ISSN 0002-306X. Proc. of the RA NAS and NPUA Ser. of tech. sc. 2018. V. LXXI, N4.

UDC 004.457

INFORMATION SYSTEMS, ELECTRONICS AND SCIENTIFIC INSTRUMENTATION

L.M. HOVSEPYAN

IMAGE RECOGNITION APPROACH FOR MOBILE PLATFORMS

Image recognition, object detection and many more very complex problems can be solved by deep learning algorithms. However, those algorithms are extremely computation intensive and can be carried out by powerful general-purpose GPUs. Despite the rise of semiconductor industry, there is still very limited computational capacity on mobile devices, hence, most algorithmic solutions that have been very successful on desktop computers and servers cannot be directly deployed on them. However, based on experiments, we see that depthwise neural network models can work efficiently enough for mobile devices.

Keywords: deep learning, image recognition, convolutional neural network, depthwise convolutions.

Introduction. Currently, AI is advancing rapidly and deep learning is one of the contributors to that. Deep learning is a sub-field of machine learning, dealing with algorithms inspired by the structure and function of the brain called artificial neural networks [1]. Those algorithms are similar to how nervous system is passing information through the structure, where each neuron is connected to the other. Deep learning models work in layers, and a typical model has at least three layers, where each layer accepts information from previous and passes to the next one. It finds complex structures in large datasets by using backpropagation algorithms [2], to indicate how a machine needs to change its internal parameters that are used to calculate the representation in each layer from the representation in the previous layer. In contrast to other known learning algorithms, which stop improving after a saturation point, deep learning increases its performance with the increase of the data amount.

The trends to make recognition accuracy higher [3, 4], made networks deeper and more complicated, therefore, recognition tasks became hard to carry out in a timely fashion on computationally limited platforms. Intensive calculations often overheat mobile devices and drain their battery, which makes the usage of deep learning, with convolutional neural networks, inefficient. For this purpose, in this article, an efficient network architecture for mobile platform and a comparison with convolutional neural network will be discussed.

Convolutional and depthwise neural networks. A convolutional neural network (CNN) is a class of networks that uses convolutional layers to filter the inputs and detect valuable information. This kind of networks consist of an input layer, an output layer and one or more hidden layers. Convolutional layers have weight and bias values, which are modified in the learning process, to extract the most possible useful information from the input. Convolutional networks have become very popular since AlexNext [5], when was clearly demonstrated, how CNN is capable of achieving record-breaking results on a highly challenging dataset of 1.2 million high-resolution images. Also their experiments showed that the depth of the network is really important for achieving good results. By removing just one layer, the performance dropped significantly. Another good use-case is demonstrated in [6], where deep convolutional networks are used for the model of the face verification system. Fig. 1 illustrates how standard cnvolutional filters work.

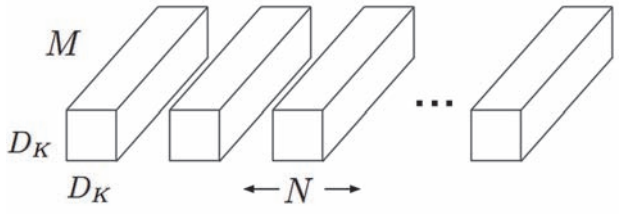

Fig. 1. Standard convolutional filters

Besides convolutional networks, in deep learning algorithms separable convolutional networks can be used. Those networks are actively discussed in [7] and [8]. They consist of depthwise and pointwise [9] convolutions. Depthwise convolution is a form of factorized standard convolution. It applies a single filter to each input channel. A standard convolution both filters and combines inputs with a new set of outputs in one step, while depthwise separable convolution splits this into two layers for filtering and for combining. The factorization radically reduces the computation size. Fig 2 illustrates how factorized filters work. Depthwise networks serve as the base for MobileNets [8], which are a set of efficient convolutional neural networks.

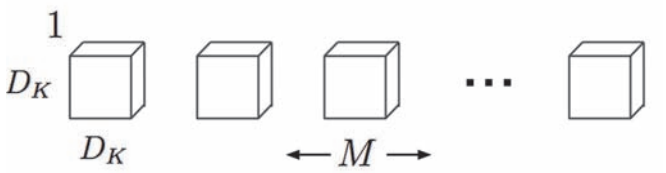

Fig. 2. Depthwise convolutional filters

Comparison Experiment. As discussed above, convolution neural networks are used to detect fixed-size features in an image, and for each desired feature it has a separate filter. The most common filter sizes are - $2x2$, $3x3$, $4x4$ and $5x5$. In the process of feature extraction, the input data of an image is being separated into three color channels - red, green and blue. After, each filter is operating on all color channels, hence, a $2x2$filter will have $2x2x3$ operations for an RGB image. The diagram shown in Fig 3, illustrates how filters are applied on all color channels in convolutional neural networks.

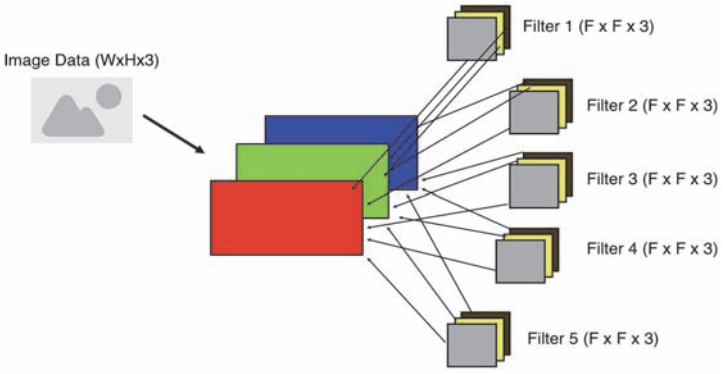

Fig. 3. A standard convolutional network

Based on the above mentioned, the general computational cost of a standard convolutional network model can be calculated with the formula (1):

$$Cost = F * F * N * C * W_0 * H_0, \quad (1)$$

where the parameters are: $F = Filter\ Size$,
$N = Number\ of\ filters$,
$C = Number\ of\ color\ channels$,
$W_0 = Image\ output\ width$,
$H_0 = Image\ output\ height$.

In the example illustrated in Fig 3, the input image has parameters $WxHx3$, which run into a convolutional network with 5 filters, to detect 5 distinct features. In real applications the number of filters depends on a certain problem and may be a few dozens, which will dramatically increase calculations, so let us take 128 filters for this example. Each of the filters has a size of FxF size, so it will effectively require $FxFx3$ flops. And finally, let's take the input image with a size of $112x112$x3. Based on formula (1):

$$Cost = 3 * 3 * 128 * 112 * 112 * 3 \approx 43.5\ million\ flops,$$

Now let's calculate the cost of depthwise convolutional networks. As we discussed above, depthwise convolutional networks operate in a similar way, but with some differences. In depthwise networks, each filter operates on a single channel and the number of filters is equal to the number of image input channels. These two differences imply that every filter operates on each channel separately. Thus, an image with three channels will need three filters, and each filter will have an effective size of FxF. The diagram shown in Fig 4, illustrates how depthwise convolutional networks work. By considering those differences, the general cost of depthwise neural networks can be calculated with the formula (2):

$$Cost = F * F * C * W_0 * H_0. \quad (2)$$

By applying the same $112x112$x3 sized image, the cost becomes:

$$Cost = 3 * 3 * 3 * 112 * 112 \approx 0.34\, million\ flops.$$

This basic experiment shows, that in comparison to convolutional neural network models, depthwise neural network models have an astonishing difference, and can be applied to solve image processing tasks on low computational power platforms like mobile phones and embedded controllers.

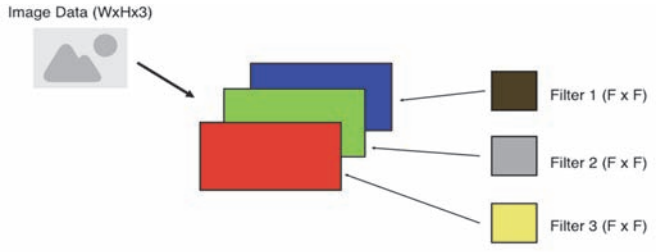

Fig. 4. A depthwise convolutional network

Conclusion. In this paper, depthwise neural networks are considered as a preferable solution for mobile platforms, and the deep learning tasks. A comparative calculation has been made, and the results have proved that the number of calculation operations in depthwise neural networks are incomparably low in contrast to standard convolutional networks, and their application will be feasible for low power mobile devices. Generally speaking, standard convolutions still outperform depthwise convolutions. However, for mobile devices with limited computing capacity, depthwise convolutions are the best.

REFERENCES

1. **Yoshua B**. Learning deep architectures for AI // Foundations and Trends R in Machine Learning. – 2009. - Vol. 2, № 22.

2. **Guo T., Dong J., Li H. and Gao Y.** Simple convolutional neural network on image classification // 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). - Beijing, 2017. – P. 721-724.
3. **Simonyan K., Zisserman A.** Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. - 2014.
4. **Szegedy C., Vanhoucke V., Ioffe S., Shlens J. and. Wojna Z**. Rethinking the inception architecture for computer vision. arXiv preprint arXiv:1512.00567. - 2015.
5. **Krizhevsky, Alex, Sutskever, Ilya, Hinton E., Geoffrey**. ImageNet Classification with Deep Convolutional Neural Networks // Neural Information Processing Systems. 25. 10.1145/3065386. - 2012.
6. **Sun Y., Wang X., Tang, X.** Deep Learning Face Representation from Predicting 10,000 Classes // IEEE Conference on Computer Vision and Pattern Recognition.- 2014.-P.1891-1898.
7. **Chollet F.** Xception: Deep Learning with Depthwise Separable Convolutions // IEEE Conference on Computer Vision and Pattern Recognition (CVPR).- 2017.- P. 1800-1807
8. **Howard G., Andrew, Zhu, Menglong, Chen, Bo, Kalenichenko, Dmitry, Wang, Weijun, Weyand, Tobias, Andreetto, Marco, Adam, Hartwig.** MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.- 2017.
9. **Zhang, Jing, Cao, Yang, Wang, Yang, Zha, Zheng-Jun, Wen, Chenglin, Chen, Chang-Wen**. Fully Point-wise Convolutional Neural Network for Modeling Statistical Regularities in Natural Images.-2018.

National Polytechnic University of Armenia. The material is received 01.02.2019.

Լ.Մ. ՀՈՎՍԵՓՅԱՆ

ՇԱՐԺԱԿԱՆ ՊԼԱՏՖՈՐՄՆԵՐԻ ՀԱՄԱՐ ՊԱՏԿԵՐՆԵՐԻ ՃԱՆԱՉՄԱՆ ՄՈԴԵԼ

Խոր ուսուցման ալգորիթմերը լուծում են պատկերների ճանաչման, օբյեկտների հայտնաբերման և շատ այլ խնդիրներ: Սակայն այդ ալգորիթմները պահանջում են հաշվողական շատ մեծ ռեսուրսներ և իրագործելի են միայն ընդհանուր նշանակության հզոր գրաֆիկական պրոցեսորների վրա: Չնայած կիսահաղորդչային տեխնոլոգիաների զարգացմանը, բջջային սարքավորումների հաշվողական հզորությունները շատ սահմանափակ են, և հետևաբար՝ շատ ալգորիթմական լուծումներ, որոնք հաջողությամբ կիրառվում են ընդհանուր նշանակության համակարգիչներում և սերվերներում, չեն կարող ուղղակիորեն կիրառվել դրանց վրա: Այնուամենայնիվ, հիմնվելով փորձերի վրա, տեսնում ենք, որ ճիշտ ընտրված խոր ուսուցման մոդելը կարող է բավարար չափով արդյունավետ լինել շարժական պլատֆորմներում կիրառվելու համար:

Առանցքային բառեր. խոր ուսուցում, պատկերների ճանաչում, convolutional neural network, depthwise convolutions:

Л.М. ОВСЕПЯН

МОДЕЛЬ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ ДЛЯ МОБИЛЬНЫХ ПЛАТФОРМ

Алгоритмы глубокого обучения могут решить проблемы распознавания изображений, обнаружения объектов и др. Однако эти алгоритмы требуют больших вычислительных ресурсов и могут выполняться мощными графическими процессорами общего назначения. Несмотря на рост индустрии полупроводников, вычислительные возможности мобильных устройств все еще очень ограничены. Поэтому большинство алгоритмических решений, которые были довольно успешными на настольных компьютерах и серверах, не могут быть непосредственно использованы в них. Эксперименты показали, что использование depthwise нейронных сетей может быть достаточно эффективным для мобильных платформ.

Ключевые слова: глубокое обучение, распознавание изображений, сверточные нейронные сети, depthwise нейронные сети.