

Распознавание тест-объектов на тепловизионных изображениях

А.В. Мингалев¹, А.В. Белов¹, И.М. Габдуллин¹, Р.Р. Агафонова¹, С.Н. Шушарин¹

¹АО «Научно-производственное объединение «Государственный институт прикладной оптики», Казань, Россия

Аннотация

Представлен сравнительный анализ нескольких способов распознавания тест-объектов на тепловизионном изображении при настройке и проверке характеристик тепловизионных каналов в автоматизированном режиме. Рассмотрены способы распознавания изображений на основе корреляционного сопоставления изображений, на основе метода Виолы–Джонса, на основе классифицирующей сверточной нейронной сети LeNet, на основе классифицирующей сверточной нейронной сети GoogleNet (Inception v. 1), на основе детектирующей сверточной нейронной сети глубокого обучения типа Single Shot Multibox Detector (SSD) VGG16. Самое высокое значение функционала качества получено с использованием детектирующей сверточной нейронной сети глубокого обучения типа SSD VGG16. К основным достоинствам данного способа следует отнести инвариантность к изменению размеров тест-объектов, высокие значения таких параметров, как точность и полнота, а также отсутствие необходимости применения дополнительных методов для локализации областей интереса.

Ключевые слова: классификация изображений, детектирование объектов на изображениях, распознавание изображений, сверточные нейронные сети глубокого обучения, тепловизионное изображение, тепловизионный прибор.

Цитирование: Мингалев, А.В. Распознавание тест-объектов на тепловизионных изображениях / А.В. Мингалев, А.В. Белов, И.М. Габдуллин, Р.Р. Агафонова, С.Н. Шушарин // Компьютерная оптика. – 2019. – Т. 43, № 3. – С. 402-411. – DOI: 10.18287/2412-6179-2019-43-3-402-411.

Введение

В задачах автоматизации технологических процессов настройки и проверки характеристик тепловизионных приборов одним из основных является вопрос обнаружения и распознавания тест-объектов на тепловизионном изображении без участия оператора.

При выполнении работ по настройке тепловизионных приборов и проведении проверок характеристик тепловизионных приборов в автоматизированном режиме, когда без участия оператора происходит распознавание тест-объектов, вычисление координат положения тест-объекта на тепловизионном изображении с субпиксельной точностью, вычисление значений требуемого параметра или характеристики тепловизионного прибора, оператор контролирует корректность результатов по полученным отчетам. От точности автоматического распознавания и получения координат положения тест-объектов на тепловизионном изображении зависит качество настройки, достоверность измеряемых параметров и характеристик тепловизионных приборов.

Процесс выполнения проверок характеристик тепловизионных каналов в автоматизированном режиме можно разделить на два основных этапа: распознавание тест-объекта на тепловизионном изображении и измерение требуемой характеристики.

В свою очередь, распознавание тест-объектов на тепловизионном изображении включает этап распознавания и вычисления координат положения тест-объекта с погрешностью в несколько пикселей и этап вычисления координат тест-объекта с субпиксельной точностью. Это обусловлено тем, что известные способы определения координат объектов с субпиксель-

ной точностью, оперируя дифференцированием значений яркости изображения и операциями преобразований Фурье, требуют локализации обрабатываемой области на изображении.

В данной статье представлен сравнительный анализ некоторых способов распознавания и вычисления координат положения тест-объектов на тепловизионном изображении, применяемых на первом этапе распознавания тест-объектов с погрешностью в несколько пикселей: способ на основе корреляционного сопоставления изображений [1], способ на основе метода Виолы–Джонса [2, 3], способ на основе классифицирующей сверточной нейронной сети LeNet [4, 5], способ на основе классифицирующей сверточной нейронной сети GoogleNet (Inception v. 1) [6], способ на основе детектирующей сверточной нейронной сети глубокого обучения типа Single Shot Multibox Detector (SSD) VGG16 [7].

1. Методика проведения экспериментов и расчет функционала качества

В качестве тестовых изображений сформирован основной набор данных из 400 кадров с изображениями тест-объектов типа «четырёхштриховая мира» различных угловых размеров. Часть кадров в основном наборе данных получена из видеопотоков, формируемых тепловизионными приборами второго поколения, серийно выпускаемыми АО «НПО ГИПО», часть кадров получена из имеющихся изображений путем выполнения аугментации в виде изменения яркости и контраста, а также внесения в имеющиеся изображения искусственного шума с различными распределениями.

В каждом из 400 кадров основного набора данных с участием человека-оператора произведена разметка изображений тест-объектов (рис. 1). В процессе тестирования ответы алгоритмов представленных способов распознавания сравниваются с данной разметкой.

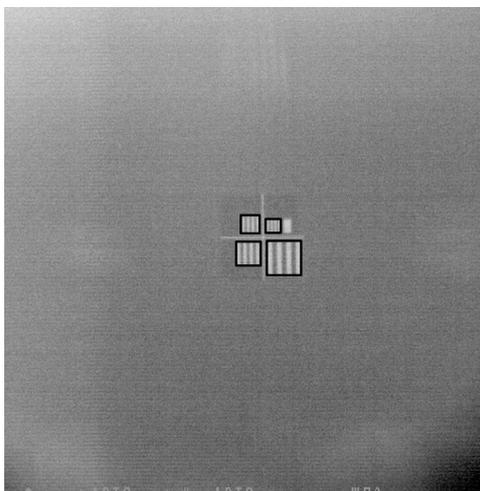


Рис. 1. Пример разметки изображений тест-объектов в тепловизионном кадре, выполненной оператором

Кроме того, в ходе выполнения сравнительного анализа для тестирования качества распознавания нейронных сетей также сформирована дополнительная выборка, состоящая из 200 кадров, в которой присутствуют тест-объекты с большими размерами по сравнению с основным набором данных, а именно с размерами порядка 100×100 пикселей.

В данной работе по дополнительной выборке не выполнялось тестирование способа на основе корреляционного сопоставления изображений при допущении, что качество распознавания данного способа при корректном подборе шаблона не зависит от геометрических размеров тест-объектов. На дополнительной выборке способ на основе метода Виолы–Джонса проверялся в составе с классифицирующими сверточными нейронными сетями LeNet и GoogleNet, качество распознавания данного способа при этом было сопоставимо с качеством, полученным данным способом на основном наборе данных, состоящем из 400 кадров.

Качество распознавания тест-объектов с применением представленных способов оценивается по значению площади F под кривой precision/recall [8], вычисляемой по формуле (1):

$$\gamma(x, y) = \frac{\sum_s \sum_t [w(s, t) - \bar{w}] \times \sum_s \sum_t [f(x + s, y + t) - \bar{f}(x + s, y + t)]}{\left\{ \sum_s \sum_t [w(s, t) - \bar{w}]^2 \times \sum_s \sum_t [f(x + s, y + t) - \bar{f}(x + s, y + t)]^2 \right\}^{\frac{1}{2}}}, \tag{4}$$

где $x = 0, 1, 2, \dots, M-1$; $y = 0, 1, 2, \dots, N-1$; M – ширина изображения f ; N – высота изображения f ; t – ширина шаблонного изображения w ; s – высота шаблонного изображения w ; \bar{w} – среднее значение элементов шаблонного изображения, вычисляемое один раз; \bar{f} – среднее значение элементов изображения f в области,

$$F = \frac{1}{2} \sum_{k=1}^L [P(k-1) + P(k)] \Delta R(k), \tag{1}$$

где P (precision) – точность распознавания, вычисляемая по формуле (2); R (recall) – полнота распознавания, вычисляемая по формуле (3); L – количество всех ответов алгоритма; k – номер текущего ответа алгоритма; $\Delta R(k) = R(k) - R(k-1)$; $R(0) = 0$; $P(0) = 0$.

$$P(k) = \frac{\sum z_{1:k}}{k}; \tag{2}$$

$$R(k) = \frac{\sum z_{1:k}}{n}, \tag{3}$$

где $\sum z_{1:k}$ – сумма первых k элементов вектора z , являющегося бинарным вектором, элемент z_j которого является индикатором того, что ответу алгоритма с индексом j соответствует тест-объект, выделенный человеком-оператором, n – количество всех тест-объектов, выделенных человеком-оператором.

Для тестирования и настройки рассматриваемых способов распознавания выбран персональный компьютер с процессором Intel Core i7, с объемом оперативной памяти 16 Гб, жестким диском SSD-типа объемом 240 Гб и графическим ускорителем NVidia Quadro K5200.

2. Способ, основанный на корреляционном сопоставлении изображений

Корреляционное сопоставление изображений, часто называемое поиском по шаблону, позволяет количественно определить степень схожести отдельных участков изображения с шаблонным изображением данного объекта.

В общем случае принцип работы данного способа заключается в поэлементном (попиксельном) сравнении шаблонного изображения с некоторой областью изображения, совпадающей по размерам с шаблонным изображением. Шаблонное изображение перемещают по изображению, изменяя координаты сравниваемой области изображения, вычисляют коэффициент корреляции по формуле (4) [1], изменяют координаты сравниваемой области изображения, снова вычисляют коэффициент корреляции и так далее до тех пор, пока не будет вычислен коэффициент корреляции для каждого элемента (пикселя) изображения, на котором требуется найти объект.

совпадающей с текущим положением w , суммирование ведется по всем парам координат, общим для f и w .

Таким образом, формируется корреляционная карта, равная по размерам изображению, на котором требуется найти объект. Значения элементов данной карты равны значениям коэффициента корреляции $\gamma(x, y)$, вычисленным по формуле (4) в диапазоне зна-

чений $[-1; 1]$. Чем ближе значение элемента $\gamma(x, y)$ к единице, тем ближе схожесть области изображения, на котором требуется найти объект, с шаблонным изображением.

Проведены эксперименты по определению качества распознавания тест-объектов с использованием шаблонных изображений, соответствующих размерам тест-объектов из набора кадров (рис. 2а–в), и подобрано значение коэффициента корреляции:

$$\gamma(x, y) = 0,55, \quad (5)$$

при котором значение площади F под кривой precision/recall, вычисленное по формуле (1), принимает максимальное значение:

$$F_{\text{шаблон}} = 0,733219. \quad (6)$$

При этом значения точности P и полноты R получились равными:

$$P_{\text{шаблон}} = 0,963843; \quad (7)$$

$$R_{\text{шаблон}} = 0,765574. \quad (8)$$

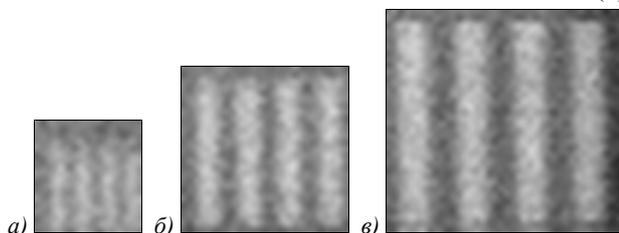


Рис. 2. Пример шаблонных изображений, соответствующих размерам тест-объектов из тестового набора кадров в увеличенном масштабе отображения

Кривая precision/recall для данного способа представлена на рис. 12 (см. кривую «Поиск по шаблону»).

К достоинствам данного способа следует отнести высокую точность и полноту распознавания, стабильность распознавания при незначительных изменениях яркости изображений. К недостаткам – чувствительность способа к внесению в изображения каких-либо изменений, например, шума с различным распределением, добавление которого в исходные изображения приводит к резкому ухудшению качества распознавания.

Также следует отметить, что при расположении тест-объектов вблизи края изображения в случае, когда часть тест-объектов выходит за пределы поля зрения, данный способ выдает ложные положительные ответы при наличии близких по размеру шаблонных изображений.

На рис. 3 представлена центральная часть кадра с изображением четырех тест-объектов различных угловых размеров, распознанных верно с применением шаблонных изображений, представленных на рис. 2.

На рис. 4 показан пример работы данного способа, на котором приведена часть кадра, соответствующая правому краю поля зрения, с изображением двух тест-объектов различных угловых размеров, где нижний тест-объект распознан верно, а верхний тест-объект распознан два раза при определении уровня корреляции по параметру $\gamma(x, y) = 0,55$ с двумя шаблонными изображениями, близкими по угловым раз-

мерам, что приводит к необходимости применения дополнительных проверок для определения конкретного типа тест-объекта.

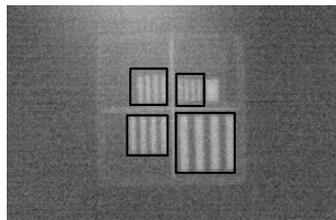


Рис. 3. Пример центральной части кадра с изображением четырех тест-объектов различных угловых размеров, распознанных верно с использованием шаблонных изображений

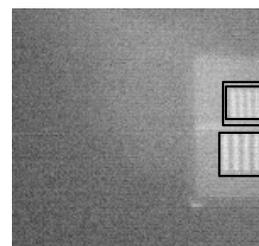


Рис. 4. Пример правой части кадра с изображением двух тест-объектов различных угловых размеров, где верхний тест-объект распознан дважды

Кроме того, для получения необходимых результатов с применением данного способа в конкретных условиях эксплуатации определенного типа тепловизионных приборов в заданном поле зрения в режиме внешнего воздействия заданных значений температур необходимо использовать отдельное шаблонное изображение, что ведет к необходимости хранения достаточно большого количества шаблонных изображений, а также создает дополнительные сложности сопровождения программного обеспечения в условиях его эксплуатации при возникновении необходимости проверки нового типа тепловизионных приборов, а также при необходимости изменения условий проведения испытаний.

3. Способ, основанный на применении метода Виолы–Джонса

Данный способ был разработан и представлен в 2001 году Полом Виолой и Майклом Джонсом и хорошо зарекомендовал себя в области распознавания лиц на изображениях [2, 3]. В основе метода Виолы–Джонса лежат такие ключевые моменты, как интегральное представление изображения, признаки Хаара, бустинг [9, 10], каскад классификаторов.

Применение способа состоит из двух этапов – этапа обучения (построения каскада классификаторов) и этапа распознавания. Для обучения и тестирования данного способа подготавливают две выборки изображений: положительную с изображениями объектов, которые необходимо распознать, и отрицательную, состоящую из изображений без объектов, которые необходимо распознать.

С целью проведения экспериментов по определению качества распознавания тест-объектов на тепло-

визионных изображениях с использованием способа, основанного на применении метода Виолы–Джонса, подготовлена положительная выборка, содержащая 2000 изображений тест-объектов в различных масштабах с различным уровнем контраста (рис. 5), а также отрицательная выборка, содержащая 4500 изображений, на которых тест-объекты отсутствуют (рис. 6).

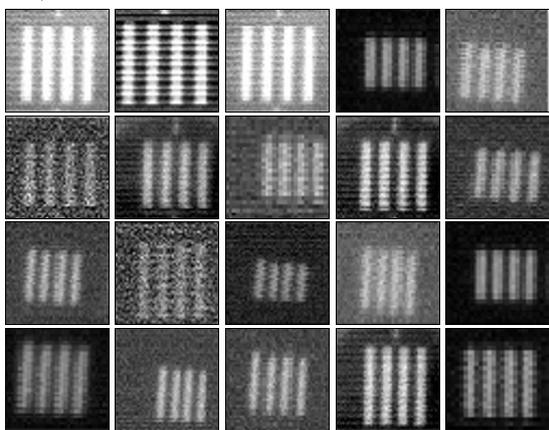


Рис. 5. Примеры изображений из положительной выборки, используемой при обучении классификатора на основе метода Виолы–Джонса

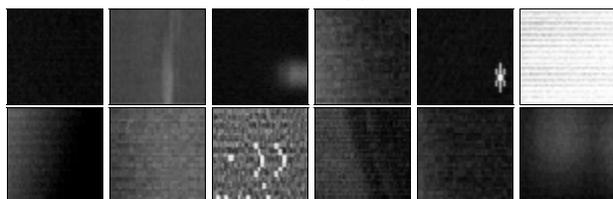


Рис. 6. Примеры изображений из отрицательной выборки, используемой при обучении классификатора на основе метода Виолы–Джонса

Этап обучения на персональном компьютере с процессором Intel Core i7 и объемом оперативной памяти 16 ГБ занял 160 часов.

С использованием данного способа проведены эксперименты по определению качества распознавания тест-объектов, в результате которых каскад классификаторов настроен на получение максимального значения площади F под кривой precision/recall, рассчитываемого по формуле (1), при этом:

$$F_{\text{Виола-Джонс}} = 0,567923. \quad (9)$$

Значения точности P и полноты R получились равными:

$$P_{\text{Виола-Джонс}} = 0,815299; \quad (10)$$

$$R_{\text{Виола-Джонс}} = 0,716393. \quad (11)$$

Кривая precision/recall, полученная при данных значениях, представлена на рис. 12 (см. кривую «Виола–Джонс»).

Пример работы обученного каскада классификаторов на этапе распознавания тест-объектов представлен на рис. 7.

К достоинствам данного способа следует отнести инвариантность к изменению размера изображений тест-объектов, то есть для распознавания тест-объектов различных размеров используется один и

тот же классификатор без подстройки каких-либо параметров (рис. 7).

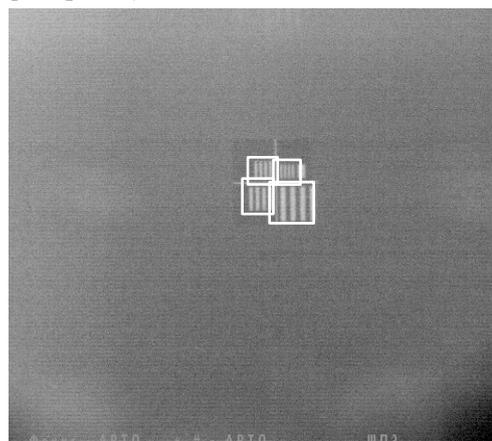


Рис. 7. Пример кадра с изображением четырех тест-объектов различных угловых размеров, распознанных верно с использованием метода Виолы–Джонса

Так же, как и первый способ распознавания, основанный на корреляционном сопоставлении, данный способ показывает хорошую стабильность распознавания при незначительных изменениях яркости изображений, на которых требуется найти объект.

К недостаткам данного способа следует отнести то, что при настройке параметров каскада классификаторов на получение максимального значения площади под кривой precision/recall: $F_{\text{Виола-Джонс}} = 0,567923$, получаются недостаточно высокие значения точности и полноты распознавания: $P = 0,815299$; $R = 0,716393$ соответственно. По сравнению со способом, основанным на корреляционном сопоставлении, данный способ уступает в качестве распознавания (рис. 12).

Однако способ, основанный на применении метода Виолы–Джонса, можно настроить, чтобы при незначительном уменьшении значения площади под кривой precision/recall получить высокие значения полноты R (рис. 12 кривая «Виола–Джонс Макс. R »), что является предпосылкой для использования данного способа с целью предварительного отбора областей интереса совместно с другими типами классификаторов, например, классифицирующими нейронными сетями глубокого обучения.

4. Способ, основанный на применении сверточной нейронной сети LeNet

Сверточная нейронная сеть LeNet (полное название LeNet-5) представлена Яном Лекуном в 1998 году [4, 5]. Данная сеть разработана для распознавания рукописных цифр из базы данных MNIST [11]. Принцип построения большинства современных сверточных нейронных сетей глубокого обучения основан на использовании тех же блоков, что и сеть LeNet-5, так что эту сеть можно считать одной из первых опубликованных сверточных нейронных сетей глубокого обучения.

С целью проведения экспериментов по определению качества распознавания тест-объектов на тепловизионных изображениях использована сверточная

нейронная сеть LeNet, входящая в состав пакета фреймворка Caffe [12]. Для обучения сети использована выборка, представленная в параграфе 3, содержащая 2000 изображений класса «тест-объект» (рис. 5), а также 4500 изображений класса «не тест-объект» (рис. 6). Все изображения преобразованы к размеру 28×28 пикселей. До момента начала обучения сети из выборки было взято по 500 изображений каждого класса, которые предназначались исключительно для тестирования сети. В процессе обучения сети пройдено 200 эпох, сеть обучена для распознавания двух классов изображений: «тест-объект» и «не тест-объект».

Оценка качества распознавания сверточной нейронной сетью LeNet выполнена по следующей схеме: для выделения областей интереса выбран способ на основе метода Виолы–Джонса, выделенные области преобразовывались к размеру 28×28 пикселей и подавались на вход сверточной нейронной сети LeNet. Сверточная нейронная сеть LeNet выдавала значение вероятности отношения классифицируемого изображения к классу «тест-объект» в диапазоне значений $[0; 1]$.

Параметры метода Виолы–Джонса, а также пороговое значение вероятности отношения классифицируемого изображения нейронной сетью LeNet к классу «тест-объект» были настроены на получение максимального результирующего функционала качества. При этом значение площади F под кривой precision/recall метода Виолы–Джонса получилось равным:

$$F_{\text{Виола-Джонс}} = 0,50832, \quad (12)$$

значения точности P и полноты R метода Виолы–Джонса равны:

$$P_{\text{Виола-Джонс}} = 0,603571; \quad (13)$$

$$R_{\text{Виола-Джонс}} = 0,831967, \quad (14)$$

а пороговое значение вероятности отношения классифицируемого изображения нейронной сетью LeNet к классу «тест-объект» равно значению 0,98.

Кривая precision/recall, полученная в результате экспериментов, показана на рис. 12 (кривая «Виола-Джонс + LeNet»).

По формуле (1) рассчитано значение площади F под кривой precision/recall:

$$F_{\text{Виола-Джонс}} = 0,586555. \quad (15)$$

Значения точности P и полноты R получились равными:

$$P_{\text{Виола-Джонс+LeNet}} = 0,90401; \quad (16)$$

$$R_{\text{Виола-Джонс+LeNet}} = 0,610656. \quad (17)$$

К достоинствам данного способа следует отнести инвариантность к изменению размера тест-объектов (рис. 8). Данный способ показывает достаточно высокую точность распознавания при незначительных изменениях яркости изображений, на которых требуется найти объект, однако полнота распознавания при этом уменьшается (рис. 9). Также к достоинствам данной сети следует отнести высокую скорость распознавания, которая с использованием графиче-

ского ускорителя NVidia Quadro K5200 составляет порядка десятка миллисекунд на один объект.

К недостаткам данного способа следует отнести невысокое значение площади под кривой precision/recall: $F_{\text{Виола-Джонс+LeNet}} = 0,586555$.

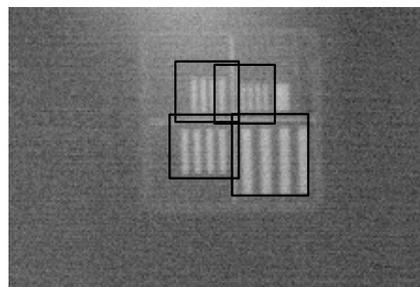


Рис. 8. Пример центральной части кадра с изображением четырех тест-объектов различных угловых размеров, распознанных верно с использованием сверточной нейронной сети LeNet

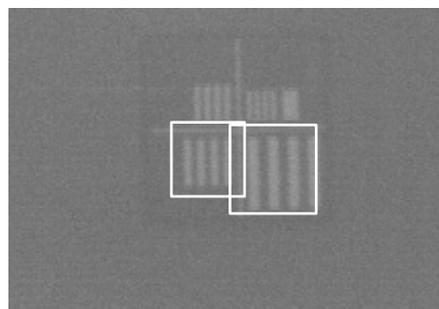


Рис. 9. Пример центральной части кадра с изображением четырех тест-объектов различных угловых размеров, два из которых распознаны верно, а два не распознаны, в качестве классификатора применена сверточная нейронная сеть LeNet

5. Способ, основанный на применении сверточной нейронной сети GoogLeNet

Сверточная нейронная сеть GoogLeNet [6] победила в соревновании ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14) в 2014-м году с результатом 6,67 % top 5 error. К отличительным особенностям данной сети следует отнести достаточно небольшой размер модели, высокую скорость вычислений при хорошем качестве распознавания.

С целью проведения экспериментов по определению качества распознавания тест-объектов на тепловизионных изображениях использована сверточная нейронная сеть GoogLeNet (Inception v.1), входящая в состав пакета фреймворка Caffe [12]. Для обучения сети использована выборка, представленная в параграфе 3, содержащая 2000 изображений класса «тест-объект» (рис. 5), а также 4500 изображений класса «не тест-объект» (рис. 6). Все изображения преобразованы к размеру 224×224 пикселей. До момента начала обучения сети из выборки было взято по 500 изображений каждого класса, которые предназначались исключительно для тестирования сети. В процессе обучения сети пройдено 200 эпох, сеть обучена для распознавания двух классов изображений: «тест-объект» и «не тест-объект».

Оценка качества распознавания сверточной нейронной сетью GoogLeNet выполнена по следующей схеме: для выделения областей интереса выбран способ на основе метода Виолы–Джонса, выделенные области преобразовывались к размеру 224×224 пикселей и подавались на вход сверточной нейронной сети GoogLeNet. Сверточная нейронная сеть GoogLeNet выдавала значение вероятности отношения классифицируемого изображения к классу «тест-объект» в диапазоне значений [0; 1].

Параметры метода Виолы–Джонса, а также пороговое значение вероятности отношения классифицируемого изображения нейронной сетью GoogLeNet к классу «тест-объект» были настроены на получение максимального результирующего функционала качества. При этом значение площади F под кривой precision/recall метода Виолы–Джонса получилось равным:

$$F_{\text{Виола-Джонс}} = 0,50832, \tag{18}$$

значения точности P и полноты R метода Виолы–Джонса равны:

$$P_{\text{Виола-Джонс}} = 0,603571; \tag{19}$$

$$R_{\text{Виола-Джонс}} = 0,831967, \tag{20}$$

а пороговое значение вероятности отношения классифицируемого изображения нейронной сетью GoogLeNet к классу «тест-объект» равно значению 0,999.

Кривая precision/recall, полученная в результате экспериментов, показана на рис. 12 (кривая «Виола–Джонс + GoogLeNet»).

По формуле (1) рассчитано значение площади F под кривой precision/recall:

$$F_{\text{Виола-Джонс+GoogLeNet}} = 0,501443. \tag{21}$$

Значения точности P и полноты R получились равными:

$$P_{\text{Виола-Джонс+GoogLeNet}} = 0,788462; \tag{22}$$

$$R_{\text{Виола-Джонс+GoogLeNet}} = 0,638525. \tag{23}$$

Следует отметить, что применение сверточной нейронной сети GoogLeNet повышает точность P распознавания способа на основе метода Виолы–Джонса (рис. 12 кривая «Виола–Джонс + GoogLeNet»). Однако значение площади F под кривой precision/recall с применением сети GoogLeNet остается примерно таким же, как и при использовании способа на основе метода Виолы–Джонса без применения дополнительных классификаторов (рис. 12 кривая «Виола–Джонс Макс. R »), так как значение полноты R распознавания с добавлением сети GoogLeNet уменьшается (рис. 12 кривая «Виола–Джонс + GoogLeNet»).

Кроме того, по результатам проведенных экспериментов значение площади F (15) под кривой precision/recall с использованием сверточной нейронной сети LeNet оказалось выше, чем значение площади F (21) под кривой precision/recall с использованием сверточной нейронной сети GoogLeNet. В данной ситуации напрашивается вывод о том, что более простая по своей структуре сверточная нейронная сеть LeNet показывает лучшие результаты по сравнению с более

глубокой сверточной нейронной сетью GoogLeNet. Однако при подробном какадровом рассмотрении результатов работы сверточной нейронной сети GoogLeNet было обнаружено, что данная сеть способна распознавать даже такие тест-объекты, которые частично выходят за края изображения, в подобных ситуациях сетью LeNet тест-объекты не распознаются.

Кроме того, после проверки обеих сверточных нейронных сетей по качеству распознавания на дополнительной выборке, состоящей из двухсот изображений, в которой присутствовали тест-объекты с большими размерами, чем в исходной выборке, а именно порядка 100×100 пикселей, сверточная нейронная сеть GoogLeNet (рис. 10а) показала лучшие результаты по сравнению со сверточной нейронной сетью LeNet (рис. 10б) за счет меньшего количества ложных срабатываний и, соответственно, более высокого значения точности P (табл. 1).

Табл. 1. Результаты проведенных экспериментов, полученные на дополнительной выборке для нейронных сетей LeNet и GoogLeNet

Параметр	Знач., отн. ед.	Параметр	Знач., отн. ед.
$P_{\text{Виола-Джонс+GoogLeNet}}$	0,80	$P_{\text{Виола-Джонс+LeNet}}$	0,759
$R_{\text{Виола-Джонс+GoogLeNet}}$	0,82	$R_{\text{Виола-Джонс+LeNet}}$	0,817
$F_{\text{Виола-Джонс+GoogLeNet}}$	0,67	$F_{\text{Виола-Джонс+LeNet}}$	0,600

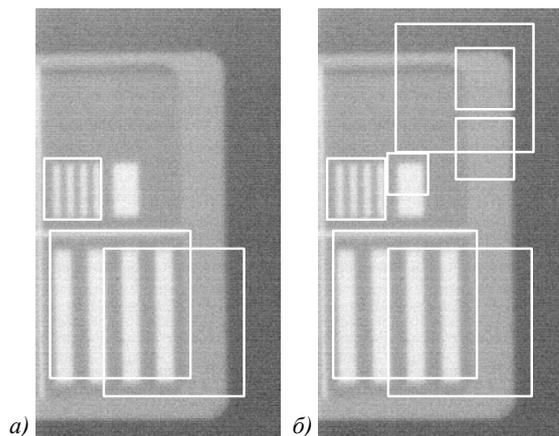


Рис. 10. Пример изображения части кадра, соответствующей левому краю поля зрения: выделенные области классифицированы сетью GoogLeNet как тест-объекты (а), выделенные области классифицированы сетью LeNet как тест-объекты (б)

К достоинствам способа на основе сверточной нейронной сети GoogLeNet следует отнести инвариантность к изменению размера тест-объектов, хорошую точность распознавания при незначительных изменениях яркости изображений, на которых требуется найти тест-объект, высокую скорость классификации, которая составляет порядка нескольких десятков миллисекунд на одну выделенную область с использованием графического ускорителя NVidia Quadro K5200.

Недостатком данного способа является невысокое значение площади под кривой precision/recall:

$$F_{\text{Виола-Джонс+GoogLeNet}} = 0,673788 \text{ — на дополнительной выборке.}$$

$$F_{\text{Виола-Джонс+GoogLeNet}} = 0,501443 \text{ на основной выборке.}$$

6. Способ, основанный на применении сверточной нейронной сети глубокого обучения типа Single Shot Multibox Detector

Сверточные нейронные сети глубокого обучения типа Single Shot Multibox Detector (SSD) [7] используются в качестве детекторов каких-либо объектов на изображении. Сети данной конфигурации построены таким образом, что они способны самостоятельно выделять области интереса, а затем присваивать данным областям вероятности принадлежности к одному из известных сети классов. При использовании данных типов сетей нет необходимости использовать методы предварительного выделения областей интереса. При обучении на вход сети подаются изображения заданного конфигурацией сети размера, например 512×512 пикселей с разметкой в виде прямоугольных областей, в которых находятся объекты того или иного класса. При детектировании на вход сети подаются изображения того же размера, что и при обучении, например 512×512 пикселей, сеть пропускает через себя данные изображения, применяя веса, полученные при обучении, и на выходе формирует разметку кадров с вероятностями принадлежности к одному из известных сети классов.

Для обучения данной сети была подготовлена выборка из 8500 изображений размером 512×512 пикселей, с разметкой тест-объектов на каждом изображении в виде ограничивающих прямоугольников. Часть изображений выборки получена из видеопотоков, формируемых тепловизионными приборами, часть получена из имеющихся изображений путем выполнения аугментации в виде изменения яркости и контраста, изменения масштаба, зеркального горизонтального отображения, а также внесения в имеющиеся изображения искусственного шума с различными распределениями.

В качестве начальных весов использовались веса предобученной модели сверточной нейронной сети VGG16 «models_VGGNet_VOC0712_SSD_512x512» типа SSD. Сеть обучалась на протяжении 50 эпох.

На рис. 11 в качестве примера показан один из сложных случаев для распознавания тест-объектов в проводимых экспериментах, где сверточной нейронной сетью SSD не был распознан лишь правый верхний тест-объект в центральной части кадра, однако следует отметить, что человеком-оператором данный тест-объект также трактуется как нераспознаваемый.

По результатам экспериментов с использованием сверточной нейронной сети SSD, построена кривая precision/recall (рис. 12 кривая «SSD») и по формуле (1) рассчитано значение площади F под кривой precision/recall:

$$F_{SSD} = 0,781971. \quad (24)$$

При этом значения точности P и полноты R получились равными:

$$P_{SSD} = 0,993757; \quad (25)$$

$$R_{SSD} = 0,783607. \quad (26)$$

На дополнительной выборке изображений, состоящей из 200 изображений, в которой присутствовали тест-объекты больших размеров по сравнению с основной выборкой, порядка 100×100 пикселей, сверточная нейронная сеть глубокого обучения SSD показала следующие результаты:

$$F_{SSD} = 0,959615. \quad (27)$$

При этом значение точности P и полноты R получились равными:

$$P_{SSD} = 1; \quad (28)$$

$$R_{SSD} = 0,961538. \quad (29)$$

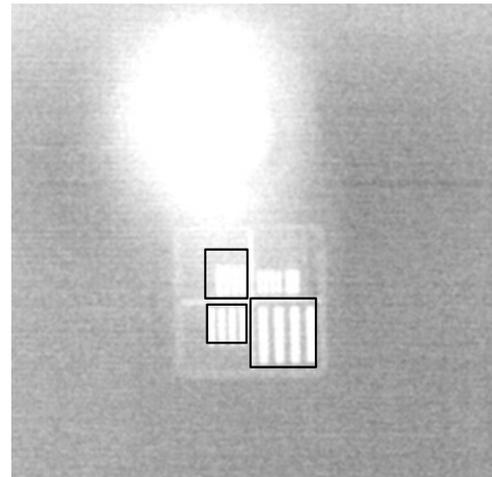


Рис. 11. Пример детектирования тест-объектов с использованием сети SSD

К основным достоинствам способа распознавания на основе сверточной нейронной сети глубокого обучения SSD следует отнести инвариантность к изменению размеров тест-объектов, высокую точность распознавания при незначительных изменениях яркости изображений, на которых требуется найти объект, а также высокое значение площади под кривой precision/recall: $F_{SSD} = 0,781971$ – на основной выборке, $F_{SSD} = 0,959615$ – на дополнительной выборке. К достоинствам также следует отнести то, что в отличие от нейронных сетей LeNet и GoogleNet данный способ не требует применения дополнительных методов для локализации областей интереса.

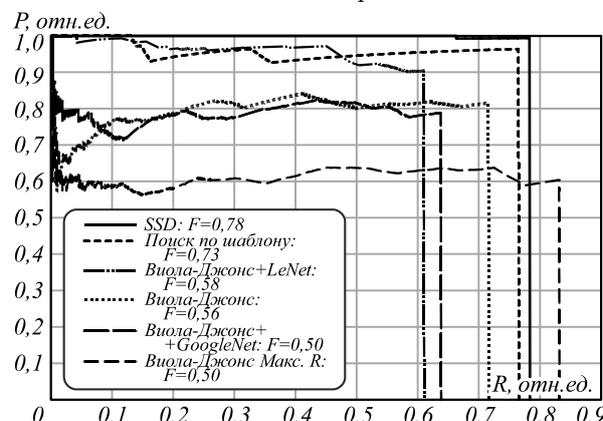


Рис. 12. График кривых precision/recall для рассмотренных способов по основной выборке изображений

К недостаткам способа следует отнести невысокую скорость распознавания, которая с использованием графического ускорителя NVidia Quadro K5200 составляет несколько кадров в секунду, а также достаточно длительный процесс обучения сети.

Выводы

График кривых precision/recall для всех рассмотренных выше способов на основном наборе данных представлен на рис. 12.

В табл. 2 представлены значения точности, полноты распознавания, а также площади под кривой precision/recall, полученные на основном наборе данных для всех рассмотренных выше способов распознавания тест-объектов в порядке убывания значения площади под кривой precision/recall. В табл. 3 представлены значения точности, полноты распознавания, а также площади под кривой precision/recall, полученные на дополнительной выборке изображений для всех рассмотренных выше способов, кроме корреляционного сопоставления, в порядке убывания значения площади под кривой precision/recall.

Табл. 2. Результаты проведенных экспериментов, полученные на основном наборе данных

Способ распознавания	Точность P , отн. ед.	Полнота R , отн. ед.	Площадь под кривой Precision/recall F , отн. ед.
SSD	0,993757	0,783607	0,781971
Корреляционное сопоставление	0,963843	0,765574	0,733219
Виола-Джонс + Lenet	0,904010	0,610656	0,586555
Виола-Джонс	0,815126	0,715574	0,508320
Виола-Джонс + GoogLeNet	0,788462	0,638525	0,501443

Среди рассмотренных способов распознавания тест-объектов наибольшее значение площади под кривой precision/recall $F=0,781971$ на основном наборе данных получено при использовании способа на основе сверточной нейронной сети глубокого обучения SSD, при этом способ на основе корреляционного сопоставления изображений позволил получить близкое значение площади под кривой precision/recall $F=0,733219$, однако применение способа на основе корреляционного сопоставления изображений в реальных условиях эксплуатации приводит к необходимости хранения достаточно большого количества шаблонных изображений, а также к необходимости добавления новых шаблонных изображений в случае проведения проверки нового типа тепловизионных каналов. Кроме того, способ на основе корреляционного сопоставления изображений крайне неустойчив к шумам на изображении, что говорит о недостаточной универсальности данного способа.

Способ на основе сверточной нейронной сети глубокого обучения SSD позволил получить высокие значения точности распознавания $P_{SSD}=0,993757$,

полноты распознавания $R_{SSD}=0,783607$ на основной выборке изображений, а также значение точности распознавания $P_{SSD}=1$ и значение полноты распознавания $R_{SSD}=0,961538$ на дополнительной выборке изображений. При анализе результатов распознавания тест-объектов с использованием данного способа видно, что тест-объекты, которые не были распознаны нейронной сетью, человеком-оператором также не распознаются.

Табл. 3. Результаты проведенных экспериментов, полученные на дополнительной выборке

Способ распознавания	Точность P , отн. ед.	Полнота R , отн. ед.	Площадь под кривой precision/recall F , отн. ед.
SSD	1	0,961538	0,959615
Виола-Джонс + GoogLeNet	0,80	0,82	0,673788
Виола-Джонс + Lenet	0,759	0,817	0,60

Применение способа на основе сверточной нейронной сети глубокого обучения SSD с целью вычисления координат положения тест-объектов на тепловизионном изображении на первом этапе распознавания с погрешностью в несколько пикселей позволит выполнять работы по настройке и проведению проверок характеристик тепловизионных каналов в таком режиме, когда весь процесс выполнения измерительных операций проводится без непосредственного участия человека-оператора.

Литература

1. **Gonzalez, R.** Digital image processing / R. Gonzalez, R. Woods. – 3rd ed. – Prentice Hall, Inc., 2008. – 976 p.
2. **Viola, P.** Rapid object detection using a boosted cascade of simple features / P. Viola, M.J. Jones // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001). – 2001.
3. **Viola, P.** Robust real-time face detection / P. Viola, M.J. Jones // International Journal of Computer Vision. – 2004. – Vol. 57, Issue 2. – P. 137-154.
4. **LeCun, Y.** Gradient based learning applied to document recognition / Y. LeCun, L. Bottou, Y. Bengio, P. Haffner // Proceedings of the IEEE. – 1998. – Vol. 86, Issue 11. – P. 2278-2324.
5. **LeCun, Y.** Convolutional networks for images, speech and time series / Y. LeCun, Y. Bengio. – In: The handbook of brain theory and neural networks / ed. by M.A. Arbib. – Cambridge, MA: MIT Press, 1998. – P. 255-258.
6. **Szegedy, Ch.** Going deeper with convolutions [Electronical Resource] / Ch. Szegedy, W. Liu, Ch. Hill, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich // arXiv:1409.4842v1 [cs.CV]. – URL: <https://arxiv.org/abs/1409.4842> (request date 14.03.2019).
7. **Liu, W.** SSD: Single shot multibox detector [Electronical Resource] / W. Liu, D. Anguelov, D. Erhan, Ch. Szegedy, S. Reed, Ch.-Y. Fu, A.C. Berg // arXiv:1512.02325v5 [cs.CV]. – URL: <https://arxiv.org/abs/1512.02325> (request date 14.03.2019).
8. **Saito, T.** The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets / T. Saito, M. Rehmsmeier // PLoS One. – 2015. – Vol. 10, Issue 3. – e0118432.

9. Šochman, J. AdaBoost / J. Šochman, J. Matas. – Prague: Center for Machine Perception, Czech Technical University, 2010.
10. Freund, Y. A short introduction to boosting / Y. Freund, R.E. Schapire. – Shannon Laboratory, USA, 1999. – P. 771-780.
11. The MNIST database of handwritten digits [Electronical Resource]. – URL: <http://yann.lecun.com/exdb/mnist> (request date 12.09.2017).
12. Caffe [Electronical Resource]. – URL: <http://Caffe.berkeleyvision.org> (request date 02.10.2017).

Сведения об авторах

Мингалев Александр Владимирович, 1981 года рождения, начальник сектора по разработке программного обеспечения.

Белов Андрей Вячеславович, 1993 года рождения, инженер-программист сектора по разработке программного обеспечения.

Габдуллин Ильдар Масхутович, 1993 года рождения, инженер-программист сектора по разработке программного обеспечения.

Агафонова Регина Ренатовна, 1992 года рождения, инженер-программист III категории сектора по разработке программного обеспечения.

Шушарин Сергей Николаевич, 1970 года рождения, начальник отдела.

ГРНТИ: 50.47.02

Поступила в редакцию 17 июня 2018 г. Окончательный вариант – 17 марта 2019 г.

Test-object recognition in thermal images

A.V. Mingalev¹, A.V. Belov¹, I.M. Gabdullin¹, R.R. Agafonova¹, S.N. Shusharin¹
¹JSC “Scientific and Production Association “State Institute of Applied Optics”, Kazan, Russia

Abstract

The paper presents a comparative analysis of several methods for recognition of test-object position in a thermal image when setting and testing characteristics of thermal image channels in an automated mode. We consider methods of image recognition based on the correlation image comparison, Viola-Jones method, LeNet classificatory convolutional neural network, GoogleNet (Inception v.1) classificatory convolutional neural network, and a deep-learning-based convolutional neural network of Single-Shot Multibox Detector (SSD) VGG16 type. The best performance is reached via using the deep-learning-based convolutional neural network of the VGG16-type. The main advantages of this method include robustness to variations in the test object size; high values of accuracy and recall parameters; and doing without additional methods for RoI (region of interest) localization.

Keywords: image classification, object detection in images, image recognition, deep-learning-based convolutional neural network, thermal image, thermal imaging device.

Citation: Mingalev AV, Belov AV, Gabdullin IM, Agafonova RR, Shusharin SN. Test-object recognition in thermal images. Computer Optics 2019; 43(3): 402-411. DOI: 10.18287/2412-6179-2019-43-3-402-411.

References

- [1] Gonzalez R, Woods R. Digital image processing. 3rd ed. Prentice Hall Inc; 2008.
- [2] Viola P, Jones MJ. Rapid object detection using a boosted cascade of simple features. Proc IEEE Conf on Comp Vision and Pattern Recogn (CVPR 2001) 2001.
- [3] Viola P, Jones MJ. Robust real-time face detection. Int J Comp Vision 2004; 57(2): 137-154.
- [4] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient based learning applied to document recognition. Proc IEEE 1998; 86(11): 2278-2324.
- [5] LeCun Y, Bengio Y. Convolutional networks for images, speech and time series. In Book: Arbib MA, ed. The handbook of brain theory and neural networks. Cambridge, MA: MIT Press; 1998: 255-258.
- [6] Szegedy Ch, Liu W, Hill Ch, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. arXiv:1409.4842v1 [cs.CV]. Source: (<https://arxiv.org/abs/1409.4842>).
- [7] Liu W, Anguelov D, Erhan D, Szegedy Ch, Reed S, Fu Ch-Y, Berg AC. SSD: Single shot multibox detector. arXiv:1512.02325v5 [cs.CV]. Source: (<https://arxiv.org/abs/1512.02325>).
- [8] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015; 10(3): e0118432.
- [9] Šochman J, Matas J. AdaBoost. Prague: Center for Machine Perception, Czech Technical University; 2010.
- [10] Freund Y, Schapire RE. A short introduction to boosting», Shannon Laboratory, USA; 1999: 771-780.

-
- [11] The MNIST database of handwritten digits. Source: <http://yann.lecun.com/exdb/mnist>. | [12] Caffe. Source: <http://Caffe.berkeleyvision.org>.
-

Authors' information

Aleksandr Vladimirovich Mingalev, 1981 year of birth, chief of Software Design department.

Andrey Vyacheslavovich Belov, 1993 year of birth, programming engineer of Software Design department.

Ildar Maskhutovich Gabdullin, 1993 year of birth, programming engineer of Software Design department.

Regina Renatovna Agafonova, 1992 year of birth, the third category programming engineer of Software Design department.

Sergey Nikolaevich Shusharin, 1970 year of birth, chief of department.

Received June 17, 2018. The final version – March 17, 2019.
