Proceedings of CECNet 2021
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA210424

Classification of Knee Osteoarthritis Severity Using Modified Masks to Preprocess X-ray Images in a Deep Learning Model

Ching-Chung YANG[1]
Department of Digital Media Design, Tatung Institute of Technology, Taiwan

Abstract. We propose a concise approach to facilitate the deep learning model for medical image classification of knee osteoarthritis severity. The characteristics of the input X-ray images are sharpened by a modified 5×5 mask before training and testing in this work. We compare the inference accuracies of two experiments using the same architecture with images sharpened and not sharpened respectively. And we find it tangible that the former performs much better than the latter. This technique could also be helpful when applied onto the edge devices for object detection and image segmentation.

Keywords. Image classification, deep learning, image sharpening

1. Introduction

Image classification has been inspiring the rapid development of deep learning these years. And it also becomes as important as to adapt the applications to various fields, such as medical treatment, traffic analysis, automated optical inspection, etc.

For medical image diagnosis employing artificial intelligence measures, the preparation of the dataset is sometimes demanding owing to the limited patient cases. In the meanwhile, the detected objects in the input images are subject to be confined within merely several classes. In consequence, there is no need to adopt large size neural network to fulfill the training and testing. A concise model with delicate structure might not only be adequate but easier to install in the edge devices.

However, the challenge is that the differences among the designated classes are usually quite small or even subtle. This inherent nature makes the inference accuracy of the medical deep learning schemes tend to dwell at an unacceptable level.

In this study, we would like to establish a compact model to achieve a sensible precision for classifying knee osteoarthritis severity. The datasets were originally downloaded from the website provided by Chen Ping-jun *etal.* [1]. We use a much simpler architecture to accomplish the similar accuracy by introducing a novel image sharpening technique.

[1] Corresponding author, Department of Digital Media Design, Tatung Institute of Technology, 253 Mi-Tuo Road, Chiayi City, Taiwan.; E-mail: yang10.cc@msa.hinet.net.

Most of the well-known contributions made to develop the deep neural networks focus on the construction of model layers [2-4]. While in this work we'll emphasize the importance at the data preprocessing stage.

The prevalent image preprocessing approaches consist of Gray-scale conversion, Gamma adjustment, Contrast-limited adaptive histogram equalization (CLAHE), etc. These methods mainly improve the global or local contrast of the original images to tell apart objects with distinct appearances. However, the processing of fine characteristics might help more when deciphering the classified objects with similar look. In this article, we demonstrate one unique image sharpening technique to uncover such importance. By applying our algorithm upon the input images for the neural network, the model precision is liable to be evidently escalated with less computing resource consumption.

The commonly used Laplacian is often considered a second-order derivative to sharpen an image's edges. This operator is regarded isotropic because of its rotation invariant for enhancement. But our previous work has proven that the mask filtering is actually a superposition of directional elements [5-13]. It seems promising to exploit this concept onto the image classification scope, in which a number of images had better be preprocessed in order to acquire sensible training weights.

2. Algorithm

It is common to use 3×3 masks to process image objects. While in this work, three masks derived from a 5×5 one is developed to flexibly enhance images.

For a 5×5 mask of [0 -1 0 -1 0; -1 -1 -1 -1 -1; 0 -1 16 -1 0; -1 -1 -1 -1 -1; 0 -1 0 -1 0], there are sixteen coefficients environing the central pixel at coefficient 16. When sharpening an image f (i, j), the second derivative is now formulated as the following:

$$\begin{aligned}\nabla^2 f(i,j) = 16 f(i,j) - [&f(i+1,j) + f(i-1,j) + f(i,j+1) + f(i,j-1) \\ &+ f(i+1,j+1) + f(i-1,j+1) + f(i+1,j-1) + f(i-1,j-1) \\ &+ f(i+2,j+1) + f(i+1,j+2) + f(i-1,j+2) + f(i-2,j+1) \\ &+ f(i-2,j-1) + f(i-1,j-2) + f(i+1,j-2) + f(i+2,j-1)]\end{aligned} \tag{1}$$

The result of Eq. (1) is supposed to be imposed on the original image to obtain an enhanced one g (i, j) as the following:

$$g(i,j) = f(i,j) + \nabla^2 f(i,j) \tag{2}$$

referring to our previous works, it is reasonable to divide Eq. (1) into three parts as the following [14,15]:

$$\begin{aligned}\nabla^2 f(i,j) &= \{4 f(i,j) - [f(i+1,j) + f(i-1,j) + f(i,j+1) + f(i,j-1)]\} \\ &+ \{4 f(i,j) - [f(i+1,j+1) + f(i-1,j+1) + f(i+1,j-1) + f(i-1,j-1)]\} \\ &+ \{8 f(i,j) - [f(i+2,j+1) + f(i+1,j+2) + f(i-1,j+2) + f(i-2,j+1) \\ &\qquad + f(i-2,j-1) + f(i-1,j-2) + f(i+1,j-2) + f(i+2,j-1)]\} \\ &= \nabla^2 f_\alpha(i,j) + \nabla^2 f_\beta(i,j) + \nabla^2 f_\gamma(i,j)\end{aligned} \tag{3}$$

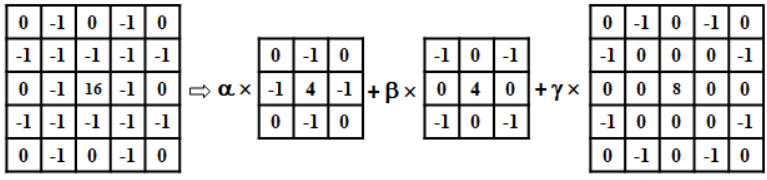
Eq. (3) means that a traditional 5×5 mask with 17 coefficients can now be dissembled into three masks, of which the second derivatives are just like summations of first derivatives along different directions as the following Eqs. (4), (5) and (6):

$$\begin{aligned}\nabla^2 f_\alpha(i,j) &= 4f(i,j) - [f(i+1,j) + f(i-1,j) + f(i,j+1) + f(i,j-1)] \\ &= [f(i,j) - f(i+1,j)] + [f(i,j) - f(i-1,j)] \\ &+ [f(i,j) - f(i,j+1)] + [f(i,j) - f(i,j-1)]\end{aligned} \tag{4}$$

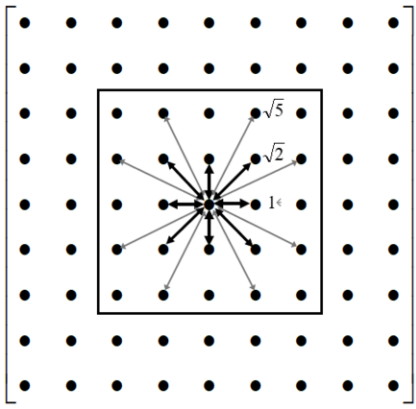
$$\begin{aligned}\nabla^2 f_\beta(i,j) &= 4f(i,j) - [f(i+1,j+1) + f(i-1,j+1) + f(i+1,j-1) + f(i-1,j-1)] \\ &= [f(i,j) - f(i+1,j+1)] + [f(i,j) - f(i-1,j+1)] \\ &+ [f(i,j) - f(i+1,j-1)] + [f(i,j) - f(i-1,j-1)]\end{aligned} \tag{5}$$

$$\begin{aligned}\nabla^2 f_\gamma(i,j) &= 8f(i,j) - [f(i+2,j+1) + f(i+1,j+2) + f(i-1,j+2) + f(i-2,j+1) \\ &\quad + f(i-2,j-1) + f(i-1,j-2) + f(i+1,j-2) + f(i+2,j-1)] \\ &= [f(i,j) - f(i+2,j+1)] + [f(i,j) - f(i+1,j+2)] \\ &+ [f(i,j) - f(i-1,j+2)] + [f(i,j) - f(i-2,j+1)] \\ &+ [f(i,j) - f(i-2,j-1)] + [f(i,j) - f(i-1,j-2)] \\ &+ [f(i,j) - f(i+1,j-2)] + [f(i,j) - f(i+2,j-1)]\end{aligned} \tag{6}$$

Eq. (4) denotes the mask with four coefficients along the vertical and horizontal directions, Eq. (5) denotes the mask with four coefficients in the diagonal directions, and Eq. (6) is another one that has eight coefficients along other eight different directions. The three masks shown on the right-hand side of Fig. 1(a) essentially have different intrinsic properties due to the different distances from their coefficients to the target pixel, which is indicated in Fig. 1(b). Eq. (4) could extract the finest characteristics from the original image, Eq. (5) implies the higher spatial frequencies, and Eq. (6) contains more high-intermediate spatial frequencies. In order to flexibly adapt to different image compositions, Eq. (3) is more sensible to be modified as the following:



(a)



(b)

Figure 1. (a) The 5×5 mask divided into three masks with assigned weights. (b) The distances to the target pixel from different mask coefficients.

$$\nabla^2 f(i,j) \Rightarrow \alpha \times \nabla^2 f_\alpha(i,j) + \beta \times \nabla^2 f_\beta(i,j) + \lambda \times \nabla^2 f_\gamma(i,j) \tag{7}$$

And Eq. (2) could be rewritten as

$$g(i,j) = f(i,j) + \alpha \times \nabla^2 f_\alpha(i,j) + \beta \times \nabla^2 f_\beta(i,j) + \lambda \times \nabla^2 f_\gamma(i,j) \tag{8}$$

The above enhancement technique is developed here to preprocess the datasets used in this work. Then we subsequently build a neural network similar to Lenet-5 to illustrate the merit of our proposed method.

3. Experiment

Fig. 2 shows our entire neural network architecture. Images with and without sharpening are about to be input to the model respectively for comparison. Every two convolutions are followed by a maxpooling. The flattening converts the image matrix to a one-dimensional full connection layer. Afterward, a dropout processing is combined with the hidden layer to help reduce the over-fitting phenomenon. The output layer performs the classification of five classes.

The intended input data in Fig. 2 are all gray-scale images with a single channel. The first two convolutions containing 3×3 masks are with 64 and 32 output channels respectively. And the last two are both with 16 output channels. Because the two maxpoolings totally rescale an image to one sixteenth its original size, the flattened layer would contain the same pixels number as the input image of 128×128 = 16384.

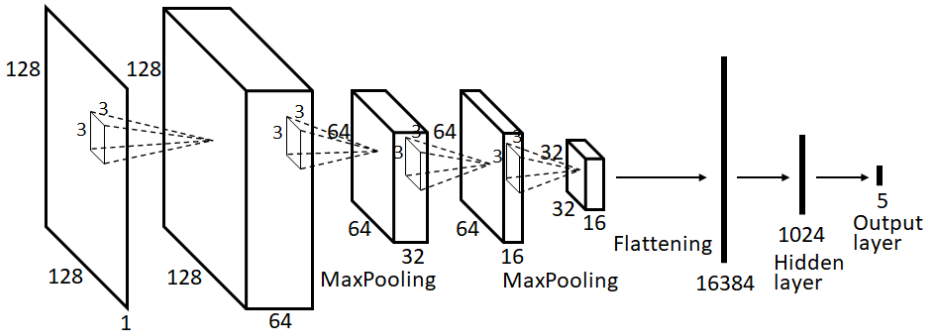


Figure 2. The CNN model and its parameters used in this work.

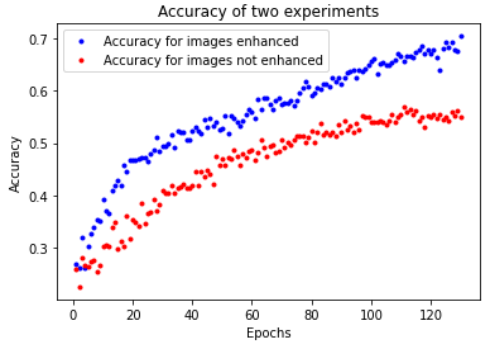
The images in the original datasets were all rescaled to be with the same size of 128×128 in this work. The severity of knee osteoarthritis is classified with five levels from 0 to 4. Images in the folders were rearranged so that every level contains 1500 images, with the exception that the folder labelled 4 only has 500 images. This is because that patients with most severe knee osteoarthritis are fewer than others. And the fewer images don't seem to degrade the experiment results too much.

Then we complete the desired datasets from all the labelled images. The training dataset includes 80% of them, the testing 10%, and the validation 10%. For comparison, we proceed twice the experiment. The first time is with datasets not sharpened, and the second time with datasets sharpened in advance by our method. Both of them exploit the same neural network model with identical parameters.

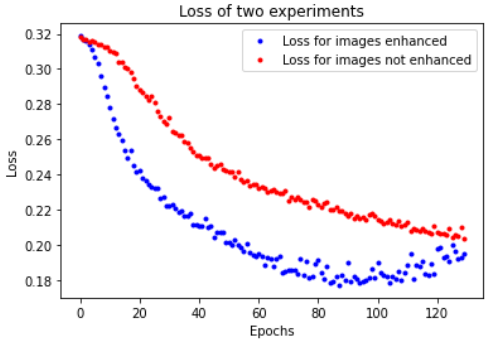
The constants used for sharpening images at the second time are (α, β, γ) = (0.4, 1.6, 0.15) in Eq. (8). In real practice, f (i, j) in Eq. (8) can be multiplied by a constant near 1 to slightly adjust the brightness of the final image. We emphasize the weight of β as to magnify more higher spatial frequencies, which are considered important for revealing the characteristics of the knee osteoarthritis. All the constants are decided by empirical trial and error.

4. Result

Fig. 3(a) shows that the acquired inference accuracy is about 56% after 130 epochs of training at the first time. And it is about 71% at the second time. These outcomes were calculated by implementing Keras 2.4 and TensorFlow 1.14.0 in Python 3.7. Although the inference accuracy is possible to get higher after more epochs, Fig. 3(b) tells that the cross-entropy loss is also increasing in the meantime. In this situation, we have to take into account the cost of sacrificing the prediction confidence in case of running more epochs.



(a)



(b)

Figure 3. (a) The prediction accuracy diagram of the knee testing dataset for the two experiments. (b) The prediction loss diagram for the same two experiments.

At last, we load the trained weights trying to predict 50 labelled images in the testing dataset. Fig. 4 shows those images without enhancement and Fig. 5 shows the same images with enhancement. Table 1 elaborates the confusion matrices of these two experiments. It is conspicuous to observe the improvement in the matrix by our method.

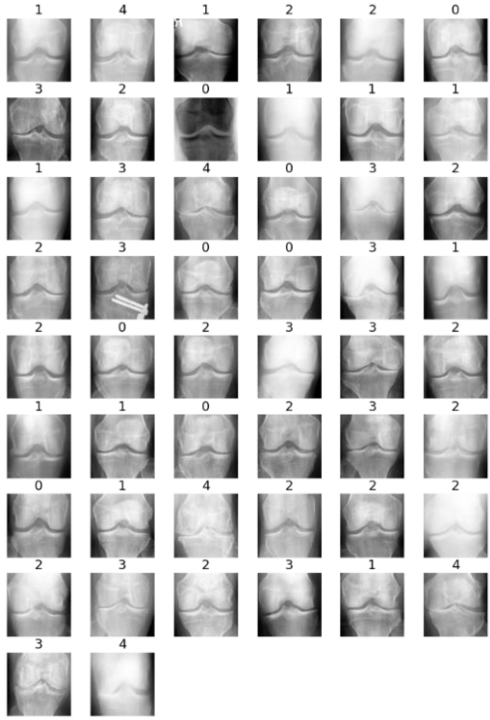


Figure 4. The 50 labelled knee testing images without enhancement.

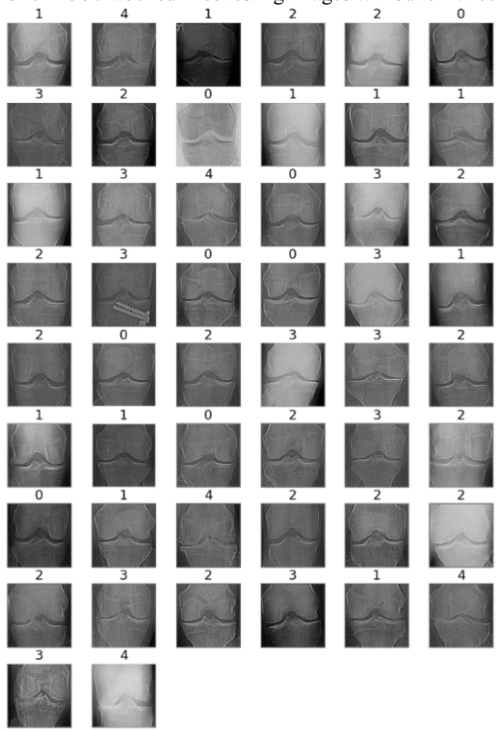


Figure 5. The same 50 labelled knee testing images with enhancement.

Table 1. The confusion matrices of the two experiments. (a). With images not enhanced. (b). With images enhanced.

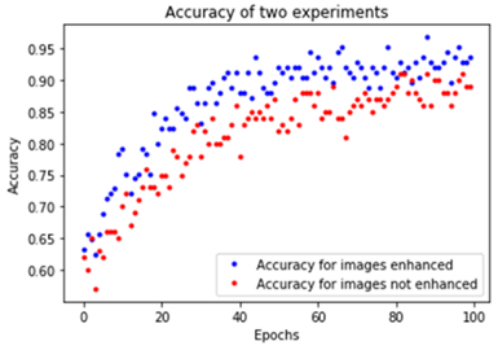
predicted / labelled	0	1	2	3	4
0	5	1	2	0	0
1	3	2	5	1	0
2	2	6	7	0	0
3	0	0	1	10	0
4	0	0	0	2	3

(a)

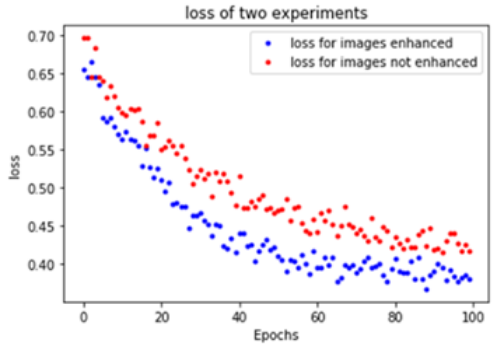
predicted / labelled	0	1	2	3	4
0	6	0	2	0	0
1	2	5	4	0	0
2	3	3	9	0	0
3	0	0	1	10	0
4	0	0	0	0	5

(b)

5. Discussion and Conclusion



(a)



(b)

Figure 6. (a) The prediction accuracy diagram of the liver testing dataset for the two experiments. (b) The prediction loss diagram for the same two experiments.

To ensure the merit of our proposed method, we employ other datasets to train the model and also observe the improvement after the image enhancement. The practiced data are acquired from liver-cancer MRI scanning by Tungs' Taichung MetroHarbor Hospital in Taiwan. Livers inside patients' abdomens were scanned from top to bottom with image separation of several millimeters. The images had been with pseudonymization and de-identification when adopted in this work. We built the datasets by use of the separated scanned images, which are labelled 1 or 0 representing images with or without cancer focal area. Totally 664 images are labelled 0 and 336 images are labelled 1, for that the images with focal areas are fewer than those without.

Figure 6 shows the accuracy and loss curves for the training. The accuracy has been elevated from 90% to 95%. Figure 7 are the testing images without enhancement, and Figure 8 are the same images after enhancement. The constants used for Figure 8 are (α, β, γ) = (1.5, 0.1, 0.1). We emphasize the weight of α in this place because liver is composed of soft tissue with more sophisticated structures.

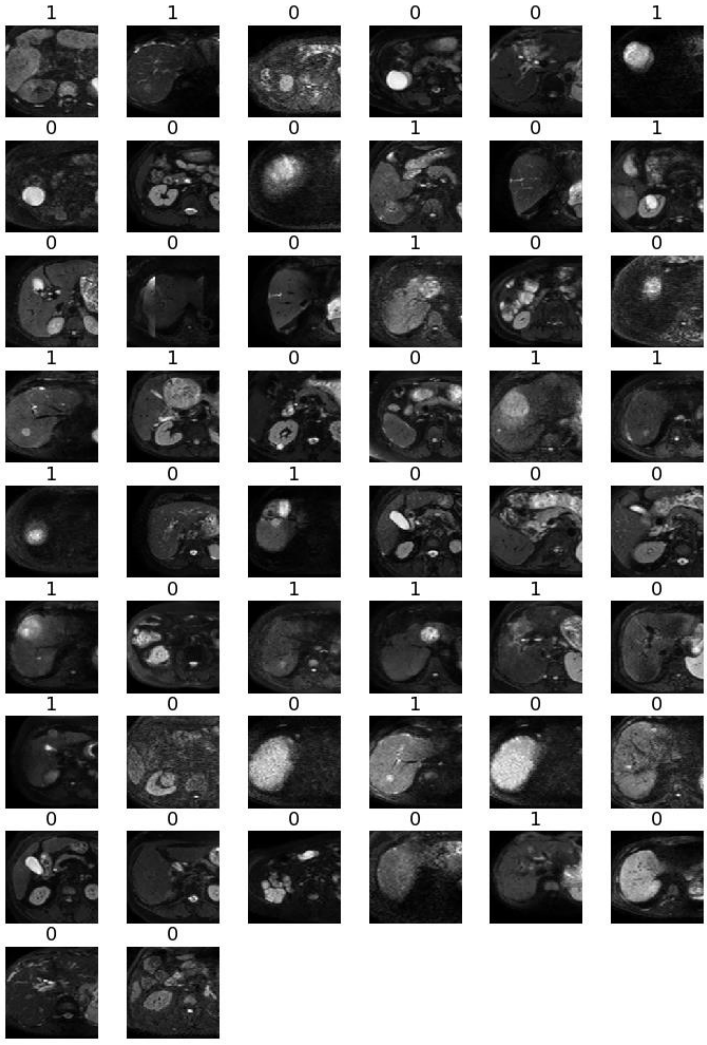


Figure 7. The 50 labelled liver testing images without enhancement.

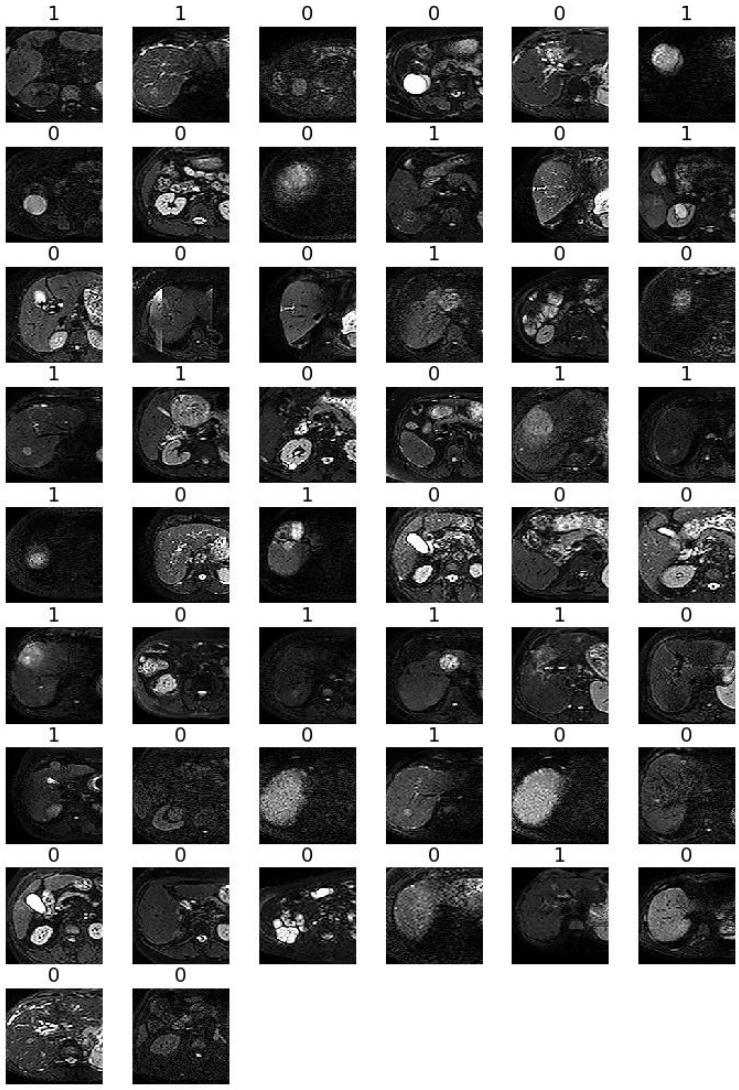


Figure 8. The same 50 labelled liver testing images with enhancement.

In conclusion, the image sharpening technique mentioned in this article is mainly based on mask-filtering approach, which has some similarity with the convolution operation in deep neural networks. Although one of them is with fixed coefficients and the other with trained coefficients in multi-channels, both of them extract features of the image.

Along with the emerging demand for edge computing software embedded in smaller devices, it seems promising that the recommended strategy serves as a preferential alternative. The proposed tactic not only reduces computing consumption, but provides the flexibility to accommodate to variant image compositions.

References

[1] Chen PJ, Gao LL, Shi XS, Kyle A, Yang L. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Computerized Medical Imaging and Graphics 2019;75:84.
[2] Yann L, Léon B, Yoshua B, Patrick H. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998;86:2278.
[3] Krizhevsky A, Sutskever I, Geoffrey EH. ImageNet classification with deep convolutional neural networks. Communications of the ACM 2017;60:84.
[4] Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations 2015 May.
[5] Yang CC. Improving the sharpness of an image with non-uniform illumination. Opt. Laser Technol. 2005;37(3):235-8.
[6] Yang CC. Image enhancement by modified contrast-stretching manipulation. Opt. Laser Technol 2006;38(3):196-201.
[7] Yang CC. Image enhancement by adjusting the contrast of spatial frequencies. Optik 2008;119(3):143-6.
[8] Yang CC. Image enhancement by the modified high-pass filtering approach. Optik 2009;120(17):886-9.
[9] Yang CC. A modification for the mask-filtering approach by superposing anisotropic derivatives in an image. Optik 2011;122(18):1684-7.
[10] Yang CC. Color image enhancement by a modified mask-filtering approach. Optik 2012;123(19):1765-7.
[11] Yang CC. Improving the overshooting of a sharpened image by employing nonlinear transfer functions in the mask-filtering approach. Optik 2013;124(17):2784-6.
[12] Yang CC. Improving the sharpness of a non-uniformly illuminated image by an integral mask-filtering approach. Optik 2013;124(21):5049-51.
[13] Yang CC. Finest image sharpening by use of the modified mask filter dealing with highest spatial frequencies. Optik 2014;125(8):1942-4.
[14] Yang CC. Improving the sharpness of a color image with slowly-varying hues by undertaking its intensity and saturation factors. Optik 2014;125(22):6730-2.
[15] Yang CC. Ultra-fine characteristics of a sharpened image rendered by decreasing the critical low-intermediate frequencies as well as increasing the critical higher frequencies. Optik 2015;126(23):3803-6.