УДК 004.931

Существующие способы распознавания образов

Нурпейсова Ардак Алданышовна
доктор PhD, старший преподаватель кафедры «Информационные системы»Казахский агротехнический университет им. С. Сейфуллина
e-mail: naa11317@mail.ru
Республика Казахстан, г. Астана

Nurpeisova A. A.PhD, senior lecturer of the Department "Information Systems" S. Seifullin Kazakh Agrotechnical University
e-mail: naa11317@mail.ru
Republic of Kazakhstan, Astana

Шаушенова Анаргуль Гимрановна
к.т.н., старший преподаватель кафедры «Информационные системы»
Казахский агротехнический университет им. С. Сейфуллина
e-mail: shaushenova_78@mail.ru
Республика Казахстан, г. Астана

Shaushenova A. G., candidate of Technical Sciences,
senior lecturer of the Department "Information Systems"
S. Seifullin Kazakh Agrotechnical University
e-mail: shaushenova_78@mail.ru
Republic of Kazakhstan, Astana

Смаилова Лайла Кудайбергеновна
старший преподаватель кафедры «Информационные системы»
Казахский агротехнический университет им. С. Сейфуллина
e-mail: smailova_laila@mail.ru
Республика Казахстан, г. Астана

Smailova L.K., senior lecturer of the Department "Information Systems" S. Seifullin Kazakh Agrotechnical University
e-mail: smailova_laila@mail.ru
Republic of Kazakhstan, Astana

Муталова Жазира Саткановна
магистр, старший преподаватель Институт экономики, информационных технологий и профессионального образования,
e-mail: zhazira77@mail.ru
Республика Казахстан, г. Уральск

Mutalova Zh.S., senior lecturer, Master's degree
Institute of Economics, Information Technology and Vocational Education
e-mail: zhazira77@mail.ru
Republic of Kazakhstan, Uralsk

Аннотация. Машинное обучение широко применяется в компьютерном зрении и обработке изображений. Уровень развития машинного обучения и наличие разнообразных комплексных моделей могут позволить с достаточной точностью обнаруживать на изображении людей, что обеспечит возможность их подсчета. Видеопоток с заведений может быть получен с систем видеонаблюдения, которых с каждым годом становится все больше. Точность модели будет являться одним из основных показателей, на основе которого можно сделать вывод о пригодности модели для ее практической эксплуатации. До 2014 года технология распознавания образов основывалась на традиционных алгоритмах; после 2014 года технология распознавания образов в основном основывалась на алгоритмах глубокого обучения. Традиционные методы основаны на характеристиках моделей, созданных вручную, таких как описания текстур краев, и на методах, включающих методы машинного обучения, такие как анализ основных компонентов, линейный дискриминантный анализ и методы опорных векторов. Трудность, которую необходимо преодолеть из-за функциональности этого проекта, заключается в том, чтобы приспособить различную изменчивость к неограниченным средам, что заставило исследователей сосредоточиться на поиске конкретных методов для каждого типа изменчивости, таких как контроль неизменности возраста, неизменность позы, контроль одинаковых условий освещения и так далее. Сверточные нейронные сети являются наиболее распространенным алгоритмом глубокого обучения, применяющим несколько сверхточных слоев и вычислений. Они предоставляют эффективные способы извлечения признаков, а также являются лучшим выбором для решения проблем обнаружения объектов.

Ключевые слова: распознавание лиц, эксперты по распознаванию лиц, искусственный интеллект, машинное распознавание и машинное обучение, обнаружение образов.

Existing methods of pattern recognition

Annotation. Machine learning is widely used in computer vision and image processing. The level of development of machine learning and the availability of a variety of complex models can allow detecting people in the image with sufficient accuracy, which will make it possible to count them. The video stream from the establishments can be obtained from video surveillance systems, which are becoming more and more every year. The accuracy of the model will be one of the main indicators, on the basis of which it can be concluded that the model is suitable for its practical operation. Until 2014, pattern recognition technology was based on traditional algorithms; after 2014, pattern recognition technology was mainly based on deep learning algorithms. Traditional methods rely on handcrafted model characteristics such as edge texture descriptions and methods that incorporate machine learning techniques such as Principal Component Analysis, Linear Discriminant Analysis, and Support Vector Machines. The difficulty that needs to be overcome due to the functionality of this project is to accommodate different variability to unrestricted environments, which led the researchers to focus on finding specific methods for each type of variability, such as control for age invariance, posture invariance, control for uniform lighting conditions, and so on. Convolutional Neural Networks are the most common deep learning algorithm, applying multiple ultra-precise layers and calculations. They provide efficient ways to extract features, and are also the best choice for solving object detection problems.

Keywords: face recognition, face recognition experts, artificial intelligence, machine recognition and machine learning, pattern detection.

More than 83% of the information received by a person comes through vision. The eyes play an important role in human activities for obtaining information, adapting to the environment and changing the world, and are also the most important sense organs. But in the ever-changing world of nature, the ability to perceive the human eye is very limited. With the expansion of human activity and the increasingly rapid pace of life and production, it is not enough to rely on the human eye alone to obtain information to meet needs. Machine vision appeared to better understand and change the world.

Computer vision is an area of artificial intelligence that allows computers and systems to obtain meaningful information from images, videos and other visual data and take actions or provide recommendations based on this information. If artificial intelligence gives computers the ability to think, then computer vision gives them the ability to detect, observe and understand. The advantage of the human visual system is that it can train the ability to distinguish objects, the distance to the object, the movement of the object or not, as well as the presence of image problems in the appropriate environment.

Computer vision trains machines to perform these functions, but they rely on cameras, data and algorithms to complete their work in a shorter period of time, unlike humans who rely on the retina, optic nerve and visual cortex of the brain. Systems trained to inspect products or monitor production assets can analyze thousands of products or processes every minute and detect extremely subtle defects or problems, so computer vision capabilities quickly surpass human ones. Computer vision is widely used in many industries, such as energy, utilities, manufacturing and automotive, and the market continues to expand. It is expected that by 2022 that market capitalization will reach 48.6 billion US dollars [1].

Object detection refers to a technique that allows a computer to find multiple specific objects in an image by writing specific algorithm codes. Target detection contains two values:

- determine which target objects are in the image and solve the problem of the existence of target objects;

- Determine the specific position of the target object in the image and solve the problem of where the target object is.

The biggest difference between target detection and image classification is that target detection requires more detailed evaluation not only to determine if a target is enabled, but also to determine the specific location of each target. Despite the wide variety of presented algorithms, it is possible to single out the general structure of the face recognition process.

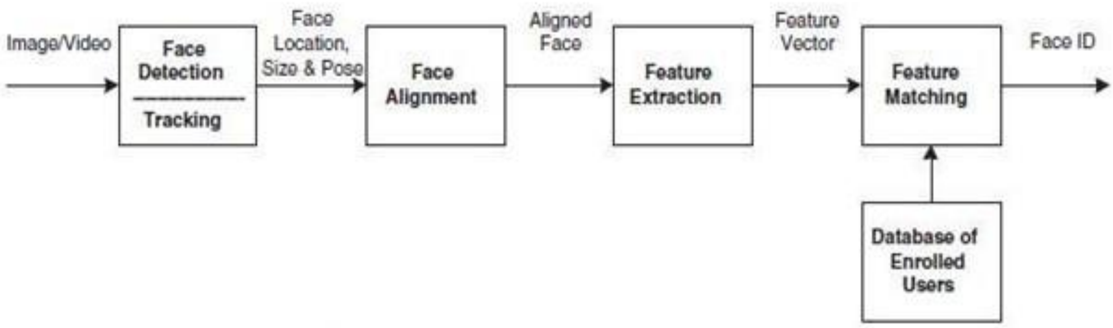

Figure 1. The general process of processing the face image during recognition

At the first stage, the face is detected and localized in the image. At the stage of recognition, the face image is aligned (geometric and brightness), the features are calculated, and the recognition itself is performed - the calculated features are compared with the standards stored in the database. The main difference between all the presented algorithms will be the calculation of features and comparison of their sets with each other (Figure 1) [2].

Over the past two decades, target detection technology can be divided into two stages: before 2014, target detection tasks were based on traditional algorithms, after 2014, target detection tasks were mainly based on deep learning algorithms. During the traditional algorithm stage, the detectors rely entirely on hand-designed functions. Among them, the emergence of multiple detectors has had an important impact on the development of object detection technology.

Traditional algorithms

(1) Viola Jones algorithms

The Viola-Jones algorithm was proposed by P. Viola and M. Jones for face detection scenarios [3]. With the same accuracy of the algorithm, the Viola-Jones detector works tens and hundreds of times faster than other algorithms over the same period of time. The Viola-Jones detector uses the most direct sliding window method, and the detection frame goes through all scales and positions in the image to see if the detection frame contains a target face. This sliding window seems simple, but requires a lot of computing time. The advantage of the Viola-Jones detector is that it uses the strategies of integral imaging, feature screening and cascade detection, which significantly increases the speed of the algorithm.

(2) HOG algorithm

The HOG (Histogram of Oriented Gradients) function, an important scale-independent function, was proposed by N. Dalal and B. Triggs in 2005 [4]. The basic idea is to use a regular grid to divide the image into equally sized sub-blocks and compute a gradient-oriented histogram in each sub-block. In this way, zoom and lighting effects can be largely eliminated. For quite a long time, HOG has been an important feature to solve detection problems, especially in pedestrian detection scenarios, HOG has an extremely important application.

(3) The DPM algorithm

DPM (Deformable Part Model, Deformable Component Model) is a component—based detection algorithm proposed by P. Felsenswalb in 2008 [5], and later R. Girshik [6] made many important improvements to it. DPM extends the classical HOG function at the function level, and also uses the sliding window method for SVM-based classification, Its main idea is to divide the target to be detected into a number of parts and transform the problem of detecting a complex target into a problem of detecting several simple parts. For example, the detection of a car is converted to the detection of windows, bodywork and wheels, respectively. Although the performance of detection algorithms far exceeds that of DPM, the idea of task separation adopted in DPM has had a profound impact on the development of subsequent detection algorithms. Now many detection tasks are being developed and solved based on this idea.

Deep Learning Algorithms

(1) R-CNN algorithm

In 2014, R. Girshik et al. proposed the R-CNN algorithm [7]. The idea behind the R-CNN algorithm is very simple. First, several candidate fields are selected from the original image based on a selective search method, and then a fixed image zoom scale in each candidate field is sent to the convolutional network for feature extraction. Features are classified using the support vector machine (SVM). In the VOC2007 dataset, the R-CNN algorithm has greatly improved performance (from 33.7% to 58.5%) over the previous discovery algorithm, which is a major breakthrough in the discovery algorithm.

(2) SPP-Net algorithm

In 2014, He Yumin and others proposed the SPP-Net algorithm (Spatial pyramid unification networks, Spatial pyramid Unification Network). The SPP-Net algorithm is based on the spatial pyramid union layer (SPP level), which discretizes image fragments into fixed-size images regardless of the size of the input image fragment. In the process of using SPP-Net, only one convolutional network calculation for the source image is required for detection. After receiving the feature map of the entire image, each region of the candidate frame (ROI) is divided into subimages of the same size through the spatial pyramid pooling layer, and each subimage of the same size is sent to the subsequent network for feature extraction. The extracted features have the same dimension and are finally sent to a fully connected layer for classification. SPP-Net does not require multiple convolutional network calculations, compared to R-CNN, the algorithm speed is significantly improved by 20 times, with the algorithm accuracy unchanged [8].

(3) Faster R-CNN algorithm

In 2015, S. Ren et al. proposed the Faster R-N algorithm, which is the first end-to-end algorithm and the first target detection algorithm close to deep learning in real time. Using the ZF-Net network skeleton, on the VOC 2007 dataset, the mAP reached 73.2%, and the algorithm speed reached 17 frames per second. The main contribution of Faster-CNN is to detect block proposals using convolutional networks. Detection of candidate blocks, prediction of target categories, regression of target position offsets and sharing of basic convolutional functions — the entire process of the algorithm is integrated into a complete end-to-end learning structure. A faster RCN overcomes a bottleneck in the detection rate of candidate frames and is a very important two-step algorithm [9].

(4) The FPN algorithm

In 2017, Y. Lin et al. proposed the FPN algorithm (Feature Pyramid Networks, functional pyramid strategy) based on Faster R-CNN [10]. Prior to FPN, most deep learning detectors were calculated based on the topmost feature map of the convolutional network. Deep functions contain global information, but weaken detailed information Using deep functions has a big disadvantage when detecting small targets, especially with precise positioning. The FPN uses a top-down structure and a side-join method to combine deep and surface functions, so that the functions contain both global information and detailed information. In addition, the detection calculation is also based on a multilayer feature map of the feature pyramid, which has a stronger multiscale adaptability. The Faster R-CNN algorithm based on FPN showed the best performance in the COCO dataset. Currently, FPN has become the basic strategy for building detection algorithms.

(5) The YOLO algorithm

In 2015, R.Joseph et al. proposed the YOLO (You Look Only Once) algorithm, which is the first one-step algorithm in the field of deep learning. As the name suggests, there are no two steps in YOLO to extract candidate frames and check the classification in a two-step algorithm, the entire prediction process can be completed by sending an image to a neural network. The implementation of the YOLO algorithm is to first divide the original image into a grid, and then regress the probability of the category and the coordinates of the target position based on each grid cell. One of the biggest advantages of YOLO as a one-step algorithm is its speed: on the VOC2007 dataset, the mAP is 63.4%, and the speed of the detection algorithm can reach 45 frames per second, the accelerated version of YOLO has a map of 52.7%, and the speed can reach an astounding 155 frames per second (Figure 2) [11].

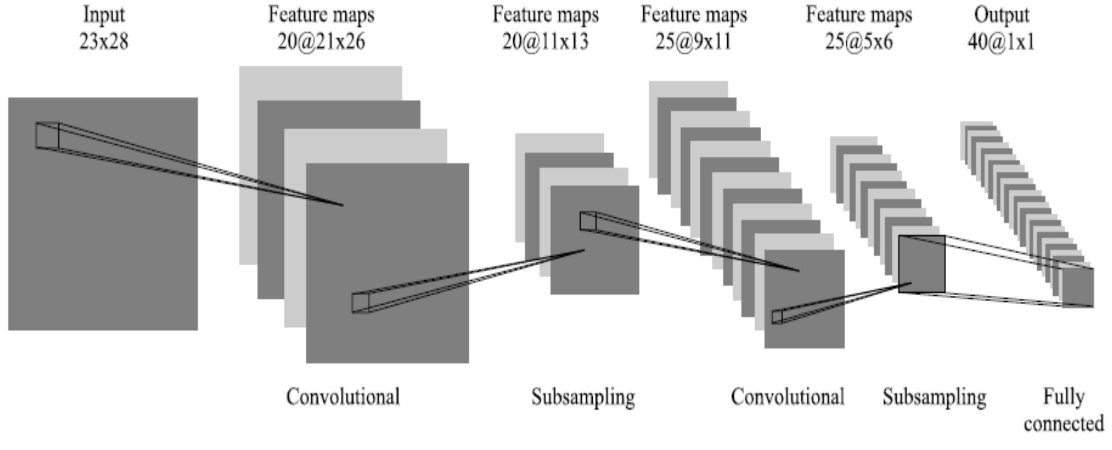

Figure 2. Schematic representation of the convolutional neural network architecture [11]

An important method is to conduct an assessment on the same publicly available data set and approximate determination of the pros and cons of the algorithm based on the performance indicators obtained as a result of the assessment [12]. Databases play a vital role in the field of algorithms. Table 1 shows several publicly available recognition evaluation databases that have a great impact in the field of object detection.

Table 1. Databases for evaluating face recognition, pedestrian detection, text detection

Database	Date	Description
		Database for evaluating facial recognition
FDDB	2010	Yahoo data contains 2,800 images and 5,000 faces. Covers include various postures and occlusions. Link to the resource: http://vis-www.cs.umass.edu/fddb/index.html
AFLW	2011	The Flickr data contains 22,000 images and 26,000 faces covering the key points of the face. Link to the resource: https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/aflw/
IJB	2015	Contains 50,000 images, including two detection and recognition tasks. Link to the resource:https://www.nist.gov/programs-projects/face-challenges
Wider Face	2016	One of the largest facial recognition assessment kits, comprising 32,000 images, 394,000 faces, include various poses and occlusions. Link to the resource: http://mmlab.ie.cuhk.edu.hk/projects/WIDERFace/
UFDD	2018	Contains 6,000 images and 11,000 test samples covering scenes such as motion blur and focus blur. Link to the resource: http://www.ufdd.info/
		A database for assessing pedestrian detection
MIT Ped	2000	A database for evaluating pedestrian detection, including 500 training images and 200 test images, link to the resource: http://cbcl.mit.edu/software-datasets/PedestrianData.html
INRIA	2005	A base for assessing the early detection of pedestrians, a link to the resource: http://pascal.inrialpes.fr/data/human/
Caltech	2009	The pedestrian detection assessment database contains 190,000 training and 160,000 test samples, link to the resource: http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/

KITTY	2012	The well-known estimated base for traffic analysis contains data on 100,000 pedestrians, link to the resource: http://www.cvlibs.net/datasets/kitti/index.php
City Persons	2017	Based on the CityScapes database, it contains 19,000 training and 11,000 test samples. Link to the resource: https://www.cityscapes-dataset.com/
EuroCity	2018	A large-scale database for pedestrian detection, data comes from 31 cities in 12 European countries, including 47,000 images, for a total of 238,000 object instances. Link to the resource: https://eurocity-dataset.tudelft.nl/
A base for evaluating text detection		
ICDAR	2003	An early publicly available database for text detection. Link to the resource: http://rrc.cvc.uab.es/
STV	2010	Данные из Google Street View, содержащие 350 изображений, 720 текстовых экземпляров. Ссылка на ресурс: http://tc11.cvc.uab.es/datasets/SVT_1
MSRA-TD500	2012	Contains 500 indoor and outdoor images with text in English and Chinese. Link to the resource: http://www.iapr-tc11.org/mediawiki/index.php/
IIIT5k	2012	Contains 1100 images, 5000 words from street view and digital images only. Link to the resource: http://cvit.iiit.ac.in/projects/SceneTextUnderstanding/IIIT5K.html
COCOText	2016	A set for evaluating large-scale text detection based on MS-COCO, including 63,000 images and 173,000 annotated texts. Link to the resource: https://bgshih.github.io/cocotext/

When choosing a single public dataset, it is impossible to be sure that it was not used when training or configuring the algorithm. In this case, the accuracy of the algorithm will be overestimated. The probability of this event can be reduced by comparing the results on different datasets.

References

1. Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2009). pp. 807–816.

2. Abbas H.H., Ahmed B.Z., Abbas A.K. 2019. 3D Face Factorisation for Face Recognition Using Pattern Recognition Algorithms. Cybernetics and

Information Technologies. Volume 19, Issue 2. Page 28-37. DOI10.2478/cait-2019-0013

3. P. Viola and M.J. Jones, «Rapid Object Detection using a Boosted Cascade of Simple Features», proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), 2001

4. Dalal N., Triggs B. Histograms of oriented gradients for human detection. Conference on Computer Vision and Pattern Recognition (CVPR), 2005. P. 886-893.

5. Felzenszwalb, P.F., Girshick, R., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. PAMI (2010)

6. Ross Girshick. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision. 2015.

7. He, K. Mask r-cnn / K. He, G. Gkioxari, P. Dollar, R. Girshick // IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October 2017.

8. Shao W., Chen Z.C., Lu G.B., Tu X.K., Fang Y.C. 2018. Sketch Synthesized Face Recognition with Deep Learning Models. Lecture Notes in Artificial Intelligence. Volume 10996. Page 387-398. DOI10.1007/978-3-319-97909-0_42

9. S. Ren, K. He, R. Girshick, and J. Sun, «Faster R-CNN: Towards real-time object detection with region proposal networks,» in Neural Information Processing Systems (NIPS), 2015.

10. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick. Microsoft COCO: Com- ´ mon objects in context. In ECCV, 2014.

11. Redmon J. YOLO9000: Better, Faster, Stronger / J. Redmon, A. Farhadi // arXiv:1612.08242v1. – 2016. – 9 p.

12. Kim J., Yu S., Lee S. 2014. Random-Profiles-Based 3D Face Recognition System. Sensors. Volume 14. Issue 4. Page 6279-6301. DOI10.3390/s140406279

This research was funded by project No. AP09259657 “Research and development of an automated proctoring system for monitoring students’ knowledge in the context of distance learning” under program 217 "Development of science", under program 102 "Grant financing of scientific research".