

УДК 004.932
DOI: 10.15827/0236-235X.117.130-137

Дата подачи статьи: 16.12.16
2017. Т. 30. № 1. С. 130–137

АЛГОРИТМ ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ НА ФОТОСНИМКАХ С НИЗКИМ КАЧЕСТВОМ ИЗОБРАЖЕНИЯ

*А.С. Викторов, аспирант, alsevictor@mail.ru
(Костромской государственной университет,
ул. Дзержинского, 17, г. Кострома, 156005, Россия)*

В статье рассматривается набор алгоритмов, применяемых для распознавания объектов определенного класса на фотоснимках с некачественным изображением, полученных с видеокамеры низкого разрешения.

Особенностью рассматриваемой методики детектирования объектов является возможность обнаружения объектов, размеры изображений которых на фотоснимках не превышают нескольких десятков пикселей. Исследуемое изображение сканируется скользящим окном, считывающим участки изображения с заданным перекрытием между соседними участками. Сканируемые участки изображения предварительно обрабатываются дискриминативным автокодировщиком, извлекающим вектор признаков из участка изображения, который анализируется мультиклассовым классификатором, построенным на основе вероятностной модели регрессии, на предмет наличия изображения или части изображения объекта. Для каждого сканируемого участка изображения классификатор вычисляет значение вероятности обнаружения детектируемого объекта определенного класса на данном участке. На основании результатов сканирования изображения делается вывод о наличии изображения объекта и о его наиболее вероятном положении на фотоснимке.

Для повышения точности обнаружения границ изображения значение вероятности обнаружения детектируемого объекта определенного класса интерполируется для каждого анализируемого пикселя изображения. После детектирования пикселей на основании их распределения на изображении уточняются границы изображения детектируемого объекта.

В ходе проведенного исследования было обнаружено, что использование дискриминативного автокодировщика значительно повысило робастность алгоритма детектирования. В статье дано подробное описание процесса обучения и настройки параметров алгоритмов, используемых в процессе детектирования. Результаты данного исследования могут найти широкое применение для автоматизации различных процессов, например для сбора и анализа информации в различных аналитических системах.

Ключевые слова: каскадный шумоподавляющий автокодировщик, метод релевантных векторов, нейронная сеть, функция потерь, обучающая выборка, признаковое описание объекта, детектирование объектов, функции правдоподобия.

Основной подход, предлагаемый в современной литературе для детектирования объектов на изображении, заключается в использовании 2D-признаков, инвариантных к возможному искажению изображения объекта, вызванному аффинными преобразованиями [1], деформациями сдвига, изменением освещения или масштаба изображения объекта [2] и т.д. Для извлечения признаков из изображений, например, широко используется алгоритм SIFT (Scale-invariant feature transform), впервые рассмотренный в [3]. Для достижения качественного результата в процессе обучения классификатора при использовании признакового описания объекта путем поиска особых точек и вычисления дескрипторов их окрестностей требуется специальная подготовка обучающего набора исходных изображений и отбора уникальных признаков объекта для формирования визуального словаря, например, по методике, описанной в [4]. После извлечения 2D-признаков из тестового изображения производится их сравнение с признаками [5], хранящимися в БД. Чтобы распознать некоторый известный объект, система распознавания сначала извлекает множество характерных точек из изображений объекта, сделанных с различного ракурса, и запоминает извлеченные из них признаки в проиндексированную структуру, например та-

кую, как дерево поиска. Во время распознавания признаки извлекаются из проверяемого изображения и сравниваются с сохраненными признаками объекта. Каждый раз, когда заданное число признаков, извлеченных из тестового изображения, совпадает с признаками, описывающими детектируемый объект, система распознавания вызывает процедуру верификации – проверку на взаимное совпадение пространственного положения множества точек на тестовом изображении, из которого были извлечены признаки, с взаимным положением характерных точек, описывающих объект.

При формировании признакового описания объекта по данной методике возникает сложность процедуры формирования набора признаков, по которому однозначно можно идентифицировать объект, что приводит к появлению ложных распознаваний. Появление ложных распознаваний связано с тем, что исходные изображения определенного класса объектов могут содержать множество помех и похожие признаки могут принадлежать объектам различных классов. Поэтому для повышения надежности распознавания требуется извлечь из тренировочного набора изображений как можно больше признаков и сформировать словарь визуальных слов [6] для формирования набора уникальных признаков, что невозможно эффективно

реализовать для задачи детектирования объектов при низком разрешении изображения данных объектов. Задача поиска объектов определенного класса на фотоснимках, полученных с видеокамеры наружного наблюдения, например задача распознавания транспортных средств на этих фотоснимках, связана с детектированием объектов, как правило, имеющих низкое разрешение. Для решения данной задачи необходим иной подход к формированию признакового описания объекта.

Обзор современной литературы, описывающей методику формирования признакового описания объекта, показал, что наиболее передовой подход при формировании описания объекта заключается в использовании автокодировщика [7], являющегося нейронной сетью с симметричной архитектурой. Автокодировщик состоит из входного и выходного слоев одинаковой размерности и скрытых слоев с меньшей размерностью, чем у входного и выходного слоев. Автокодировщик имеет скрытый слой, называемый bottleneck-слоем, у которого самая наименьшая размерность из всех скрытых слоев. Размерность слоев автокодировщика при переборе от входного слоя к bottleneck-слою последовательно уменьшается, а при переборе от bottleneck-слоя к выходному слою последовательно увеличивается. Набор слоев автокодировщика от входного до bottleneck-слоя называется кодировщиком, который осуществляет преобразование некоторого входного сигнала $\bar{x}_0 \in R^n$ в сигнал $\bar{x}_{bottleneck} \in R^r$, где r – количество нейронов в bottleneck-слое. Набор слоев автокодировщика от bottleneck-слоя до выходного называется декодировщиком, он осуществляет преобразование выходного сигнала с кодировщика $\bar{x}_{bottleneck} \in R^r$ в $\bar{x}_L \in R^n$, где L – количество слоев в автокодировщике. В общем случае задача обучения автокодировщика заключается в поиске такой конфигурации весов слоев автокодировщика $\bar{W} = [\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L]$ и пороговых значений $\bar{b} = [\bar{b}_1, \bar{b}_2, \dots, \bar{b}_L]$,

которая при подаче на его вход многомерного сигнала обеспечивала бы на выходе отклик, наиболее близкий к входному сигналу. В качестве функции активации для слоев автокодировщика используют нелинейную функцию активации, например сигмоид, что позволяет автокодировщику аппроксимировать сложные зависимости. Автокодировщик может применяться для уменьшения размерности входного сигнала. Уменьшение размерности входного сигнала при его подаче на вход автокодировщика осуществляется при считывании с bottleneck-слоя выходного сигнала, составляющие которого слабо коррелируют или не коррелируют между собой. Выходной сигнал с bottleneck-слоя автокодировщика далее подается на вход модуля мультиномиальной регрессии RVM (relevance vector

machine) [8], который вычисляет вероятность наличия изображения объекта определенного класса на извлеченном участке фотоснимка. По результатам сканирования всего фотоснимка производится анализ наличия изображений объектов определенного класса на фотоснимке и определение областей фотоснимка, содержащих данные изображения.

Разработка архитектуры автокодировщика и способа его обучения

Обозначим через $\bar{x}_0 \in R^n$ входной сигнал, имеющий размерность n , а через $\bar{x}_{i+1} = f_{i+1}(\bar{x}_i) = \sigma(\bar{W}_{i+1} \cdot \bar{x}_i + \bar{b}_{i+1})$, где $i = 0, \dots, L-1$, функциональную зависимость между сигналами на входе \bar{x}_i и выходе \bar{x}_{i+1} $i+1$ -го слоя, где \bar{W}_{i+1} – весовые коэффициенты $i+1$ -го слоя; \bar{b}_{i+1} – пороговые значения $i+1$ -го слоя; $\sigma(\bar{x}) = \frac{1}{1 + e^{-\bar{x}}}$ – функция ак-

тивации, в качестве которой согласно [9] выбрана сигмоид-функция. Далее запишем функциональную зависимость между входом и выходом автокодировщика в виде $\bar{x}_L = DF(\bar{x}_{bottleneck}) = DF(EF(\bar{x}_0))$,

где $EF(\bar{x}_0) = f_k(f_{k-1}(\dots(f_1(\bar{x}_0))))$ – функциональная зависимость между сигналами на входе и выходе кодировщика; $DF(\bar{x}_{bottleneck}) = f_L(f_{L-1}(\dots(f_{k+1}(\bar{x}_{bottleneck}))))$ – функциональная зависимость между сигналами на входе и выходе декодировщика; L – число слоев автокодировщика.

Задача настройки автокодировщика заключается в поиске такой конфигурации $C(\bar{W}, \bar{b})$ весов \bar{W} и пороговых значений \bar{b} , которая минимизировала бы выбранную целевую функцию. В качестве целевой функции используется функция потерь $L(\bar{x}_0, \bar{x}_L)$, определяющая меру несоответствия сигнала на выходе автокодировщика от ожидаемого сигнала при подаче на его вход заданного сигнала. Для стандартного автокодировщика, предназначенного только для сжатия информации, функция потерь имеет вид

$$L(\bar{x}_0, \bar{x}_L) = \|\bar{x}_0 - \bar{x}_L\|_2^2 = \|\bar{x}_0 - DF(EF(\bar{x}_0))\|_2^2,$$

а задача обучения автокодировщика [10] сводится к минимизации целевой функции вида

$$C(\bar{W}, \bar{b}) = \arg \min_{\bar{W}, \bar{b}} \frac{1}{2} \cdot \sum_{i=0}^{N-1} \|\bar{x}_{0,i} - DF(EF(\bar{x}_{0,i}))\|_2^2 + \frac{1}{2} \cdot \lambda \cdot \|\bar{W}\|_F^2, \text{ где } N - \text{размер обучающей выборки } X.$$

Для обучения автокодировщика дополнительному функционалу, а именно способности извлекать информацию о заданной категории объектов, необходимо использовать другую функцию потерь, которая учитывала бы наличие объекта, относящегося к определенной категории на изображении.

Для обучения автокодировщика, обладающего описанным функционалом, имеется обучающий набор изображений $X \in \chi^+ \cup \chi^-$, состоящий из N изображений. Каждому изображению из набора назначена метка класса $l(x_0) \in \{-1, 1\}$. Каждое изображение из набора имеет размер $m \times m$, при этом $m^2 = n$, где n – размерность входа автокодировщика. Изображение из набора имеет метку 1, если содержит изображение всего или части объекта, принадлежащего к классу детектируемых объектов, в противном случае изображение имеет метку -1 . В качестве функции потерь для выбранного автокодировщика предлагается модифицировать стандартную целевую функцию путем добавления в нее дополнительной функции потерь, например, как в работе [11], для учета принадлежности входного сигнала, подаваемого на вход автокодировщика к детектируемому классу объектов.

Обозначим через $\bar{z}^t = \{\bar{z}_{-1}^t, \bar{z}_1^t\}$ набор состояний bottleneck-слоя автокодировщика при его обучении, где $\bar{z}_l^t = \{z_{l,1}^t, \dots, z_{l,j}^t, \dots, z_{l,J}^t\}$ – состояние скрытого k -го bottleneck-слоя автокодировщика при подаче на его вход изображений из обучающего набора, имеющих определенную метку из набора $l \in \{-1, 1\}$, где $j=1, \dots, J$ – номер образца изображения из поднабора изображений, имеющих определенную метку; t – номер итерации процесса обучения. Тогда дополнительную функцию потерь можно записать в виде

$$L(\bar{z}^t) = \sum_{x \in \chi^+ \cup \chi^-} \max \left(0, l(x) \cdot \left(\frac{1}{|\chi^{l(x)}| - 1} \times \right. \right. \quad (1)$$

$$\left. \left. \times \sum_{\forall k | k \neq j} \left\| z_{l(x), j}^t - z_{l(x), k}^t \right\| - \log \frac{1}{|\chi^{-l(x)}|} \sum_{\forall k} e^{\left(\left\| z_{l(x), j}^t - z_{-l(x), k}^t \right\| \right)} - thr \right) \right),$$

где $|\chi^{l(x)}|$ – мощность подмножества образцов из обучающего набора с заданной меткой; \log – обозначение десятичного логарифма; $\left\| z_{l(x), j}^t - z_{l(x), k}^t \right\|$ – евклидово расстояние между состояниями bottleneck-слоя автокодировщика (при подаче на вход двух различных изображений); thr – пороговое значение, задающее диапазон разброса расстояний между состояниями bottleneck-слоя автокодировщика.

Так как на снимках, полученных с видеокамеры наружного наблюдения, размер изображений объектов, как правило, не превышает 100×100 пикселей, для их эффективного поиска методом скользящего окна экспериментально выбран размер окна 20×20 . Для обнаружения объекта изображение сканируется скользящим окном, которое последовательно перемещается по заданной направляющей сетке, обеспечивающей перекрытие между соседними сканируемыми участками изображения в 40%. Так как окно имеет размер 20×20 пикселей, перекрытие составляет 8 пикселей. Полученные при помощи аппаратуры видеонаблюдения изображения могут содержать сильную шумовую компоненту, например из-за плохой видимости или внешнего освещения, поэтому для качественной реализации решения задачи классификации необходимо производить предварительную фильтрацию изображений.

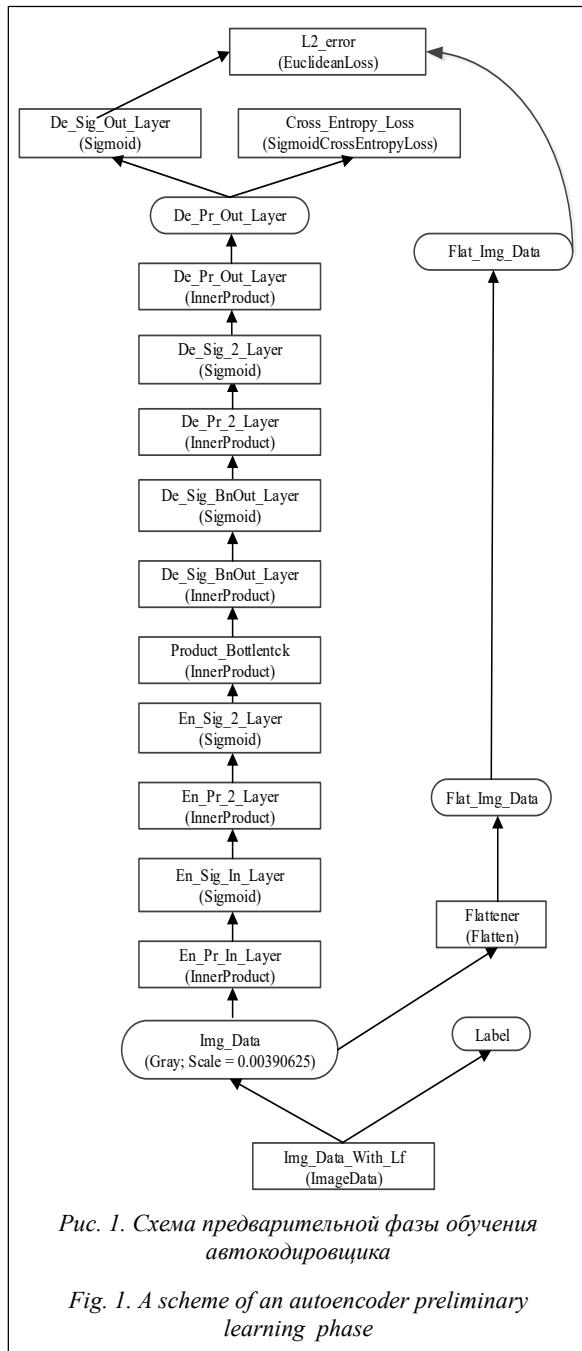
В результате поиска возможных архитектурных решений для реализации автокодировщика и методик обучения автокодировщика была выбрана концепция автокодировщика SDA (stacked denoising autoencoder) [12], который позволяет производить не только уменьшение размерности входных данных, но и их фильтрацию от возможных помех. За основу реализации процесса обучения автокодировщика была выбрана методика процесса обучения, описанная в статье [13].

Для обучения автокодировщика разработано ПО, для реализации которого используется фреймворк Caffe, который был разработан в BLVC (Berkeley Vision and Learning Center). Обучение производится в две стадии. На первой стадии автокодировщик учится восстанавливать образ, поданный на его вход, как стандартный автокодировщик, по схеме, изображенной на рисунке 1.

Для обучения автокодировщика во время проведения первой фазы обучения из обучающего набора используются только изображения, относящиеся к классу детектируемых объектов (то есть имеющие метку 1). После завершения первой фазы производится обучение обученного стандартному функционалу автокодировщика дополнительному функционалу с использованием функции потерь (1), которая производит анализ сигнала на выходе bottleneck-слоя, и функции потерь из [11], которая производит анализ сигнала выходного слоя автокодировщика (на рисунке 1 слой с названием De_Sig_Out_Layer).

Для обучения автокодировщика цветное изображение предварительно преобразуется в черно-белое, после чего производится трансформация значений интенсивности пикселей изображения из формата $[0, 255]$ в $[0, 1]$ путем умножения значений на масштабирующий коэффициент $sk = 0,00390625$.

После обучения автокодировщика производится обучение классификатора RVM, с которым



автокодировщик будет работать совместно, в соответствии со схемой, изображенной на рисунке 2.

После подачи на вход автокодировщика участка исходного изображения с его bottleneck-слоя считывается многомерный сигнал $\bar{x}_{bottleneck} = EF(\bar{x}_{0, new}) \in R^{30}$, который является сжатым и отфильтрованным образом участка исходного изображения. Полученный сигнал подается на вход классификатора RVM, обучение которого производится на маркированном наборе образов, считанных с bottleneck-слоя обученного автокодировщика при подаче на его вход изображений из обучающего набора. Изображения, используемые

для обучения автокодировщика, могут быть искусственно синтезированы из исходных изображений обучающего набора путем искусственных трансформаций и добавления шумовой составляющей.

Метод релевантных векторов для задачи распознавания

Пусть $(\bar{t}, \bar{x}) = \{t_n, \bar{x}_n\}_{n=1}^N$ – обучающая выборка, полученная из обучающего набора изображений после их обработки автокодировщиком, где $\bar{x}_n = x_{bottleneck, n} \in R^r$ – вектор признаков, извлеченный из изображения обучающего набора при помощи автокодировщика; N – число изображений в обучающем наборе; $\bar{t}_n = (t_0, \dots, t_K) \in R^K$ – вектор, j -й элемент которого равен 1, а остальные элементы нулевые, что указывает на принадлежность рассматриваемого образца к j -му классу объектов, при этом метки от 1 до K соответствуют различным классам детектируемых объектов (например, класс детектируемых объектов «автотранспортные средства» состоит из подклассов «легковые», «грузовые» средства и т.д.).

Необходимо определить, содержит ли изображение, представленное своим вектором признаков $\bar{x} \in R^r$, извлеченным при помощи автокодировщика, изображение объектов определенного класса; если да, то вывести метку класса изображенного объекта.

Рассмотрим сущность метода RVM, предназначенного для решения задачи восстановления регрессии, при этом поиск решения осуществляется в классе линейных решающих правил [14]:

$$y = f(\bar{x}, \bar{w}) + \varepsilon = \varepsilon + \sum_{i=0}^M w_i \cdot \varphi_i(\bar{x}) = \bar{w}^T \cdot \bar{\varphi}(\bar{x}) + \varepsilon,$$

где \bar{x} – вектор признаков; y – скалярное значение;

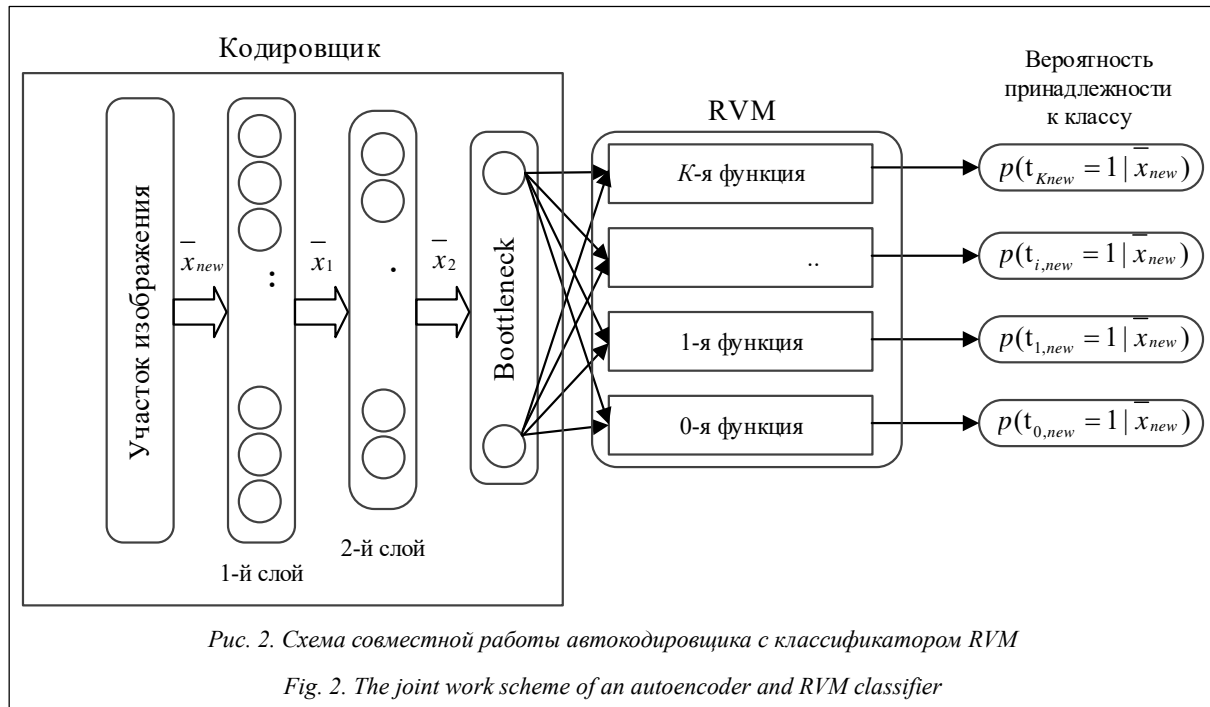
$\bar{\varphi}(\bar{x}) = (\varphi_0(\bar{x}), \dots, \varphi_M(\bar{x}))^T$ – набор базисных функций (в качестве базисной функции в данной реализации алгоритма используется гауссова радиальная базисная функция);

$\bar{w} = (w_0, \dots, w_M)^T$ – весовые коэффициенты, которые определяют вклад, вносимый определенной базисной функцией;

$\varepsilon \sim N(0, \sigma^2)$ – аддитивный гауссовский шум с нулевым математическим ожиданием.

Для обучения рассматриваемой модели используется метод максимального правдоподобия, который позволяет производить оценку весовых коэффициентов по обучающей выборке: $\bar{w}_{est} = \arg \min_w (\|\bar{y} - \bar{\Phi} \cdot \bar{w}\|)^2$,

где $\bar{y} = (y_1, \dots, y_N)^T$ – набор скалярных значений, каждое из которых соответствует определенному вектору \bar{x}_n из обучающей выборки $\bar{x} = (\bar{x}_1, \dots, \bar{x}_N)^T$;



$\bar{\Phi} = (\bar{\varphi}(x_1), \dots, \bar{\varphi}(x_N))^T$ – матрица, сформированная из значений базисных функций, вычисленных для обучающей выборки. Для адаптации RVM к задаче бинарной классификации в статье [14] рассматривается функция правдоподобия, основанная на распределении Бернулли:

$$p(\bar{t} | \bar{w}) = \prod_{i=1}^N \sigma \left\{ \left(y(\bar{x}_i) \right) \right\}^{t_i} \cdot \left[1 - \sigma \left\{ \left(y(\bar{x}_i) \right) \right\} \right]^{1-t_i}, \quad (2)$$

где $y(\bar{x}) = \sigma(f(\bar{x}, \bar{w})) = \frac{1}{1 + \exp(-\bar{w}^T \cdot \bar{\varphi}(\bar{x}))}$ – логистическая функция; $(\bar{t}, \bar{x}) = \{t_n, \bar{x}_n\}_{n=1}^N$ – обучающая выборка; $t_n = t(\bar{x}_n) \in \{0, 1\}$ – элементы вектора \bar{t} .

Стандартный подход к решению задачи мультиклассовой классификации, рассматриваемой в данной статье, – мультиномиальная логистическая регрессия $p(t_{j, new} = 1 | \bar{x}_{new}) = \frac{\exp(y_j(\bar{x}_{new}))}{\sum_{k=1}^K \exp(y_k(\bar{x}_{new}))}$, где

$$y_j(\bar{x}) = \bar{w}^T \cdot \bar{\varphi}(\bar{x}); j=0, \dots, K. \text{ Тогда аналогичным выражению (2) образом вводим следующую функцию правдоподобия:}$$

$$p(\bar{t} | \bar{w}) = \prod_{i=1}^N \left(\frac{\exp(\bar{w}_{til}^T \cdot \bar{\varphi}_{til}(\bar{x}_i))}{\sum_{k=0}^K \exp(\bar{w}_k^T \cdot \bar{\varphi}_k(\bar{x}_i))} \right),$$

где til – метка класса i -го образца. Для оценки апостериорного максимума выведенной функции

правдоподобия используем технику, описанную в статье [14]. При этом делаем допущение о распределении весовых коэффициентов $p(\bar{w} | \bar{\alpha}) = \prod_{i=0}^M N(w_i | 0, \alpha_i^{-1})$, где $\bar{\alpha} = (\alpha_0, \dots, \alpha_M)^T$ – вектор гиперпараметров, и шума измерения $\bar{\varepsilon} \sim N(0, \beta^{-1} \cdot \bar{I})$, а также вводим эквивалентную объектную функцию $J(\bar{w}) = \sum_{i=1}^N \log p(t_i | \bar{w}_{til}) + \sum_{i=1}^N \log p(\bar{w}_{til} | \alpha_{til}^*)$ [14].

Реализация процесса сканирования изображения

Процесс детектирования объектов определенного класса разбит на несколько фаз. Во время первой фазы фотоснимок сканируется скользящим окном, которое перемещается по регулярной сетке, обеспечивая 40 %-ное перекрытие между соседними сканируемыми участками фотоснимка.

После первой, предварительной, фазы детектирования производится отбор участков фотоснимка, для которых предсказанное RVM-значение вероятности $p(t_k = 1 | \bar{x})$ того, что данный участок фотоснимка содержит изображение детектируемого объекта определенного класса k , больше некоторого заданного порогового p_{thrMax} значения. Для каждого отобранного участка вычисляется положение центрального пикселя, которому назначается метка класса k , выявленная модулем RVM. Далее для всех пикселей, соседних с центральным пикселем, вычисляется принадлежность к классу объектов с меткой k . Вычисление принадлежности рас-

смастриваемых пикселей к данному классу k производится на основании значения вероятности, которое вычисляется путем интерполяции методом natural neighbor [15]. Интерполяция значений вероятности $p(t_k = 1 | \bar{x})$ принадлежности к данному классу объектов k производится на основе данных о значении $p(t_k = 1 | \bar{x})$ для центральных участков пикселей, соседних с рассматриваемым участком. Если значение $p(t_k = 1 | \bar{x})$ для пикселя, принадлежность которого к классу объектов k выясняется, больше некоторого заданного порогового p_{thrMin} значения, данному пикселю назначается метка класса объектов k . Если выясняется, что исследуемый пиксель принадлежит к классу объектов k , производится выявление принадлежности или непринадлежности к классу объектов k соседних с ним пикселей, принадлежность которых еще не выявлена. В противном случае, если исследуемый пиксель не принадлежит к классу объектов k , процедура выявления принадлежности соседних с ним пикселей не производится.

В результате описанной процедуры формируется множество пикселей, для которых вычисляется минимальный ограничивающий прямоугольник, определяющий границы изображения детектируемого объекта.

Экспериментальная часть

Для оценки качества детектирования объектов, имеющих низкое разрешение, и целесообразности использования разработанного алгоритма был проведен эксперимент, в котором для обучения детектора и оценки качества его работы использовался полученный с веб-ресурса набор изображений аэрофотосъемки HRO_2012_6_Inch_Orthophotography [16]. Изображения из набора имеют разрешение $12,5 \times 12,5$ см на пиксель.

Из данного набора были сформированы два набора изображений, один из которых содержит изображения транспортных средств, имеющих размеры примерно 15×35 пикселей. Второй набор изображений, полученный из исходного набора, содержит различные изображения земной поверхности без изображений транспортных средств.

Обучение детектора и оценка качества детектирования производились на персональном компьютере, обладающем следующими характеристиками: центральный процессор – Intel Core i7-6700K; оперативная память – 8 Гб; видеопроцессор – NVIDIA GEFORCE GTX 980 Ti; операционная система – Ubuntu 14.04 LTS. Для измерения качества детектирования использовались следующие показатели: MAP (mean average precision); 0.01 FPPI (false positive per image); 0.1 FPPI; 1 FPPI. Сравнение качества детектирования разработан-

ного алгоритма производилось с DPM (deformable part model) [17], детектором HOG+SVM [18], детектором RVM, работающим в связке со стандартным автокодировщиком.

Результаты оценки работы алгоритмов приведены в таблице.

Результаты эксперимента

Experiment results

Детектор	MAP	0.01 FPPI	0.1 FPPI	1 FPPI
DPM	62.7±4.3	14.4±7.1	34.5±6.1	76.5±5.2
HOG+SVM	58.9±4.0	12.9±6.2	32.1±4.8	74.8±4.7
Standard Autoencoder+RVM	59.8±3.9	13.9±5.6	32.9±4.2	75.1±4.6
Discriminative Autoencoder+RVM	79.8±4.2	31.5±5.2	58.0±4.6	89.5±2.5

Проведенный эксперимент показал, что RVM в связке со стандартным автокодировщиком работает хуже DPM, но лучше детектора HOG+SVM. Если вместо стандартного автокодировщика использовать дискриминативный автокодировщик, качество детектирования значительно возрастет, то есть данный алгоритм превзойдет остальные алгоритмы, включая алгоритм DPM. Таким образом, результаты исследования доказывают целесообразность использования предложенного алгоритма.

Выводы

В результате проведенного исследования был предложен алгоритм детектирования изображений объектов определенного класса на фотоснимках, полученных с видекамеры при малом разрешении изображений детектируемых объектов и низком качестве фотоснимков. В процессе создания алгоритма выбрана архитектура и разработана схема обучения автокодировщика, предназначенного для формирования признакового описания объектов по их изображениям.

В работе предложена оригинальная функция потерь для обучения автокодировщика дополнительному функционалу, а именно способности извлекать информацию только о заданной категории объектов.

Для детектирования объектов использована схема скользящего окна, которое сканирует фотоснимок, обеспечивая перекрытие между соседними сканируемыми участками фотоснимка в 40 %. С целью повышения точности детектирования границ изображения объекта применен мультиномиальный классификатор, построенный на основе алгоритма RVM, для вычисления вероятности нахождения изображения объекта или его частей в сканируемом участке фотоснимка. Также для по-

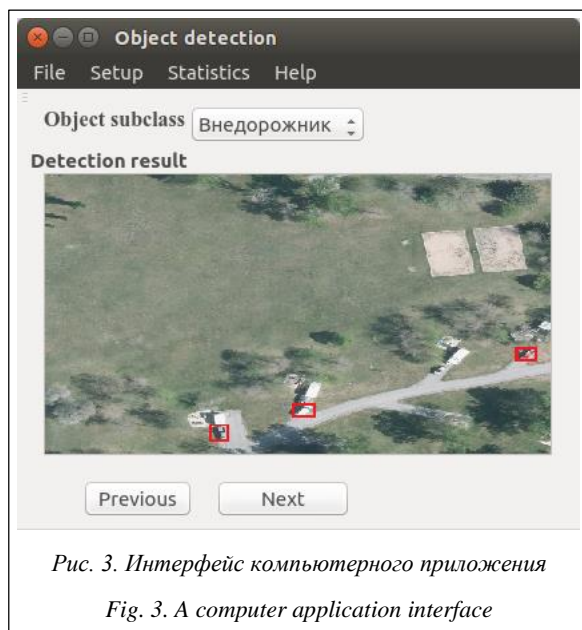


Рис. 3. Интерфейс компьютерного приложения

Fig. 3. A computer application interface

вышения точности обнаружения границ изображения детектируемого объекта было предложено интерполировать значения вероятности обнаружения детектируемого объекта определенного класса для каждого пикселя изображения, анализ принадлежности которого к изображению объекта проводится. После чего на основании распределения пикселей, которые были приняты за принадлежащие к изображению объекта определенного класса, производится оценка границ изображения детектируемого объекта. На основании результатов проведенного исследования было разработано компьютерное приложение (рис. 3).

Результаты данного исследования могут найти широкое применение для автоматизации различных процессов, например, для сбора и анализа информации в различных аналитических системах.

Литература

1. Mikolajczyk K., Schmid C. An affine invariant interest point detector. Proc. 7th ECCV'02, 2002, part I, pp. 128–142.
2. Harris C.G., Stephens M. A combined corner and edge detector. Proc. of 4th Alvey Vision Conf. 1988, pp. 147–151.

3. Lowe D.G. Distinctive image features from scale-invariant keypoints. Intern. Jour., of Comp. Vision. 2004, vol. 60, iss. 2, pp. 91–110.
4. Sivic J., Zisserman A. Video Google: a text retrieval approach to object matching in videos. Proc. ICCV'03. 2003, vol. 2, pp. 1470–1477.
5. Schaffalitzky F., Zisserman A. Automated location matching in movies. Comp. Vision and Image Understanding. 2003, vol. 92, pp. 236–264.
6. Fergus R., Fei-Fei L., Perona P., Zisserman A. Learning object categories from Google's image search. Proc. Intern. Conf. on Comp. Vision. 2005, vol. 2, pp. 1816–1823.
7. Krizhevsky A., Hinton G.E. Using very deep autoencoders for content-based image retrieval. Conf. ESANN 2011, 19th Europ. Sympos. on Artificial Neural Networks, Bruges, Belgium, 2011, pp. 44–51.
8. Tipping M.E. Sparse bayesian learning and the relevance vector machine. The Jour. of Machine Learning Research. 2001, vol. 1, pp. 211–244.
9. Kamyshanska H., Memisevic R. On autoencoder scoring. ICML (3), vol. 28 of JMLR Workshop and Conf. Proc. 2013, pp. 720–728.
10. Wenchao Yu, Guangxiang Zeng, Ping Luo, Fuzhen Zhuang, Qing He, Zhongzhi Shi. Embedding with autoencoder regularization. Machine Learning and Knowledge Discovery in Databases Europ. Conf., ECML PKDD 2013, Prague, Czech Republic, 2013, Proc., part III. Berlin Heidelberg: Springer, 2013, pp. 208–223.
11. Razakarivony S., Discriminative F.J. Autoencoders for small targets detection. ICPR '14 Proc. 22nd Intern. Conf. on Pattern Recognition. 2014, pp. 3528–3533.
12. Yoonseop Kang, Kang-Tae Lee, Jihyun Eun, Sung Eun Park, Seungjin Choi. Stacked denoising autoencoders for face pose normalization. Neural Information Processing 20th Intern. Conf., ICONIP 2013, Daegu, Korea, Proc., part III. Berlin Heidelberg: Springer, 2013, pp. 241–248.
13. Kin Gwn Lore, Adedotun Akintayo, Soumik Sarkar. LLNet: a deep autoencoder approach to natural low-light image enhancement. Cornell Univ. Library. URL: <https://arxiv.org/abs/1511.03995> (дата обращения: 05.2016).
14. Tzikas D.G., Liyang Wei, Likas A., Yongyi Yang, Galatsanos N.P. A tutorial on relevance vector machines for regression and classification with applications. EURASIP Newsletter, 2012, vol. 17, no. 2, pp. 4–23.
15. Enriquez-Cervantes C.J., Rodriguez-Dagnino R.M. A super-resolution image reconstruction using natural neighbor interpolation. Computación y Sistemas, 2015, vol. 19, no. 2, pp. 211–231.
16. UTAH AGRC. URL: <https://gis.utah.gov/data/aerial-photography/> (дата обращения: 01.2016).
17. Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, vol. 32, iss. 9, pp. 1627–1645.
18. Elhoseiny M., Bakry A., Elgammal A. MultiClass object classification in video surveillance systems – experimental study. Proc. 2013 IEEE CVPRW '13, 2013, pp. 788–793.

OBJECT DETECTION ALGORITHM IN LOW IMAGE QUALITY PHOTOGRAPHS

A.S. Viktorov¹, Postgraduate Student, alsevictor@mail.ru

¹ Kostroma State University, Dzerzhinskogo St. 17, Kostroma, 156005, Russian Federation

Abstract. The article considers a set of algorithms for specified class object recognition in low quality photographs obtained via camera with low resolution.

A special feature of the considered method of object detection is the ability to detect objects even if their sizes in images don't exceed several tens of pixels. Each processed image is scanned via sliding window of fixed width and height that reads rectangular image regions with specified overlap between neighboring regions. All scanned image regions are preliminarily processed by a discriminative autoencoder to extract feature vector from a processed image region. Further analysis of an extracted vector includes classifier means on the basis of probabilistic multinomial regression model to check the scanned region of image if there is object image or its parts. The classifier calculates the probability of detection of a certain class detectable object in each scanned image region. On the basis of an image scan result there is a conclusion on the object image presence and its most probable position in the photograph.

To improve the accuracy of calculation of detected object image boundaries, the value of a detection probability of a certain detectable object is interpolated for each pixel, which is analyzed for belonging to the image of the object. After that, on the basis of the detected pixel distribution on the image it is possible to estimate the boundaries of the detected object.

The experiment has revealed that using a discriminative autoencoder significantly increases detection algorithm robustness. The article also gives a detailed description of a learning and algorithm parameters adjustment process. The results of this research can be widely used to automate various processes, for example, to collect and analyze information in various analytical systems.

Keywords: stacked denoising autoencoders, relevance vector machine, neural net, loss function, training data set, feature vector, object detection, likelihood function.

References

1. Mikolajczyk K., Schmid C. An affine invariant interest point detector. *Proc. 7th European Conf. on Computer Vision (ECCV '02)*. Part 1, 2002, pp. 128–142.
2. Harris C.G., Stephens M. A combined corner and edge detector. *Proc. 4th Alvey Vision Conf.* 1988, pp. 147–151.
3. Lowe D.G. Distinctive image features from scale-invariant keypoints. *Int. Jour. of Computer Vision*. 2004, vol. 60, iss. 2, pp. 91–110.
4. Sivic J., Zisserman A. Video Google: a text retrieval approach to object matching in videos. *Proc. of 9th IEEE Int. Conf. on Computer Vision (ICCV '03)*. 2003, vol. 2, pp. 1470–1477.
5. Schaffalitzky F., Zisserman A. Automated location matching in movies. *Computer Vision and Image Understanding. Special iss. on video retrieval and summarization*. 2003, vol. 92, pp. 236–264.
6. Fergus R., Fei-Fei L., Perona P., Zisserman A. Learning object categories from google's image search. *Proc. Int. Conf. on Computer Vision*. 2005, vol. 2, pp. 1816–1823.
7. Krizhevsky G.E. Hinton using very deep autoencoders for content-based image retrieval. *Proc. Conf. ESANN 2011, 19th European Symp. on Artificial Neural Networks*. 2011, Bruges, Belgium, pp. 44–51.
8. Tipping M.E. Sparse bayesian learning and the relevance vector machine. *Jour. of Machine Learning Research*. 2001, vol. 1, pp. 211–244.
9. Kamyshanska H., Memisevic R. On autoencoder scoring. *Proc. ICML (3), vol. 28 of JMLR Workshop and Conf.* 2013, pp. 720–728.
10. Wenchao Yu, Guangxiang Zeng, Ping Luo, Fuzhen Zhuang, Qing He, Zhongzhi Shi. *Proc. Embedding with Autoencoder Regularization. Machine Learning and Knowledge Discovery in Databases European Conf. (ECML PKDD 2013)*, Berlin Heidelberg, Springer Publ., 2013, part III, pp. 208–223.
11. Razakarivony S., Discriminative F.J. Autoencoders for small targets detection. *Proc. 22nd Int. Conf. on Pattern Recognition (ICPR '14)*. 2014, pp. 3528–3533.
12. Kang Y.-S., Lee K.-T., Eun J.-H., Park S.-E., Choi S.-J. Stacked denoising autoencoders for face pose normalization. *Proc. Neural Information Processing 20th Int. Conf. (ICONIP 2013)*, Berlin Heidelberg, Springer Publ., 2013, part III, pp. 241–248.
13. Kin Gwn Lore, Adedotun Akintayo, Soumik Sarkar. *LLNet: a deep autoencoder approach to natural low-light image enhancement*. Cornell Univ. Library. Available at: <https://arxiv.org/abs/1511.03995> (accessed May 5, 2016).
14. Tzikas D.G., Wei L., Likas A., Yang Y., Galatsanos N.P. *A tutorial on relevance vector machines for regression and classification with applications*. Dept. of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, USA, 2012, no. 6, pp. 4–23.
15. Enriquez-Cervantes C.J., Rodriguez-Dagnino R.M. *A super-resolution image reconstruction using natural neighbor interpolation*. Monterrey, Mexico, Tecnologico de Monterrey, Electrical and Computing Engineering Dept., 2015, pp. 211–231.
16. *UTAH AGRC*. Available at: <https://gis.utah.gov/data/aerial-photography/> (accessed January 16, 2016).
17. Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 2010, vol. 32, iss. 9, pp. 1627–1645.
18. Elhoseiny M., Bakry A., Elgammal A. MultiClass Object classification in video surveillance systems – experimental study. *Proc. 2013 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW '13)*. 2013, pp. 788–793.