



Н.Г. Аксак¹, А.А. Бессонов², И.В. Новосельцев³, О.Г. Руденко⁴

¹ ХНУРЭ, г. Харьков, Украина, nataliia.axak@nure.ua

² ХНЭУ, г. Харьков, Украина, oleksandr.bezsonov@hneu.net;

³ ХНУРЭ, г. Харьков, Украина, i.v.novoseltsev@gmail.com;

⁴ ХНЭУ, г. Харьков, Украина, oleg.rudenko@hneu.net

РАСПОЗНАВАНИЕ ИЗМЕНЕНИЯ РАЗМЕРА И ЦВЕТА ИЗОБРАЖЕНИЯ НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

В статье рассмотрено применение сверточной нейронной сети Mask R-CNN для распознавания изменения размера и цвета изображения. Применение данной сети оправдано тем, что в отличие от других типов сверточных нейронных сетей данная сеть предназначена для не только для обнаружения всех объектов указанных классов и определения охватывающей рамки для каждого из них, но и для определения пикселей, принадлежащих каждому объекту каждого класса по отдельности. Описана архитектура этой сети. Предложена модификация алгоритма обучения многослойного перцептрона, входящего в состав Mask R-CNN, представляющая собой матричные варианты процедуры Качмажа (Уидроу-Хоффа). Вследствие наличия ограниченных помех предлагается использовать в алгоритме зону нечувствительности, для которой приведены процедуры настройки. Проведено экспериментальное исследование эффективности работы сети в задаче распознавания заболевания кожи — меланомы. В качестве обучающей выборки использовался набор фотографий из конкурса «ISIC 2016: Skin Lesion Analysis Towards Melanoma Detection». Полученные результаты свидетельствуют об эффективности развиваемого подхода.

РАСПОЗНАВАНИЕ ОБРАЗОВ, КЛАССИФИКАЦИЯ, СВЕРТОЧНАЯ ИСКУССТВЕННАЯ НЕЙРОННАЯ СЕТЬ, АЛГОРИТМ ОБУЧЕНИЯ, АКТИВАЦИОННАЯ ФУНКЦИЯ

Аксак Н.Г., Бессонов О.О., Новосельцев И.В., Руденко О.Г. Розпізнавання зміни розміру і кольору зображення на основі згорткової нейронної мережі. У статті розглянуто застосування згорткової нейронної мережі Mask R-CNN для розпізнавання зміни розміру і кольору зображення. Застосування даної мережі виправдано тим, що на відміну від інших типів згортальних нейронних мереж дана мережа призначена для не лише для виявлення всіх об'єктів заданих класів і визначення охоплюючої рамки для кожного з них, а й для визначення пікселів, що належать кожному об'єкту кожного класу окремо. Описана архітектура цієї мережі. Запропоновано модифікацію алгоритму навчання багатосарового перцептрона, що входить до складу Mask R-CNN, яка представляє собою матричні варіанти процедури Качмажа (Уїдроу-Хоффа). Внаслідок наявності обмежених перешкод пропонується використовувати в алгоритмі зону нечутливості, для якої наведені процедури налаштування. Проведено експериментальне дослідження ефективності роботи мережі в задачі розпізнавання захворювання шкіри - меланоми. В якості навчальної вибірки використовався набір фотографій з конкурсу «ISIC 2016: Skin Lesion Analysis Towards Melanoma Detection». Отримані результати свідчать про ефективність підходу, що розвивається.

РОЗПІЗНАВАННЯ ОБРАЗІВ, КЛАСИФІКАЦІЯ, ЗГОРТКОВА ШТУЧНА НЕЙРОННА МЕРЕЖА, АЛГОРИТМ НАВЧАННЯ, АКТИВАЦІЙНА ФУНКЦІЯ

Aksak N.G. Bezsonov O.O., Novoseltsev I.V., Rudenko O.G. Recognition of image resizing and color based on a convolutional neural network. The article describes the use of the Mask R-CNN convolutional neural network for recognition of changes in image size and color. The use of this network is justified by the fact that, unlike other types of convolutional neural networks, this network is designed not only to detect all objects of the indicated classes and determine the covering frame for each of them, but also to determine the pixels belonging to each object of each class separately. The architecture of this network is described. A modification of the learning algorithm of the multilayer perceptron, which is part of the Mask R-CNN, is a matrix version of the Kacmage (Widrow-Hoff) procedure. Due to the presence of limited interference, it is proposed to use the dead zone in the algorithm, for which tuning procedures are given. An experimental study of the effectiveness of the network in the task of recognizing skin disease - melanoma was conducted. A set of photographs from the competition "ISIC 2016: Skin Lesion Analysis Towards Melanoma Detection" was used as a training sample. The obtained results testify to the effectiveness of the developed approach.

RECOGNITION OF IMAGES, CLASSIFICATION, CLASSIFICATION, ARTIFICIAL NEURAL NETWORK, TRAINING ALGORITHM, ACTIVATION FUNCTION

Введение

На сегодняшний день одним из самых интенсивно развивающихся научных и технологических направлений является обработка и анализ изображений.

Среди важнейших задач, связанных с обработкой изображений и которые относят к задачам компьютерного зрения, следует отметить:

- **Classification** — классификация изображения по типу объекта, которое оно содержит.

- **Semantic segmentation** — определение всех пикселей объектов определённого класса или фона на изображении. Если несколько объектов одного класса перекрываются, их пиксели никак не отделяются друг от друга.

- **Object detection** — обнаружение всех объектов указанных классов и определение охватывающей рамки для каждого из них.

- **Instance segmentation** — определение пикселей, принадлежащих каждому объекту каждого класса по отдельности.

За последние годы был представлен ряд методов и алгоритмов, применяемых для решения данных задач, среди одними из наиболее эффективных являются искусственные нейронные сети (ИНС) [1]. Использование персептрона, радиально-базисных сетей, автоэнкодера, неокогнитрона, вероятностных сетей оказалось весьма эффективным при решении широкого круга задач. Однако новый толчок к интересу исследований в данной области дало появление в 1998 г. нового типа сетей – сверточных ИНС (СИНС).

Сверточная ИНС (Convolution Neural Network, CNN) впервые была предложена в [2,3] как развитие модели неокогнитрона, предназначенного для эффективного распознавания изображений. Впоследствии на основе СНС были построены сети R-CNN (Regions With CNNs) для применения CNN к задаче object detection. R-CNN создает ограничивающие рамки для каждого объекта на изображении или предложения регионов, используя процесс выборочного поиска. Fast R-CNN, увеличившая производительность R-CNN, осуществляет классификацию объектов каждого региона вместе с более жесткими ограничивающими рамками. Следующая сеть Faster R-CNN улучшила механизм генерации используемых в ней регионов-кандидатов за счет вычисления регионов не по изначальному изображению, а по карте признаков, полученных из CNN. Для этого был добавлен модуль под названием Region Proposal Network (RPN). Наконец, сеть Mask R-CNN развивает архитектуру Faster R-CNN путём добавления ещё одной ветки, которая предсказывает положение маски, покрывающей найденный объект, и, таким образом решает уже задачу instance segmentation. При получении изображения сеть выдает объекты (bbox), ограничивающие рамки, классы (class) и маски (mask).

Следует отметить, что Mask R-CNN является наиболее быстродействующей сетью на данный момент.

1. Структура сверточной нейронной сети

Первоначально структура сверточной нейронной сети создавалась с учетом особенностей строения некоторых частей человеческого мозга, отвечающих за зрение. В основу разработки таких сетей заложено три механизма:

- локальное восприятие;
- формирование слоев в виде набора карт признаков (разделяемые веса);
- субдискретизация (подвыборка).

Под локальным восприятием подразумевается, что на вход нейрона поступает не все изображение, а только некоторая его часть. Это позволяет сохранять конфигурацию изображения при переходе от слоя к слою.

Идея разделяемых весов подразумевает, что к большому числу связей применяется небольшой набор весов, т. е. каждая область изображения, на которые оно разделено, будет обработана одним и тем же набором весов. При таком искусственно созданном ограничении весов улучшается свойство сети к обобщению.

СНС состоит из слоев свертки, субдискретизации (подвыборки) и слоев полносвязной нейронной сети.

2. Слои сверточной нейронной сети

CNN получили свое название от оператора «свертки». Основная цель свертки в случае CNN — извлечь элементы из входного изображения. Свертка сохраняет пространственные отношения между пикселями, изучая особенности изображения, используя маленькие квадраты входных данных.

Каждый нейрон в плоскости сверточного слоя получает свои входы от некоторой области предыдущего слоя (локальное рецептивное поле), то есть входное изображение предыдущего слоя сканируется небольшим окном и пропускается сквозь набор весов, а результат отображается на соответствующий нейрон сверточного слоя.

Подвыборочный слой уменьшает масштаб плоскостей путём локального усреднения значений выходов нейронов. Таким образом, достигается иерархическая организация. Последующие слои извлекают более общие характеристики, меньше зависящие от искажений изображения.

За каждым сверточным слоем следует слой субдискретизации (подвыборки), или вычислительный слой, который производит уменьшение размерности изображения путём локального усреднения значений выходов нейронов (рис. 1).

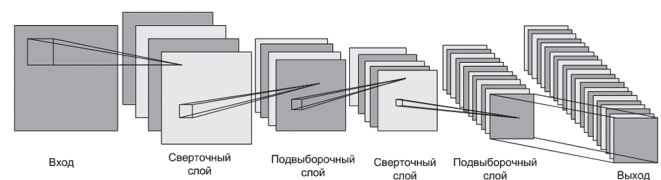


Рис. 1. Структура сверточной нейронной сети

В архитектуре сверточной сети принято считать, что наличие признака важнее информации о его расположении. Поэтому из нескольких соседних нейронов в карте признаков выбирается максимальный и его значение считается одним нейроном в карте признаков меньшей размерности.

Отличие слоя подвыборки от слоя свертки состоит в том, что в последнем области соседних нейронов перекрываются, чего не происходит в слое субдискретизации.

Таким образом, СНС строится путем чередования слоев свертки и субдискретизации. На выходе сети обычно устанавливается несколько слоев полносвязной нейронной сети, на вход которых подаются конечные карты признаков. Каждый нейрон этого слоя представляет собой персептрон, который имеет нелинейную функцию активации.

3. Методы настройки параметров сверточной нейронной сети

Для обучения сверточных нейронных сетей может применяться как стандартный метод обратного распространения ошибки, так и его различные модификации. Рассмотрим один из них.

Рассмотрим обучение многослойного персептрона (МП) с M входами и L выходами, S слоями, числом нейронов выходного слоя P , скрытого – N , все нейроны которого имеют активационную функцию вида

$$f(x) = \frac{1}{1 + e^{-\alpha x}}.$$

Введем следующие обозначения:

$$W^{*s}(k) = (W_1^{*s}(k), W_2^{*s}(k), \dots, W_p^{*s}(k))^T -$$

матрица оптимальных весов выходного слоя $P \times L$;

$$w_i^{*s}(k) = (w_{i1}^{*s}(k), w_{i2}^{*s}(k), \dots, w_{iL}^{*s}(k))^T -$$

вектор оптимальных весов i -го нейрона выходного слоя $P \times 1$;

$$W^s(k) = (W_1^s(k), W_2^s(k), \dots, W_p^s(k))^T -$$

матрица оценок весов выходного слоя МП $P \times L$;

$$w_i^s(k) = (w_{i1}^s(k), w_{i2}^s(k), \dots, w_{iL}^s(k))^T -$$

вектор оценок весов i -го нейрона выходного слоя $P \times 1$;

$$W^{*i}(k) = (W_1^{*i}(k), W_2^{*i}(k), \dots, W_p^{*i}(k))^T -$$

матрица оптимальных весов нейронов i -го скрытого слоя ($i = 1, \dots, S - 1$) $N \times P$;

$$w_i^{*i}(k) = (w_{i1}^{*i}(k), w_{i2}^{*i}(k), \dots, w_{iN}^{*i}(k))^T -$$

вектор оптимальных весов нейронов i -го скрытого слоя $N \times 1$;

$$F^s(x, w^s) = (f_1^s(w_1^{Ts}(k)x(k)), f_2^s(w_2^{Ts}(k)x(k)) \dots f_L^s(w_L^{Ts}(k)x(k)))^T$$

$$F^s(x, w^s) = (f_1^s(w_1^{Ts}(k)x(k)), f_2^s(w_2^{Ts}(k)x(k)) \dots f_L^s(w_L^{Ts}(k)x(k)))^T -$$

вектор активационных функций нейронов выходного слоя $L \times 1$;

$$\nabla f(x, w) = \text{diag}[\nabla f(w_1^T(k), x(k)), \nabla f(w_2^T(k), x(k)) \dots \nabla f(w_L^T(k), x(k))] -$$

матрица производных активационных функций слоя $L \times L$; $\nabla f(w_i^T(k), x(k))$ – первая производная активационной функции i -го нейрона.

Тогда выходные сигналы МП и ошибка аппроксимации могут быть представлены следующим образом:

– требуемый вектор выходных сигналов

$$y^*(k) = W^{*s}(k)F^s(x, w^{*s}(k)) + \xi_1(k); \quad (1)$$

– реальный вектор выходных сигналов

$$y(k) = W^s(k)F(x, w^s(k)) + \xi_2(k); \quad (2)$$

– ошибка

$$e(k) = y^*(k) - y(k) + \xi(k) = \tilde{\theta}^T(k)F(x, w(k)) + W^{*s}(k)(k)\tilde{F}(x, w(k)) + \xi_1(k) + \xi_2(k); \quad (3)$$

$$i = 1, \dots, S - 1.$$

где $\tilde{\theta}(k) = W^{*s}(k) - W^s(k)$ – матрица ошибок обучения $P \times L$; $\tilde{F}(k) = F(x, w^{*s}(k)) - F(x, w^s(k))$ – вектор ошибок активационных функций; $\xi_1(k)$, $\xi_2(k)$ – вектор ошибок выходного сигнала и нейросетевой аппроксимации соответственно.

С учетом этих обозначений градиентные процедуры обучения матриц весов выходного и i -го скрытого слоя МП, линеаризующие квадратичный функционал и представляющие собой процедуру обратного распространения ошибки, могут быть записаны следующим образом:

$$W^s(k) = W^s(k-1) - \gamma_{W^s}(k) \frac{1}{\|F(x, w^s(k-1))\|^2} \frac{\partial E(k)}{\partial W^s(k-1)} =$$

$$= W^s(k-1) + \gamma_{W^s}(k) \frac{F(x, w^s(k-1))e^T(k)}{\|F(x, w^s(k-1))\|^2}; \quad (4)$$

$$W^i(k) = W^i(k-1) -$$

$$- \gamma_{W^i}(k) \frac{1}{\|\nabla f(x(k), W^i(k-1))x(k)\|^2} \frac{\partial E(k)}{\partial W^i(k-1)} =$$

$$= W^i(k-1) + \gamma_{W^i}(k) \frac{\nabla f(x(k), w^i(k-1))e^T(k)x^T(k)}{\|\nabla f(x(k), W^i(k-1))x(k)\|^2}, \quad (5)$$

$$i = 1, \dots, S - 1,$$

т.е. представляет собой матричные варианты процедуры Качмажа (Уидроу-Хоффа).

Здесь $\gamma_{W^s}(k)$ и $\gamma_{W^i}(k)$, $i = 1, \dots, S - 1$ – коэффициенты обучения.

Для повышения вычислительной устойчивости (4), (5) они могут быть модифицированы путем введения в них параметров регуляризации, т.е.

$$W^s(k) = W^s(k-1) + \gamma_{W^s}(k) \frac{F(x, w^s(k-1))e^T(k)}{\beta_{W^s}(k) + \|F(x, w^s(k-1))\|^2}; \quad (6)$$

$$W^i(k) = W^i(k-1) + \gamma_{W^i}(k) \frac{\nabla f(x(k), w^i(k-1))e(k)x^T(k)}{\beta_{W^i}(k) + \|\nabla f(x, W^i(k-1))x(k)\|^2}, i \neq S, \quad (7)$$

где $\beta_{W^s}(k)$, $\beta_{W^i}(k), i=1, \dots, S-1$ – параметры регуляризации.

Следует, однако, отметить, что если статистические свойства помех $\xi_1(k)$ и $\xi_2(k)$ не известны, а известно, что они ограничены по амплитуде, то процедуры (4)-(7) следует модифицировать путем использования в них зон нечувствительности так, как это было сделано для случая линейной Адалины, например, следующим образом:

$$W^s(k) = W^s(k-1) + \gamma_{W^s}(k) \frac{\alpha_{W^s}(k)F(x, w^s(k-1))e^T(k)}{\beta_{W^s}(k) + \|F(x, w^s(k-1))\|^2}; \quad (8)$$

$$\text{где } \alpha_{W^s}(k) = \begin{cases} g(e(k), \Delta_{W^s}(k)), & \text{если } \|e(k)\| > \Delta_{W^s}(k); \\ 0, & \text{если } \|e(k)\| \leq \Delta_{W^s}(k). \end{cases} \quad (9)$$

Так как при вычислении оценок матрицы W используется производная функции активации, то параметр зоны нечувствительности должен содержать коэффициент α , определяющий наклон сигмоиды

$$W^i(k) = W^i(k-1) + \gamma_{W^i}(k) \frac{\alpha_{W^i}(k)\nabla f(x(k), W^s(k-1))e(k)x^T(k)}{\beta_{W^i}(k) + \|\nabla f(x, W^s(k-1))x(k)\|^2}, \quad (10)$$

где

$$\alpha_{W^i}(k) = \begin{cases} g(e(k), \Delta_{W^i}(k)), & \text{если } \frac{\nabla f_{\min}}{\alpha} \|e(k)\| > \Delta_{W^i}(k); \\ 0, & \text{если } \frac{\nabla f_{\min}}{\alpha} \|e(k)\| \leq \Delta_{W^i}(k), \end{cases} \quad (11)$$

где $\nabla f_{\min} = \min[\nabla f_1(k), \nabla f_2(k), \dots, \nabla f_L(k)] > 0$.

Настройка зон нечувствительности для процедур (8) и (10) может осуществляться следующим образом:

$$\Delta_{W^s}(k) = \Delta_{W^s}(k-1) + \frac{\alpha_{W^s}(k)\|e(k-1)\|}{\beta_{W^s}(k-1) + \|F(x, W^s(k-1))\|^2}; \quad (12)$$

$$\Delta_{W^i}(k) = \Delta_{W^i}(k-1) + \frac{\alpha_{W^i}(k)\|e(k-1)\|}{\beta_{W^i}(k-1) + \|\nabla f(x, W^i(k-1))x(k)\|^2}, i \neq S. \quad (13)$$

Эксперимент 1. Решалась задача распознавания заболевания кожи – меланомы, с помощью сверточной нейронной сети MRCNN [6]. Данная сеть является расширенной версией сети Faster RCNN, которая в свою очередь является усовершенствованной версией сети Fast RCNN.

В качестве обучающей выборки использовался набор фотографий с конкурса «ISIC 2016: Skin Lesion Analysis Towards Melanoma Detection» [7]. В данном наборе содержится 900 фотографий различных повреждений кожи и соответствующие им битовые маски, позволяющие определить точный контур повреждения. Данный набор был разбит на 600 обучающих фотографий и 300 проверочных. Примеры изображений из обучающей выборки приведены на рис. 2.

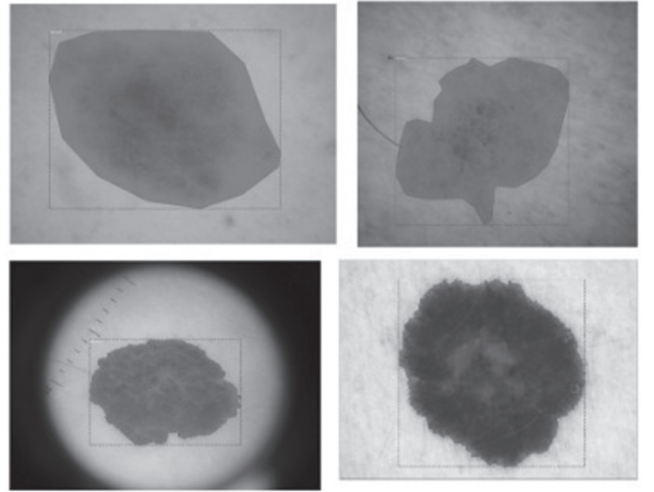


Рис. 2. Примеры изображений с наложенными масками из обучающей выборки

Обучение сети проводилось с использованием весов из предобученной на выборке “Coco” (<http://cocodataset.org>) сети resnet-101. Обучение проводилось на протяжении 30 эпох, и его результаты приведены на рис. 3. На рис. 3а показан график изменения общей ошибки обучения, состоящей из ошибки детектора «Bounding Box» – рис. 3б, ошибки классификатора 3в, ошибки предсказателя маски – 3г и ошибки модуля, предлагающего регионы, содержащие объекты для распознавания. Результаты распознавания показаны на рис. 4. Как видно из результатов, сеть MRCNN осуществляет эффективное распознавание повреждения кожи, указывая область повреждения и предлагая точную попиксельную маску. Данная сеть также позволяет отслеживать рост повреждения. Как видно на одном из фрагментов рис. 4, сеть распознала два региона – первичный и вторичный, являющийся результатом роста новообразования. Это свидетельствует о том, что предлагаемый подход позволяет отслеживать динамику развития заболевания.

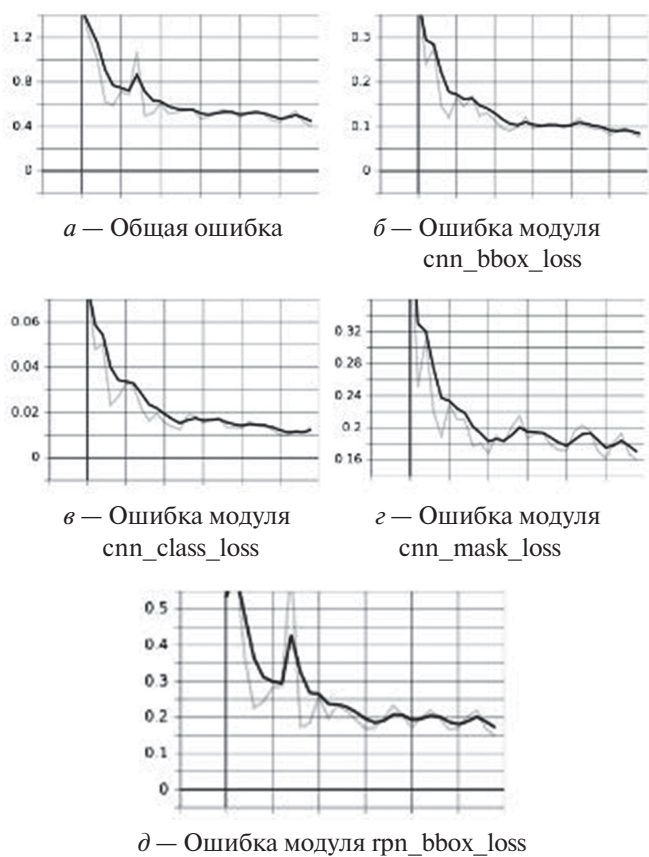


Рис. 3. Результаты обучения сети

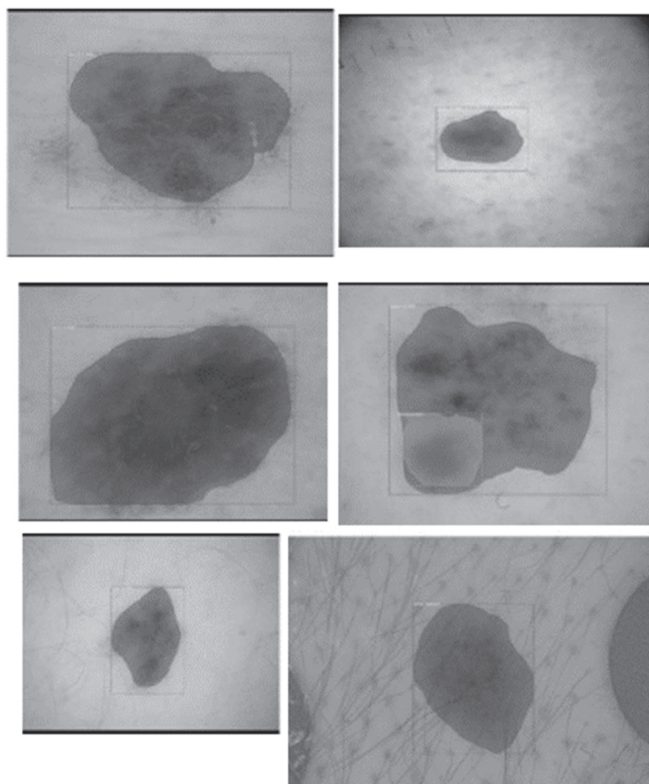


Рис. 4. Результаты распознавания

Эксперимент 2. В данном эксперименте помимо решения задачи обнаружения кожного повреждения решалась задача классификации заболевания (доброкачественное образование или

злокачественное). Для обучения использовалась база фотографий с кожными повреждениями из Эксперимента 1. К данной базе прилагалась маска, позволяющая определить границы повреждения и описание, содержащую информацию о типе заболевания. Примеры изображений, использованных для обучения представлены на рис. 5. Результаты классификации представлены на рис. 6. Зеленым цветом обозначались повреждения, которые не были распознаны сетью как злокачественные, красным — обозначены повреждения, которые с высокой вероятностью являются злокачественными. На некоторых изображениях вероятность злокачественности новообразования была близка к пороговому уровню, но не превышала его. Такие образования обозначались голубым цветом. Точность классификации после 70 эпох обучения составила 91,63%.

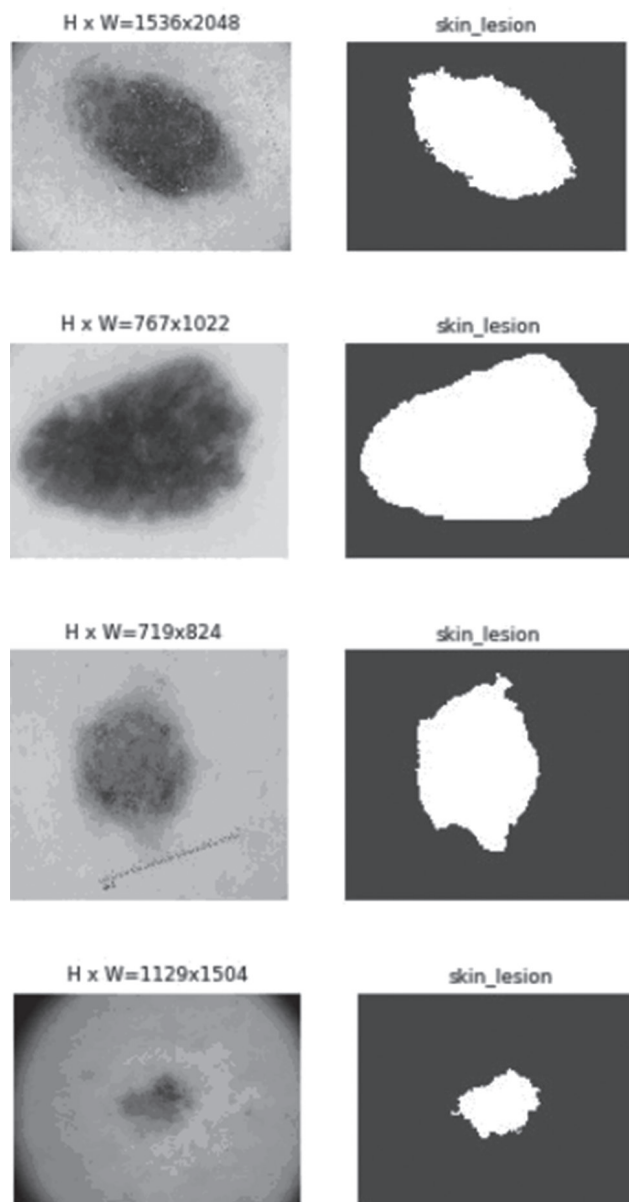


Рис. 5. Примеры изображений, использованных для обучения

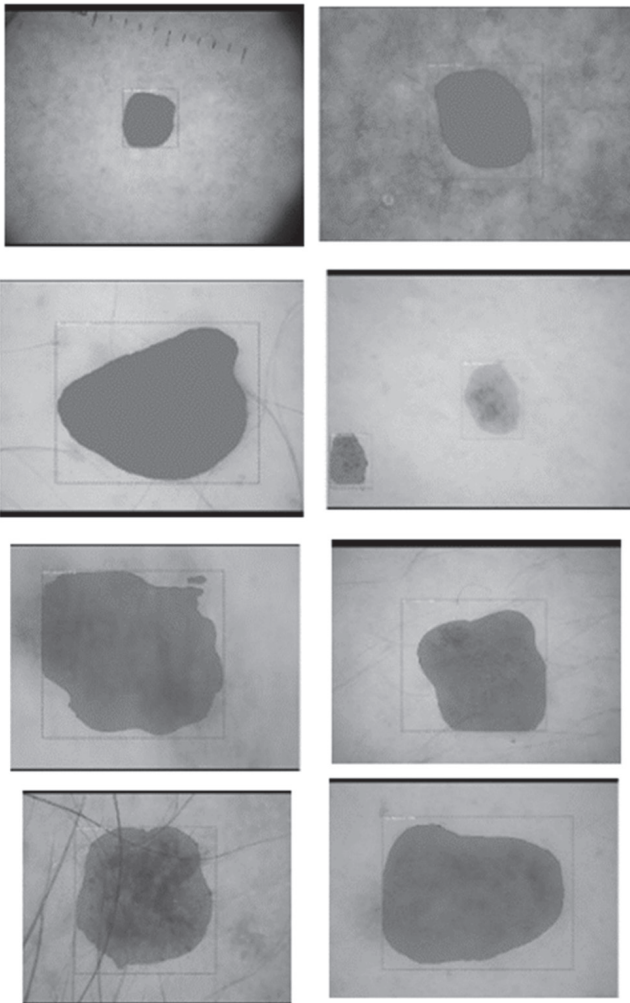


Рис. 6. Результаты классификации

Выводы

В работе рассмотрено применение сверточной нейронной сети Mask R-CNN для распознавания изменения размера и цвета изображения. Предложена модификация алгоритма обучения многослойного перцептрона, представляющая собой матричные варианты процедуры Качмажа (Уидроу-Хоффа), содержащие зону нечувствительности и приведены процедуры настройки этой зоны.

Полученные результаты проведенных экспериментов свидетельствуют об эффективности развиваемого подхода.

Список литературы:

- [1] Хайкин С. Нейронные сети: полный курс. М.: Вильямс, 2006. — 1104 с.
- [2] LeCun Y. A theoretical framework for backpropagation // Proc. of IEEE. — 1998. — P. 21-28.
- [3] LeCun Y., Bottou L., Bengio Y., Haffner P., Gradient Based Learning Applied to Document Recognition // Proc. of IEEE, 1998. V. 86, №11. — P. 2278-2324.
- [4] Hinton G.E., Osindero S., The Y.-W. A fast learning algorithm for deep belief nets. Neural Comput., 18(7):1527-1554, 2006.
- [5] Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. // In Advances in Neural Information Processing Systems, 2012. — P. 1097-1105.
- [6] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN <https://arxiv.org/abs/1703.06870>
- [7] ISIC 2016: Skin Lesion Analysis Towards Melanoma Detection <https://challenge.kitware.com/#phase/566744dccad3a56fac786787>

Поступила в редколлегию 05.12.2018