выборки, должна быть отражена в протоколах испытания/тестирования.

5.3 Объективность тестовой выборки. Тестовая выборка может считаться в той или иной степени необъективной, если:

- разработчик алгоритма/прототипа (далее — разработчик) имел доступ к данным;

- разработчик предоставил оборудование, использовавшееся при создании или обработке базы данных, особенно если подобная деятельность повлияла на характер или качество базы данных (например, удаление некоторых образцов);

- испытуемый алгоритм ранее был проверен и настроен с использованием данной базы данных.

Примечание — Данные не должны использоваться в испытании, если один или несколько разработчиков ранее имели к ним доступ. Предыдущее испытание или настройка системы с использованием выборки (всей или ее фрагмента) должны быть отражены в протоколе испытания.

5.4 Данные не должны повторно использоваться для тестирования, если тестируемый алгоритм был настроен на основе эксплуатационных характеристик, полученных в ходе предыдущего тестирования по этим данным.

5.5 При задании требований к условиям работы путем предоставления валидационной выборки тестовая выборка должна формироваться из той же генеральной совокупности.

5.6 При подтверждении характеристик работы в различных условиях тестовая выборка должна содержать для каждого изображения метаданные, содержащие информацию об условиях съемки.

5.7 Помимо изображений тестовая выборка должна содержать файлы разметки. Файлы разметки должны содержать следующую дополнительную информацию.

5.7.1 Для тестирования качества распознавания — список опасных предметов, содержащихся в сумке.

5.7.2 Для тестирования качества обнаружения — список опасных предметов, а также координаты описывающих прямоугольников, им соответствующих.

5.7.3 При наличии в выборке условно-опасных объектов каждая сумка также должна быть размечена как опасная либо безопасная.

5.8 В случае если выборка создается с целью подтверждения фиксированных вероятностных характеристик, ее размер должен быть достаточен для их подтверждения с некоторой точностью (для некоторого доверительного интервала) и доверительной вероятностью.

Примечания

1 Доверительная вероятность, а также точность оценки зависят от целей тестирования. Рекомендации по выбору доверительной вероятности и доверительного интервала описаны в приложении А.

2 Пример расчета размеров выборки приведен в приложении А.

6 Сценарные испытания

6.1 Модель интроскопа, используемого для сценарных испытаний, должна быть отражена в протоколах испытания/тестирования.

6.2 Для проведения технологических испытаний должен быть сформирован набор сумок, содержащих опасные предметы и не содержащих их. Сумки, содержащие опасные предметы, могут содержать один или несколько опасных предметов.

6.3 Для формирования набора сумок допускается использование имитаторов опасных предметов при условии, что они неотличимы от оригинальных опасных предметов.

6.4 Объективность испытаний. Испытания могут считаться в той или иной степени необъективными, если:

- разработчик алгоритма/прототипа (далее — разработчик) имел доступ к набору сумок;
- разработчик предоставил часть сумок либо имитаторов опасных предметов;
- испытуемый алгоритм ранее был проверен и настроен с использованием используемого при испытаниях набора сумок.

6.5 Данные не должны повторно использоваться для тестирования, если тестируемый алгоритм был настроен на основе эксплуатационных характеристик, полученных в ходе предыдущего тестирования по этим данным.

6.6 Перед проведением технологических испытаний должны быть согласованы порядок их проведения и методика использования прототипа. Определено понятие теста.

Примечание — Например, количество сканирований одной сумки, а также логика объединения результатов распознавания и обнаружения для нескольких прогонов одной сумки.

6.7 Сценарные испытания разделяются на тесты. Один тест подразумевает проверку одной сумки. Тесты, проводимые в рамках сценарных испытаний, делятся на два типа.

6.7.1 Тесты, эмулирующие проверку опасной сумки (далее — тест на тревогу).

6.7.2 Тесты, эмулирующие проверку безопасной сумки (далее — тест на ложное срабатывание).

6.8 В случае если испытания проводятся с целью подтверждения фиксированных вероятностных характеристик, число тестов должно быть достаточным для их подтверждения с некоторой точностью (для некоторого доверительного интервала) и доверительной вероятностью.

Примечания

1 Доверительная вероятность, а также точность оценки зависят от целей тестирования. Рекомендации по выбору доверительной вероятности и доверительного интервала описаны в приложении А.

2 Пример числа тестов приведен в приложении А.

7 Показатели качества

7.1 Показатели качества, рассчитываемые в процессе проведения технологического испытания

7.1.1 Показатели качества распознавания:

- вероятность правильного формирования сигнала тревоги (ВПФСТ) — отношение числа опасных сумок, верно идентифицированных алгоритмом/прототипом к общему числу опасных сумок;
- вероятность формирования сигнала ложной тревоги (ВФСЛТ) — отношение числа безопасных сумок, идентифицированных алгоритмом/прототипом как опасные, к общему числу безопасных сумок;
- вероятность правильного распознавания опасного предмета (ВПРОП) — отношение числа верно распознанных опасных предметов к общему числу опасных предметов в выборке;
- вероятность ложного распознавания опасного предмета (ВЛРОП) — отношение числа неверно распознанных опасных предметов к общему числу распознанных алгоритмом опасных предметов в выборке.

Примечания

1 В данном случае под сумкой подразумевается запись в тестовой выборке, соответствующая одной сумке.

2 Показатели ВПФСТ и ВФСЛТ рассчитываются только парно, например ВПФСТ при ВФСЛТ.

3 Для расчета показателя ВФСЛТ выборка должна содержать набор изображений, соответствующих безопасным сумкам.

4 Показатели ВПРОП/ВЛРОП должны быть рассчитаны для каждого опасного и условно-опасного класса предметов раздельно.

5 Показатели качества распознавания не учитывают точность локализации алгоритмом опасного предмета.

7.1.2 Показатели качества обнаружения

При тестировании качества обнаружения должны быть рассчитаны базовые показатели качества обнаружения:

- вероятность правильного обнаружения опасного предмета (ВПООП) — отношение числа верно обнаруженных опасных предметов к общему числу опасных предметов в выборке;
- вероятность ложного обнаружения опасного предмета (ВЛООП) — отношение числа неверно распознанных опасных предметов к общему числу обнаруженных алгоритмом опасных предметов в выборке.

Примечания

1 Данные показатели рекомендуется использовать в том числе для задания требований к работе алгоритмов/прототипа. Требования должны задаваться только в виде пары ВПООП при ВЛООП.

2 Для корректного расчета показателей ВПООП и ВЛООП необходимо определить методику расчета точности локализации и порог, задающий минимальную точность локализации. Рекомендации по методике расчета точности локализации и используемого порога даны в приложении Б.

7.1.3 Для ранжирования нескольких алгоритмов по показателям из п. 7.2.1 рекомендуется использовать Ф-меру, вычисляемую по формуле

$$\text{Ф-мера} = 2*(Б^2 + 1)*ВПООП*(1 - ВЛООП)/(Б^2*(1 - ВЛООП) + ВПООП)$$

Примечания

1 Коэффициент Б > 0 задает относительную значимость показателей ВПООП и ВЛООП. Если Б = 1, то ВПООП и ВЛООП имеют равную значимость. Если Б < 0, то предпочтение отдается ВЛООП, в противном случае соответственно ВПООП.

2 Показатель Ф может принимать значения в диапазоне от 0 до 1, при этом значение 1 соответствует наилучшему качеству работы.

3 При отсутствии требований по локализации и/или качеству работы для сравнительной оценки могут использоваться интегральные характеристики, описанные в приложении В.

7.2 Показатели качества, рассчитываемые в процессе проведения сценарного испытания

7.2.1 Показатели качества распознавания:

- вероятность правильного формирования сигнала тревоги (ВПФСТ) — отношение числа верно подаваемых сигналов тревоги к общему числу тестов на тревогу;
- вероятность формирования сигнала ложной тревоги (ВФСЛТ) — отношение числа подаваемых ложных сигналов тревоги к общему количеству тестов на ложные срабатывания;
- вероятность правильного распознавания опасного предмета (ВПРОП) — отношение числа верно распознанных опасных предметов к общему числу опасных предметов, использованных в тестах на тревогу;
- вероятность ложного распознавания опасного предмета (ВЛРОП) — отношение числа неверно распознанных опасных предметов к общему числу опасных предметов, зафиксированных в тестах на тревогу.

Примечания

1 Показатели ВПФСТ и ВФСЛТ рассчитываются только парно, например ВПФСТ при ВФСЛТ.

2 Показатели ВПРОП/ВЛРОП должны быть рассчитаны для каждого опасного и условно-опасного класса предметов раздельно.

3 Показатели качества распознавания не учитывают точность локализации алгоритмом опасного предмета.

4 Показатель ВЛРОП можно рассчитывать по всем тестам, однако в таком случае отношение количества тестов необходимо дополнительно учитывать в формуле для расчета ВЛРОП.

7.3 Для показателей качества по 7.1.1, 7.1.2, 7.2.1 должна быть вычислена неопределенность результатов испытаний, а также определены принципы и формулы для вычисления неопределенности.

Примечание — Пример выражений для расчета неопределенности представлен в приложении А.

8 Верификация экспертных оценок

8.1 При возникновении спорных ситуаций касательно условий проведения сценарных испытаний и изображений тестовой выборки спорные ситуации разрешаются путем привлечения группы экспертов.

8.2 Группа экспертов должна состоять из не менее трех человек, имеющих достаточную квалификацию в области анализа теневых рентгеновских снимков.

8.3 Группа экспертов принимает решение по каждому изображению тестовой выборки, вызывающему разногласия.

8.4 При анализе тестовой выборки решение принимается путем голосования. В случае если консенсус не достигнут, спорное изображение удаляется из тестовой выборки и заменяется другим.

8.5 При анализе условий сценарных испытаний решение также принимается путем голосования по каждому спорному тесту, проведенному в рамках сценарных испытаний. В случае если консенсус не достигнут, тест считается спорным и требует повторного проведения.

Приложение А
(обязательное)

Вычисление неопределенностей, необходимых размеров выборок и объема испытаний

А.1 Приведем некоторые известные факты из теории вероятностей, которые будут использованы далее для оценок.

Пусть $\varepsilon_1 ... \varepsilon_n$ — независимые ограниченные случайные величины:

$$a_i \le \varepsilon_i \le b_i,$$

и $E\varepsilon_1 ... E\varepsilon_n$ — их математические ожидания.

Тогда вероятность отклонения среднего арифметического

$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i$$

от его математического ожидания

$$E\varepsilon = \frac{1}{N}\sum_{i=1}^{N}E\varepsilon_i$$

на величину более ε может быть оценена с использованием неравенства Хефдинга [1]

$$P\left(\left|\varepsilon - E\varepsilon\right| > \varepsilon\right) \le 2\exp\left(-\frac{2N^2\varepsilon^2}{\sum_{i=1}^{N}(b_i - a_i)^2}\right).$$

При этом такая оценка не зависит от распределений $\varepsilon_1 ... \varepsilon_n$. Частным случаем описанного выше неравенства является оценка отклонения частоты $\vartheta_N(A) \in [0, 1]$ некоторого события A в серии N независимых испытаний от его вероятности $P(A) = E\vartheta_N(A)$

$$P\left(\left|\vartheta_N(A) - P(A)\right| > \varepsilon\right) \le 2\exp\left(-2N\varepsilon^2\right). \qquad (A.1)$$

Формула (А.1) позволяет для точности ε и числа испытаний N получить оценку доверительной вероятности P, и наоборот, для заданной доверительной вероятности и точности — оценку числа испытаний N. Числа испытаний для различных значений P и ε представлены в таблице А.1.

С использованием таблицы А.1 и формулы (А.1) можно вычислить неопределенности и оценить размеры тестовых выборок для показателей качества, установленных в стандарте. Точность оценки задается прежде всего требованиями к исследованию, она должна быть не менее

$$\varepsilon < 1{,}0 - P^*,$$

где P^*— подтверждаемая вероятность.

Таблица А.1 — Таблица минимального количества испытаний в зависимости от доверительной вероятности и точности

Доверительная вероятность	Точность						
	ε = 0,1	ε = 0,05	ε = 0,02	ε = 0,01	ε = 0.005	ε = 0,001	ε = 0,0001
P = 0,75	104	416	2600	10.398	41.589	1.039.721	1,04E+08
P = 0,9	150	600	3745	14.979	59.915	1.497.866	1,5E+08
P = 0,91	156	621	3877	15.506	62.022	1.550.546	1,55E+08
P = 0,92	161	644	4024	16.095	64.378	1.609.438	1,61E+08
P = 0,93	168	671	4191	16.763	67.049	1.676.204	1,68E+08
P = 0,94	176	702	4384	17.533	70.132	1.753.279	1,75E+08
P = 0,95	185	738	4612	18.445	73.778	1.844.440	1,84E+08
P = 0,96	196	783	4891	19.561	78.241	1.956.012	1,96E + 08
P = 0,97	210	840	5250	20.999	83.995	2.099.853	2,1E + 08
P = 0,98	231	922	5751	23.026	92.104	2.302.585	2,3E + 08

Окончание таблицы А.1

Доверительная вероятность	Точность						
	ε = 0,1	ε = 0.05	ε = 0.02	ε = 0,01	ε = 0,005	ε = 0,001	ε = 0,0001
P = 0,99	265	1060	6623	26.492	105.967	2.649.159	2,65E + 08
P = 0,999	381	1521	9502	38.005	152.018	3.800.451	3,8E + 08
P = 0,9999	496	1981	12.380	49.518	198.070	4.951.744	4,95E + 08

Доверительная вероятность также выбирается в зависимости от требований к исследованиям. Однако оценки, полученные с доверительными вероятностями менее 0,9, обладают недостаточной достоверностью для практического применения.

А.2 Вычисление размеров тестовой выборки либо числа тестов при тестировании качества распознавания опасных предметов

При тестировании алгоритмов распознавания опасных предметов вычисляют следующие показатели:

- вероятность правильного формирования сигнала тревоги (ВПФСТ) — отношение числа опасных сумок, верно идентифицированных алгоритмом/прототипом к общему числу опасных сумок;
- вероятность формирования сигнала ложной тревоги (ВФСЛТ) — отношение числа безопасных сумок, идентифицированных алгоритмом/прототипом как опасные, к общему числу безопасных сумок;
- вероятность правильного распознавания опасного предмета (ВПРОП) — отношение числа верно распознанных опасных предметов к общему числу опасных предметов в выборке;
- вероятность ложного распознавания опасного предмета (ВЛРОП) — отношение числа неверно распознанных опасных предметов к общему числу распознанных алгоритмом опасных предметов в выборке.

Рассмотрим пример расчета размеров тестовой выборки для параметров $\varepsilon = 0{,}02$ и $P = 0{,}9$, предъявляемых к ВПФСТ/ВФСЛТ, и $\varepsilon = 0{,}01$ и $P = 0{,}92$ — для ВПРОП.

ВПФСТ для рассматриваемых параметров по таблице А.1 $N = 3745$. Это означает, что количество изображений опасных сумок должно быть не менее 3745.

ВФСЛТ для рассматриваемых параметров по таблице А.1 $N = 3745$. Это означает, что количество изображений безопасных сумок должно быть не менее 3745.

ВПРОП для рассматриваемых параметров, таблице А.1 $N = 16.095$. Это означает, что число опасных вложений в сумки должно быть не менее 16.095. При этом если такие требования предъявляются к ВПРОП по каждому классу опасных предметов, то необходимо по N опасных предметов каждого класса.

Таким образом, тестовая выборка должна содержать не менее 3745 опасных сумок и не менее 3745 безопасных сумок. Кроме того, опасные сумки должны содержать не менее 16.095 вложений. Аналогично проводится вычисление минимального количества тестов для проведения натурного эксперимента.

ВЛРОП. Так как общее число распознанных алгоритмом опасных предметов в выборке зависит от характеристик алгоритма, то доверительная вероятность и точность для него не задаются однозначно размерами выборки. Однако с использованием таблицы А.1 либо формулы (А.1) после испытания для полученного числа распознанных опасных предметов могут быть получены значения ε и P, т. е. оценки неопределенности.

А.3 Вычисление размеров тестовой выборки при тестировании качества распознавания опасных предметов

При тестировании качества обнаружения должны быть рассчитаны базовые показатели качества обнаружения:

- вероятность правильного обнаружения опасного предмета (ВПООП) — отношение числа верно обнаруженных опасных предметов к общему числу опасных предметов в выборке;
- вероятность ложного обнаружения опасного предмета (ВЛООП) — отношение числа неверно распознанных опасных предметов к общему числу обнаруженных алгоритмом опасных предметов в выборке.

Рассмотрим пример расчета размеров тестовой выборки для параметров $\varepsilon = 0{,}02$ и $P = 0{,}9$, предъявляемых к ВПООП. Для рассматриваемых параметров по таблице А.1 $N = 3745$ это означает, что количество опасных предметов должно быть не менее 3745. При этом ВПООП определяет напрямую не количество изображений, однако все же позволяет оценить ее размер для некоторого разумного предположения о количестве опасных предметов в сумке (рекомендуется использовать 1—2 предмета в каждой опасной сумке).

ВЛООП. Так как общее число обнаруженных алгоритмом опасных предметов в выборке зависит от характеристик алгоритма, то доверительная вероятность и точность для него не задаются однозначно размерами выборки. Однако с использованием таблицы А.1 либо формулы (А.1) после испытания для полученного числа обнаруженных опасных предметов могут быть получены значения ε и P, т. е. оценки неопределенности.

Приложение Б
(справочное)

Оценка качества локализации

Задача обнаружения объектов является одной из классических задач компьютерного зрения. При этом для определения факта обнаружения объекта необходима оценка качества локализации. Наиболее популярной метрикой качества локализации, используемой практически во всех международных тестах и конкурсах (например, [2], [3], [4], [5]), является метрика «пересечение над объединением» — Пк0 (Intersection over Union)

$$\text{Пк0} = \frac{\text{П} \cap \text{П}^*}{\text{П} \cup \text{П}^*},$$

где П* — эталонная (истинная) область, занимаемая объектом на изображении (эталон).

П — результат работы алгоритма (обнаружение).

Для задачи обнаружения опасных объектов П и П* — это описывающие прямоугольники (прямоугольные области). При этом задается некоторый порог t [0 … 1]. Данный порог задает минимальную необходимую точность. Если показатель Пк0 больше данного порога, то объект считается обнаруженным. В случае если несколько обнаружений удовлетворяют условию обнаружения, в качестве правильного принимается обнаружение с максимальным Пк0, остальные считаются ложными.

Приложение В
(справочное)

Интегральные показатели качества обнаружения

В случае если не предъявляется конкретных требований к вероятностям правильного и ложного обнаружения, но при этом задаются требования по точности локализации, для сравнения алгоритмов рекомендуется использовать т. н. среднюю точность [4]. Данный показатель позволяет оценить качество работы алгоритма в среднем. Для этого тестируемый алгоритм должен каждому факту обнаружения (гипотезе) приписывать т. н. коэффициент уверенности (score) ∈ [0, 1]. При этом предполагается, что для включения гипотезы в ответы алгоритма (найденные объекты) используется простейшее пороговое правило. Таким образом, регулируя порог, мы можем гибко менять характеристики алгоритма.

Введем понятия точности и полноты:

$$\text{Точность}\,(Precesion)(t) = \frac{N^G(t)}{N^D(t)};$$

$$\text{Полнота}\,(Recall)(t) = \frac{N^G(t)}{N},$$

где $N^G(t)$ — число верных гипотез с *probability* > *t*;

$N^D(t)$ — число гипотез с *probability* > *t*;

N — число объектов в тестовой выборке.

Для каждого значения *t* от 0 до 1 существует пара значений «точность — полнота». График, задаваемый данными величинами, называется графиком «точность — полнота» или PR Curve. Данный график характеризует качество обнаружения алгоритма. Для сравнительной оценки алгоритмов используется площадь под кривой, т. н. показатель *AUC*:

$$AUC = \int_0^1 \text{Точность}(t)\text{Полнота}'(t)\mathrm{d}t.$$

Часто вместо *AUC* используется так называемая средняя точность

$$AP = \frac{1}{11}\sum_{i=0,0.1,0.2\ldots1.0} \text{Точность}_i,$$

где Точность_i — значение точности советующей значению полноты *i*.

Следует отметить, что на практике пары «точность — полнота» заданы не для всех занчений порогов *t*. Это связано прежде всего с конечностью тестовых выборок. Поэтому на практике для вычисления *AUC* и *AP* используется интерполяция между существующими парами значений.

В случае если требования по точности локализации также не задаются, рекомендуется использовать показатель «усредненная средняя точность» (*mAP* [3]):

$$mAP = \frac{1}{10}\sum_{i=0.5,0.505,0.51\ldots0.95} AP_i,$$

где AP_i — показатель *AP*, рассчитанный для минимальной точности локализации *i*.

Следует отметить, что использованный выше диапазон минимальной точности локализации рекомендуется использовать только при использовании метрики «пересечение над объединением» (приложение Б), для других метрик необходимо использовать другие диапазоны.

Приложение Г
(справочное)

Рекомендуемый список типов опасных и условно-опасных вложений для проведения испытаний

Для проведения испытаний рекомендуется использовать следующие типы опасных и условно-опасных вложений:

1 имитаторы оружия:
- нож;
- пистолет;
- автомат;

2 имитаторы патронов к оружию:
- снаряженная обойма пистолета;
- снаряженный магазин автомата;

3 имитаторы боеприпасов:
- граната;
- мина;

4 имитаторы элементов самодельных взрывных устройств;

5 заряды взрывных веществ:
- шашка;
- брикет;
- гранулит;
- лист;

6 средства инициирования:
- электродетонатор;
- запал;
- самодельный детонатор;

7 устройства приведения в действие:
- механическое;
- электрическое;

8 готовые поражающие элементы:
- шарики;
- ролики;
- гайки и т. д.

Библиография

[1] Robert J. Serfling. Probability Inequalities for the Sum in Sampling without Replacement // The Annals of Statistics. — 1974. — Т. 2, № 1. — С. 39—48.

[2] Yang, Shuo and Luo, Ping and Loy, Chen Change and Tang, Xiaoou. WIDER FACE: A Face Detection Benchmark. In CVPR, 2016

[3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick. Microsoft COCO: Common objects in context. In ECCV, 2014.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, The PASCAL visual object classes (VOC) challenge. IJCV, Jun 2010, vol. 88, no. 2, pp. 303—338

[5] I. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-ElHaija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, S. Kamali, M. Malloci, J. Pont-Tuset, A. Veit, S. Belongie, V. Gomes, A. Gupta, C. Sun, G. Chechik, D. Cai, Z. Feng, D. Narayanan, and K. Murphy. Openimages: A public dataset for large-scale multi-label and multi-class image classification. — 2017

УДК 656.7.08 ОКС 03.220.50

Ключевые слова: воздушный транспорт, технические средства, досмотр, качество распознавания, аэропорты, теневые рентгеновские изображения

БЗ 1—2020/28

Редактор *Г.Н. Симонова*
Технический редактор *И.Е. Черепкова*
Корректор *Л.С. Лысенко*
Компьютерная верстка *Е.О. Асташина*

Сдано в набор 23.12.2019. Подписано в печать 22.01.2020 Формат 60×84 1/8. Гарнитура Ариал.
Усл. печ. л. 1,86. Уч.-изд. л. 1,68.
Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении во ФГУП «СТАНДАРТИНФОРМ» для комплектования Федерального информационного фонда стандартов, 117418 Москва, Нахимовский пр-т, д. 31, к. 2.
www.gostinfo.ru info@gostinfo.ru