MACHINE LEARNING TO PROCESS IMAGES BY INCREASING THE RESOLUTION OF LOW QUALITY IMAGES

Mukhamadiyeva K.B. (Republic of Uzbekistan)
Email: Mukhamadiyeva517@scientifictext.ru

Mukhamadiyeva Kibriyo Bahodirovna - Applicant, Senior Lecturer,
DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES,
BUKHARA ENGINEERING TECHNOLOGICAL INSTITUTE,
BUKHARA, REPUBLIC OF UZBEKISTAN

Abstract: *artificial neural networks (ANNs) are currently undergoing their second birth, which is primarily due to the increased computing power of modern computers and the emergence of extra-large data sets for training, present in global networks. Solutions in the field of data classification, image segmentation, decision support, which are comparable in quality and often exceed the results obtained from classical methods of image recognition, are developed on the basis of the ANN.*

The applied field of research is related to solving problems of segmentation, classification and categorization of identification numbers of wagons that do not have a template and standard, based on the use of convolution neural networks (CNN). Recently, the research in this field has been actively conducted [1].

The work is devoted to the development of algorithms and architecture of a convolution neural network (CNN) to solve the problems of segmentation, classification and categorization of images of identification numbers of wagons.

Keywords: *image recognition, neural network, ANN, CNN, OpenCV, pixel filtr, Softmax function.*

МАШИННОЕ ОБУЧЕНИЕ ОБРАБОТКИ ИЗОБРАЖЕНИЙ ПУТЕМ ПОВЫШЕНИЯ РАЗРЕШЕНИЯ ИЗОБРАЖЕНИЙ НИЗКОГО КАЧЕСТВА

Мухамадиева К.Б. (Республика Узбекистан)

Мухамадиева Кибриё Баходировна - соискатель, старший преподаватель,
кафедра информационных и коммуникационных технологий,
Бухарский инженерно-технологический институт,
г. Бухара, Республика Узбекистан

Аннотация: *искусственные нейронные сети (ИНС) в настоящее время переходят в основное направление развития, что, в первую очередь, связано с увеличением вычислительной мощности современных компьютеров и появлением сверхбольших наборов данных для обучения, присутствующих в глобальных сетях. На базе ИНС разрабатываются*

решения в области классификации данных, сегментации изображений, поддержки принятия решений, сопоставимые по качеству и зачастую превосходящие результаты, полученные с помощью классических методов распознавания образов.
Прикладная область исследований связана с решением задач сегментации, классификации и категоризации идентификационных номеров вагонов, не имеющих шаблона и стандарта, основанных на использовании сверточных нейронных сетей (CNN). В последнее время исследования в этой области активно проводятся [1].
Работа посвящена разработке алгоритмов и архитектуры сверточной нейронной сети (CNN) для решения задач сегментации, классификации и категоризации изображений идентификационных номеров вагонов.
Ключевые слова: *распознавание образов, нейронная сеть, ИНН, CNN, OpenCV, фильтр пикселей, функция Softmax.*

Introduction .

The article is devoted to the description of the algorithm of construction and artificial increase of the marked data set on images in the conditions of small samples. The algorithm, by applying artificial geometric transformations to the initial objects, provides creation of new training examples, which, in the conditions of small samples, allows to better train the neural network and reduce retraining. The following methods of artificial data set magnification were used in the work: brightness enhancement of green colour, application of Gaussian blur filter, averaging of image pixels with the help of normalized pixel filter, image rotation relative to its center with different angles of rotation, image cropping and halving of dimensions, resizing of the central part of the image to 680*200 pixels according to CNN requirements using cubic interpolation. Algorithm of image data set construction and artificial enlargement was implemented by means of Python programming language with the use of NumPy, SciPy, etc. libraries.

The second stage describes the development of a new CNN architecture for accurate classification of the marked data set on three-channel color images of ultra-high spatial resolution in small samples. The overall architecture of the CNN consists of six convolution blocks (each includes one convolution layer). The first and third bundle blocks include pooling layers (a function to reduce the dimensional space of feature maps) with a maximum function. At the end of the CNN are two fully connected layers and one output layer. The last four bundle blocks use the ReLU activation function (1) and the output layer uses the Softmax activation function (2):

$$f\ (x) = \max(0, x), \qquad (1)$$

$$z\ =\ \omega^T x\ - \theta, \qquad (2)$$

where x is a vector-column of features of an object of Mx1 dimension, ω^T- where K is the number of object classes, and M is the number of object features.

To control the retraining of the network, it was decided to use the method of regularization for the artificial neural network Dropout, with the help of which the reduction of the model complexity was achieved, while keeping the number of its parameters at a low level. The optimal value of the regularization factor was also selected. Based on the results of the numerical experiments, it was determined that the optimal values are 0.25 after the second, fourth and fifth layers, and 0.5 before the output layer. As a target function, which should be minimized during training of the neural network, was chosen categorical cross-entropic loss between input data and the actual classification of images. This function is well suited for calculating the probability of the original image belonging to a certain category [2-3]. To date, a fairly large number of optimization algorithms have been proposed for the calculation of gradient descent in neural networks. The proposed model used ADAM optimization (adaptive moment estimation), since it is the most suitable optimizer for the problem under consideration, in particular, due to the possibility of initial calibration of an artificial neural network.

The structure of the new CNN can be formally presented in the following form:

- The input image takes into account the two-dimensional topology and consists of several matrices, where each matrix corresponds to an image of a specific color channel: red, blue and green. The input information of each specific pixel value is normalized in the range from 0 to 1 by the formula (3):

$$f(p, min, max) = \frac{p-min}{max-min} \qquad (3)$$

where *f* is the normalization function, *p* is the value of a specific pixel from 0 to 255, *min* is the minimum pixel value - 0, *max* is the maximum pixel value - 255 [4].

During CNN training, the values of each convolution layer matrix are equal to 0, and the kernel weights are set randomly in the range from -0.5 to 0.5. The kernel passes through the previous matrix pixel by pixel and performs the convolution operation by the following formula (4):

$$(f * g)[m, n] = \sum_{k,J} f[m-k, n-l] * g[k, l], \qquad (4)$$

where f is the initial image sensor, g is the convolution core, *m* is the width of the initial image sensor, *n* is the height of the initial image sensor, *k* is the width of the core, *l* is the height of the core.

Depending on the method of processing the edges of the initial image matrix at the convolution stage, the result may be smaller than the initial matrix, of the same size or larger than the formula (5):

$$x^l = f\,(x^{l-1} * \mathrm{k}^l + b^l) \qquad (5)$$

where x^l - layer output *l*, *f ( )* - activation function, *b* - layer shift coefficient *l*, * - input x convolution operation with *k* kernel.

Thus, as a result of edge effects, the original output matrix is reduced and has the following form (6):

$$x_j^l = f\left(\sum_i x_i^{l-1} * k_j^l + b_j^l\right), \qquad (6)$$

where x_i^{l-1} - feature map j (output of layer l), k_j^l- kernel of convolution j of the map, layer l .- The sub-sample layer of the developed CNN can be described by the following formula (7):

$$x^l = f\left(a^l * subsample\left(\mathbf{x}^{l-1}\right) + b^l\right), \qquad \mathbf{(7)}$$

where a^l, b^l - layer shift coefficient l, *subsample()* - operation of sampling local maximum values.

The last type of CNN layer is a full-connected layer or an ordinary multilayer perseptron. The purpose of the layer is classification, where a complex nonlinear function is modeled, and the optimization of which improves the recognition quality (8):

$$x_j^l = f(\sum_i x_j^{l-1} w_{i,j}^{l-1} + b_j^{l-1}, \qquad (8)$$

where $w_{j,j}^{l-1}$ - layer l weighting matrix.

The next step is a new method of control sampling formation on three-channel color images of ultra-high spatial resolution in order to accelerate the process of preparation of the data set under test. The structure of the proposed method includes a set of consecutive elements:

Conversion of a three-channel color image into a color palette of gray shades [5];

Blurring of the image in grayscale using a high-frequency Gaussian filter to reduce noise in images with the following function parameters: core size 11*11 pixels with a standard deviation of 0;

Create a binary image (black-and-white) from blurred images in grayscale by applying the threshold function with an optimal threshold of 100 pixels brightness of the input image;

Structuring the outlines of image elements by applying two consecutive functions (erosion and expansion) with multiple iterations for binary images to distinguish the outlines of individual characters and minimize the effect of their fusion in a single object;

Detection of image segments, which was implemented using the Green formula-based contour area calculation function.

$$S = \iint_R dxdy = \frac{1}{2}\oint_C xdy - ydx \qquad (9)$$

where S is the area of area R bounded by contour C. Symbol -$\oint_C$ indicates that the contour of the potential object is closed, and bypassing the integration along this contour is done in a counterclockwise direction. The object size for the function has been set in the range from 150*50 to 680*200 pixels according to the original image sizes of the training sample.

The use of the proposed method of control sampling allows to prepare the tested data set for a relatively short time than in analogues in a mode close to automatic. In this case, a significant acceleration of the process of preparing a test data set is achieved. Specific time characteristics vary depending on the input image size and computer configuration. Under the conditions of the experiments, it took on average about 5 minutes to form a control sample. The method of control sampling formation was implemented by means of Python programming language using OpenCV, Pillow, etc. libraries.

The next stage was a modified algorithm of image segmentation in small samples based on ResNet50 and ResNet101 CNN for building a mask of object regions. The algorithm was modified by using an artificially increased data set and fine-tuning the algorithm to solve the task at hand. A two-step approach has been implemented in this CNN. At the first stage, the input image is scanned and proposals (areas that may contain an object) are generated. At the second stage, the proposal is classified and bounding frames and masks are formed.

The R-CNN mask consists of the following structural components:

The base is the ResNet101 standard CNN, which detects low-level objects (edges and corners) on its earlier layers and detects higher-level elements (tree, man, building) on its later layers. Passing through the neural network, the image is transformed from a 1920x1080px ×3 (RGB) matrix into a 32x32x2048 map of objects. This feature map becomes an input for the following network levels. To improve the quality of object extraction by the network, the Feature Pyramid Network method was used, which takes high level CNN elements detected and transmits them to the lower layers. This allows functions at each level to access both lower and higher level functions [6-8].

2. A CNN that scans the image in a sliding window mode and finds areas containing objects. As a result, the network generates: an anchor class (foreground and background, where the first one implies the presence of a classified object) and a limiting frame with delta estimation (percentage relative to coordinate, width, and height accuracy) to specify the anchor field for better correspondence to the object [7-8].

3. Classifier and restriction window regressor. This component generates two outputs for each object: object class and bounding frame. This frame is necessary for further specification of the object location and size [7].

Since classifiers do not always handle object size determination, they usually require objects of a fixed size. For this purpose, an additional subsample layer (sub-sampling) was used, which allows to crop a part of the object map to a fixed size.

4. Segment masks are CNNs that generate masks of objects selected by the classifier from the previous step. Generated masks have low resolution: 28x28 pixels. During training, reduced object masks of up to 28x28 pixels are used to calculate losses, and during output, the predicted masks are increased to the size of the limiting object frame, which is the final mask, one per object [7].

The next step is a description of the efficiency evaluation metrics of the developed CNN architecture, with such calculated characteristics as: Accuracy (10), Precision (11), Completeness (Recall) (12) and F_score (13):

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FN+FP}} \qquad (10)$$

where the number of true positive predictions (True Positive, TP), true negative predictions (True Negative, TN), false positive predictions (False Positive, FP) and false negative predictions (False Negative, FN),

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad (11)$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad (12)$$

$$F\_score = 2x\frac{\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \qquad (13)$$

Results

Experimental application of developed CNN architecture.

After preparation of a test data set for independent testing of the new CNN architecture, the main parameters of the evaluation metrics of the developed CNN architecture were calculated using various data sets of ultra-high spatial resolution images. The calculations revealed that the use of artificial increase in the test data set for CNN training resulted in a 14.5%, 20.4%, and 34.3% improvement in the proportion of correct responses (accuracy), accuracy (precision), and F-meter (F_score), respectively.

References / Список литературы

1. *Abkar A.-A., Sharifi M.A., Mulder N.J.* Likelihood-based image segmentation and classification: a framework for the integration of expert knowledge in image classification procedures. Int. J. Appl. Earth Obs. Geoinformation 2, 2000. Pp. 104–119.
2. *Bahrampour S., Ramakrishnan N., Schott L., Shah M.* Comparative study of deep learning software frameworks. ArXiv Prepr. In: ArXiv151106435, 2015.
3. *Barbedo J.G.A.* Digital image processing techniques for detecting. quantifying and classifying plant diseases. SpringerPlus 2, 2013. Pp. 660.
4. *Davies H.G., Bowman C., Luby S.P.* Cholera – management and prevention. J. Infect., Hot Topics in Infection and Immunity in Children 74 (17). Pp. 30194–30199, 2017. [Electronic Resource]. URL: https://doi.org/10.1016/S0163-4453. S. 66–S73/ (date of access: 31.08.2020).
5. *El Hatri C., Boumhidi J.* Fuzzy deep learning based urban traffic incident detection. Cogn. Syst. Res., 2017. [Electronic Resource]. URL: https://doi.org/10.1016/j.cogsys.2017.12.002/ (date of access: 31.08.2020).
6. *Grinblat G.L., Uzal L.C., Larese M.G., Granitto P.M.* Deep learning for plant identification using vein morphological patterns. Comput. Electron. Agric., 2016, Pp. 418–424.
7. *Guo Y., Liu Y., Oerlemans A., Lao S., Wu S., Lew M.S.* Deep learning for visual understanding: A review. Neurocomputing 2016. Pp. 27–48.
8. *Boukaye Boubacar Traoré, Bernard Kamsu-Foguem, Fana Tangara*. Deep convolution neural network for image recognition, 2018.10.02. Pp. 3-14.