Koo and Cha *Hum. Cent. Comput. Inf. Sci.* (2017) 7:33
DOI 10.1186/s13673-017-0114-5

Human-centric Computing and Information Sciences

RESEARCH Open Access

Image recognition performance enhancements using image normalization

Kyung-Mo Koo and Eui-Young Cha*

*Correspondence:
eycha@pusan.ac.kr
Pusan National University,
Busan, Republic of Korea

Abstract

When recognizing a specific object in an image captured by a camera, we extract local descriptors to compare it with or try direct comparison of images through learning methods using convolutional neural networks. The more the number of objects with many features, the greater the number of images used in learning, the easier it is to compare features. It also makes it easier to detect if the image contains the feature, thus helping generate accurate recognition results. However, there are limitations in improving the recognition performance when the feature of the object to be recognized in the image is significantly smaller than the background area or when the area of the image to be learned is insufficient. In this paper, we propose a method to enhance the image recognition performance through feature extraction and image normalization called the preprocessing process, especially useful for electronic objects with few distinct recognition characteristics due to functional/material specificity.

Keywords: Image recognition, Pre-processing, Normalization, Feature Extraction, Deep learning

Introduction

As the performance of mobile devices improves and the number of functions included in the devices increases, technologies that can be implemented using these devices are also diversifying. Especially, the technology utilizing cameras has been expanding its application field, ranging from augmented reality to recognizing wine labels, book covers, packaged goods, and similar clothes employing a form of style search.

When recognizing a specific object in an image captured by a camera, it is possible to compare the existing indexed content with a local descriptor that can extract the same feature repeatedly without being affected by the size change and the shooting angle. For instance, features such as SIFT [1, 2], SURF [3], BRIEF [4], ORB [5], MSER [6, 7] or the image of the region (or object) estimated by a saliency map [8] or selective search [9] are learned and recognized using the convolutional neural networks (CNN). The more the number of extractable features in the region (or object), the easier it is to compare the presence or absence of each feature and to deliver the accurate recognition result.

In the case of printed photographs, printed book covers, and industrial packaging materials, there are many features that are easy to extract from the local descriptors such as the image itself, the logo using various colors and patterns, and the packaging design, so that a relatively accurate recognition result can be obtained. However, for consumer

goods such as TV, refrigerator, washing machine, air conditioner, etc., it is difficult to extract the local descriptors that can be used to easily compare the characteristics because of the functional (TVs, monitors, etc., whose main purpose is screen output, do not have any features on the screen. Once they are turned on, other features not related to the object to be recognized may interfere with the recognition) and material (the surfaces of refrigerators and air conditioners may be coated with light decoration) specificity.

Recently, image recognition by deep running has been getting popular. While it is true that are producing effective for some images, but they depend on the settings of a number of learning parameters in complex, nonlinear ways. Selecting good parameters is critical to the performance of the learning algorithm, but it is largely a black art [10–13].

In this paper, we propose a technique to improve the recognition performance using the preprocessing process that detects the distinguishable features of each product and normalizes them, with the aim of recognizing the manufacturer and the product name of the electronic product. In "Extraction of features", we describe the feature extraction method and normalization method for each product. In "Convolutional neural networks", we describe the method of neural network construction for normalized image recognition. The experimental results are described in "Results" and conclusions and future research plans are described in "Conclusions".

Extraction of features

First, we introduce the process of extraction for recognition in detail. Figure 1 depicts the flowchart for the recognition process. For example, in the case of a TV, once the four vertices are identified, the image is warped to a rectangle, and an edge image of the stand, extracted from the lower portion of the TV, is used for learning and recognition.

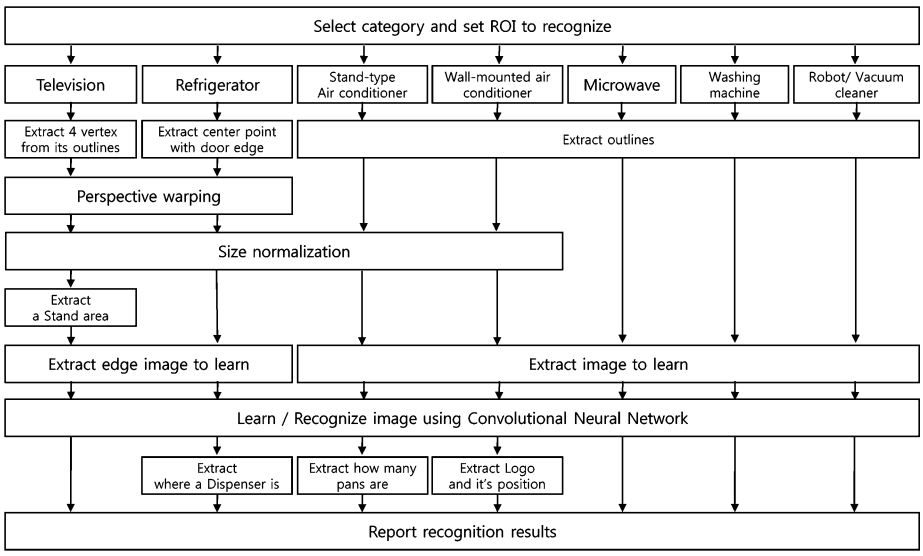

Fig. 1 Flowchart of recognition algorithm for each electronic device. This includes discrete processes for extracting features to normalize and recognize images

As mentioned earlier, electronics are limited in their possibility to extract comparable features because of their functional/material specificity. Another limitation is that the filtering/pooling layer results for the entire image are obtained only in the non-electronic components during the CNN learning process. Thus, the process of normalization of images for recognition is different for each different type of the electronic product. This process includes a preprocessing process for extracting features, a process for extracting additional individual features that facilitate recognition of the electronic product, and a process for normalizing the extracted image features.

Pre-processing

In the preprocessing step, preparations are made to extract features that are easy to recognize in each image.

At first, if you look at a television, it is difficult to distinguish it from the others because every TV is just a black screen when the screen is off. Only when the screen is turned on, it can be determined whether it can be recognized based on the features. However, these are not exactly the features of the television. Hence, the shape of the stand supporting the monitor and not the monitor is used to recognize the television. After checking all 10 models of new TVs of 3 major consumer electronics manufacturers, we confirmed that the stand shape of all products is unique, and this was a logical assumption because the shape of the stand also has a design patent. First, we extract five primary lines from the top, left, right and bottom contours of the TV monitor using Canny Edge detector [14] and Hough Transform as shown in Fig. 2 (left). Especially in this part, we divide the top contour into two parts because of the curved television display. Next, we obtain the four vertices pairs of straight lines meet. Subsequently, using the information of these points, the image is warped into a rectangular perspective of a certain size. Because we will use the structure of the stand at the bottom of the television image for recognition, rather than the image itself, to change the object closer to the front view to extract it correctly. After separating the stand image at the lower end of the normalized television image, the edge image is obtained using the Sobel filter.

Fig. 2 The process of pre-processing for recognizing Television; extract 4 vertices using 5 edges (divide the *upper edge* into two parts to account for the curved display), extract lower end from a warped perspective image and extract a Sobel filtered image

The preprocessing of the refrigerator is a process for extracting door edge information so that it can be utilized appropriately. A light-tone pattern is printed on the refrigerator, which makes it difficult to distinguish as its characteristics are not clearly seen when photographed with a camera. For this reason, we use the outline of the refrigerator door for recognition. After checking 14 models of new refrigerators of 2 major consumer electronics manufacturers, we confirmed that the door edge of all products is unique. As shown in Fig. 3, the door edge is extracted followed by the image extraction so that the distribution of the outline can be seen based on the center point where these outlines meet, and an outline image is obtained. The center points of the horizontal and vertical edges are obtained using cross points from every edge and clustering them to get the central position. Next, a principal reference line was obtained based on the center point, and a parallelogram image was obtained by extending this line in parallel. This image is then warped into perspective and normalized into a specific size. Sobel edge images are used for final learning and recognition, just like in the case of television. This takes into account unique features such as the handle of the refrigerator.

We can also check if the refrigerator has a built-in dispenser using an edge distribution of the top-left portion. This process is covered in more detail in the next chapter.

For the two aforementioned categories of electronics, the preprocessed image is used for learning, and the size is normalized only after extracting the image outlines, a process used for learning to recognize the remaining categories of objects.

Extraction of additional features

Some electronic objects look exactly the same with only one or two options being different. Sometimes, even different manufacturers need to introduce different features to distinguish their products from the competition. In this paper, we discuss how to extract features that can provide additional information on recognition results in further detail.

In case of refrigerators, the recognition of the model should not be affected by the presence of a water purifier (or ice dispenser). Hence, we attach a different label to the top-left edge enhanced image when we perform the learning.

An air-conditioner is classified into two types for the purpose of recognition. The image is cropped based on the outline and different algorithms are applied using a size-normalized image. For a stand-mounted air conditioner, the number of circles in the image is counted after extracting the edge (for wind hole) as shown in Fig. 4 and used as additional information along with the learning/recognition results. We extract a canny

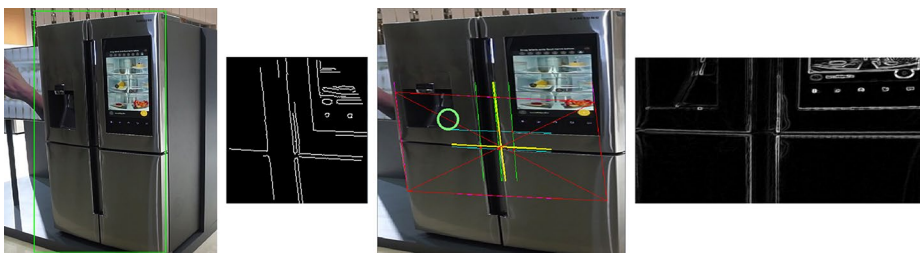

Fig. 3 The process of pre-processing for recognizing a refrigerator: extract center point using *horizontal* and *vertical edges* of the center, and then extract Sobel-filtered image from perspective-warped image

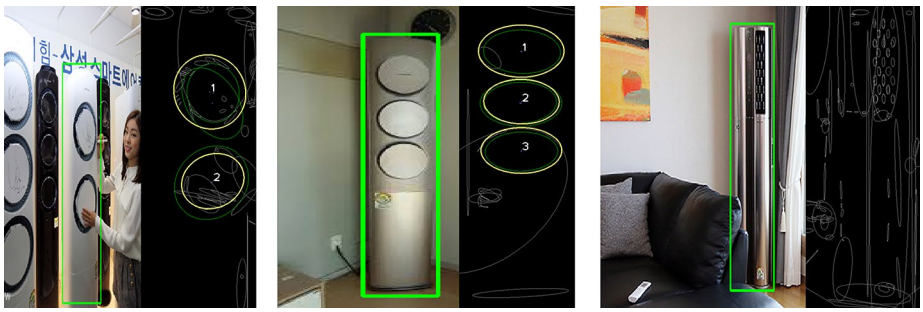

Fig. 4 The process of extracting additional information for recognizing the stand-mounted air conditioner: find fit circles from a canny edge and cluster them. Circular wind holes are arranged from one to three based on size. The fact that the wind hole is not circular also affects the final outcome

edge from a size-normalized image and find principal ellipses [15] based on inclusion relation of each ellipse. Each contour group is clustered using the radius of the circumscribed circle and the distance between the centers.

We extract the logo (including the manufacturer's name) from the center of the bottom of the image and directly learn/recognize it for wall-mounted air conditioners, as shown in Fig. 5. We use top-hat morphology to extract the logo. Morphology [16] is a method of approaching images from a space point of view and has an advantage of being easy to understand because the result of the operation can be seen visually. Typical examples of morphology operations are erosion and dilation. First, we obtain the difference image of the result of the dilation/erosion operation of the center image of the wall-mounted air conditioner image [17], and then binarize it. This is followed by a dilation operation to emphasize the logo area, a method often used to extract small characters in real-world images [18].

Convolutional neural networks

In this paper, we use a fine-tuned CaffeNet [19] model for recognition of normalized images. The CaffeNet model is based on 1.2 million high-quality images classified into 1000 categories during the Large Scale Visual Recognition Challenge (LSVRC)-2010 competition. It has updated its existing record with 37.5 and 17.0% error rates in the top-1 and top-5 categories respectively.

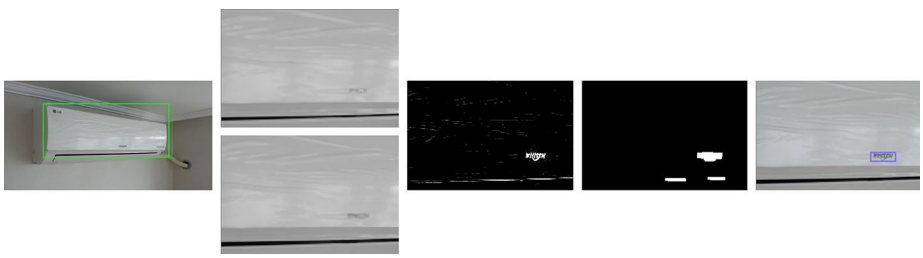

Fig. 5 The process of extracting a logo for recognizing the wall-mounted air conditioner: find character group using top-hat morphology and extract the logo using its structure

CaffeNet structure

The model structure has the following characteristics: first, the gradient vanishing problem is solved by using a rectified linear unit (ReLU) nonlinearity activation function with non-saturating characteristics and a learning speed that is faster than the activation function of the existing saturating nonlinearity characteristic.

Local reaction normalization is performed after ReLU nonlinearity. This is the brightness normalization process that is affected by the actual neurons, thereby reducing the error rates of 1.4 and 1.2% in the top-1 and top-5 categories respectively.

When the pooling size is z and the interval between the pooling units is s pixels, with 0.4 and 0.3% error rates in top-1 and top-5 categories respectively overlapping pooling is performed with the knowledge that $s < z$.

Using two GPUs reduces the error rates by 1.7 and 1.2% in top-1 and top-5 categories, respectively, when compared to usage of one GPU.

To solve the over fitting problem, set the result value of any hidden neuron to 0 so that it does not affect the learning. We use the dropout method in which all structures share the weights while learning the models of different structures every time. Combine the arbitrary partial neurons to learn more robust and useful features.

Fine tuning

The fine tuning process transforms the architecture for a new purpose based on the previously learned model, and updates the weights of the learning based on the previously learned model weights. We tuned into more than 5800 pre-processed electronic object image-sets to recognize 55 home appliances instead of object category recognition through the BVLC CaffeNet model. The CaffeNet model works well for object classification and we want to use it to recognize our electronic objects in detail.

We have more than 5800 pre-processed images to learn and have begun fine-tuning with the parameters learnt from 1,000,000 image-net images. If we provide the weights argument to the Caffe train command, the previously learned weights melt into our model, and the layers will match by name. In other words, a new data classifier will be created based on previously learned models. We changed the last layer's name of the existing CaffeNet model from fc8 to fc8_television, fc8_ refrigerator, and so on. Since there is no layer name in the existing bvlc_reference_ caffenet layer, this layer starts learning with random weights. We have created new models for all the eight categories of home appliances using fine-tuning. The results are discussed in the following section.

Results

In this paper, 55 kinds of home appliances preprocessed by the proposed method were recognized. In this section, we describe the learning set used in the test, and evaluate the proposed algorithm by comparing the recognition results of the original image, the cropped image of the recognition target part, and the preprocessed recognition target image.

Full datasets—8 categories, 55 kinds of home appliances

We collected more than 10,000 images for eight kinds of home appliances such as television, refrigerator, washing machine, air conditioner, robot cleaner, etc. from various

sources such as the manufacturer's homepage, internet, shopping mall, blogs, articles of product users, and by shooting at actual stores. They were manually labeled as one of the 55 types and the number of images per product ranged from as low as 80 to as many as 400. Few of the various collected images are shown in Fig. 6.

Normalization results

We use preprocessing especially for television and refrigerator as we described. It related to recognition accuracy, of course the better normalization will make the better the recognition rate.

First, we extract normalized television stand's edge image from source image as Fig. 7. Experimental results show that preprocessing works well for images taken at slightly oblique angles. We convert this edge image into six distorted version to input convolutional neural network as Fig. 8.

Next, we extract normalized refrigerator door's edge image from source image as Fig. 9.

Fig. 6 As you can see from the collected images, we tried to collect images from various angles in as many situations as possible. Especially for television, as mentioned above, different content and even the same content was displayed multiple times

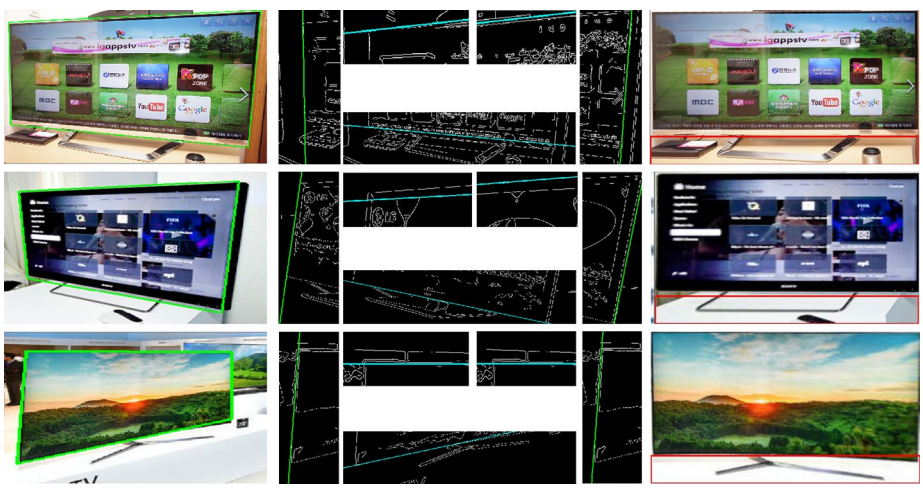

Fig. 7 Extract 4 vertices using 5 edges, extract lower end from a warped perspective image

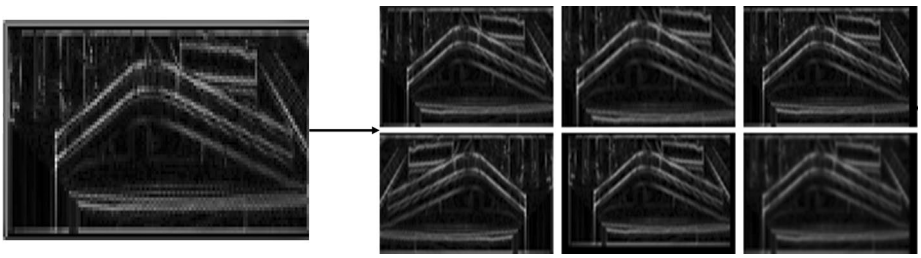

Fig. 8 Extract a Sobel filtered image and convert it into six distorted (resize, move a few pixels to the left and right, and we flip left and right for each images) images for using classification

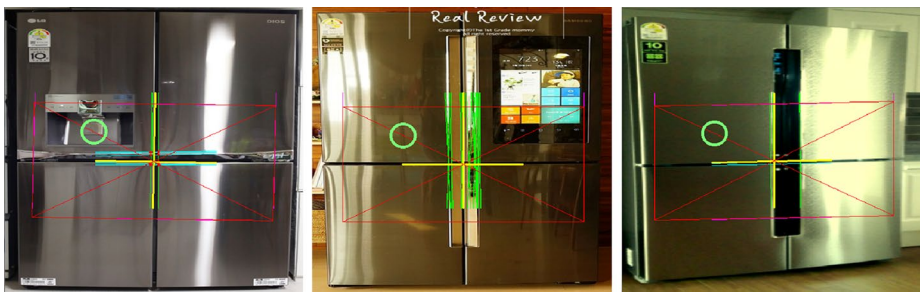

Fig. 9 Extract center point using *horizontal* and *vertical edges* of the center, and then extract Sobel-filtered image from perspective-warped image

Classification results

Classification was done for each of the kinds of electronic products. The average classification performance for the whole class is 93.24% and each classification performance is shown in detail in Table 1. The recognition rate of the television, which is judged to be the most difficult to recognize and usually displays the lowest recognition performance, was found to be 87.71% for ten types. However, it can be confirmed from the graph in Fig. 10. that the performance is much better than the process when only the original image was classified or only the stand area image was extracted.

Next, the performance of 14 kinds of refrigerator classifications was found to be 94.06% as shown in Fig. 11, the performance of seven types of washing machines was

Table 1 The number of images used for learning and recognition performance are described for each category

Category	Kind of	# of images	Performance (%)
Television	10	1888	87.71
Refrigerator	14	1181	94.06
Washing machine	7	856	96.85
Stand-type air conditioner	8	340	99.12
Wall-mounted air conditioner	2	136	90.44
Robotic vacuum cleaner	2	838	100.0
Vacuum cleaner	4	398	98.99
Microwave	8	165	96.36

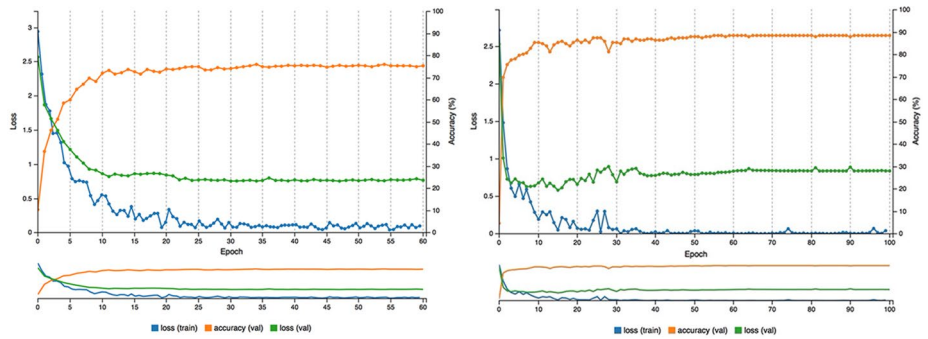

Fig. 10 Test the television in two conditions and compare the results: classifying only the objects to be recognized (75.40%), and classifying the outline only from the stand images (87.7%)

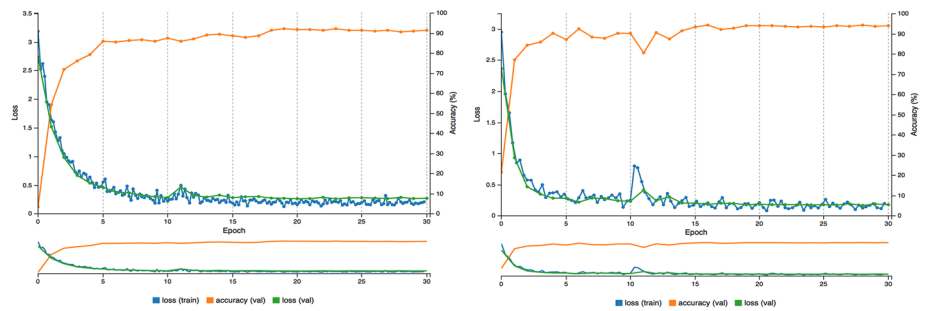

Fig. 11 Test the refrigerator in two conditions and compare the results; classifying only the objects to be recognized (90.94%), classifying the edge image at the center of the refrigerator that was extracted using the door edge information (94.96%)

found to be 96.85%, and the performance of eight kinds of stand-mounted air conditioners was found to be 99.12% as shown in Fig. 12.

Conclusions

In this paper, we have discussed the importance of preprocessing and evaluated the improvement in recognition performance when applying deep learning to the recognition of home appliances. Convolutional neural networks is a model that is optimized for vision while minimizing the complexity of the model based on three ideas: sparse weight, tied weight, and equivariant representation. The process can recognize many objects with its complex capabilities. Many types of improved techniques are being introduced routinely and will continue to be introduced.

However, most techniques do not take rotation invariance into account, for which a large amount of well-formed datasets are required, or unnecessary information has to be manually excluded from the learning data, which is not an ideal algorithm that can easily be applied to all areas. It is more desirable to specify the problem using human intelligence and the computer is supposed to do the work to help it. Therefore, it is necessary to continue the process of extracting and recognizing various features that cannot be extracted by the convolutional neural networks.

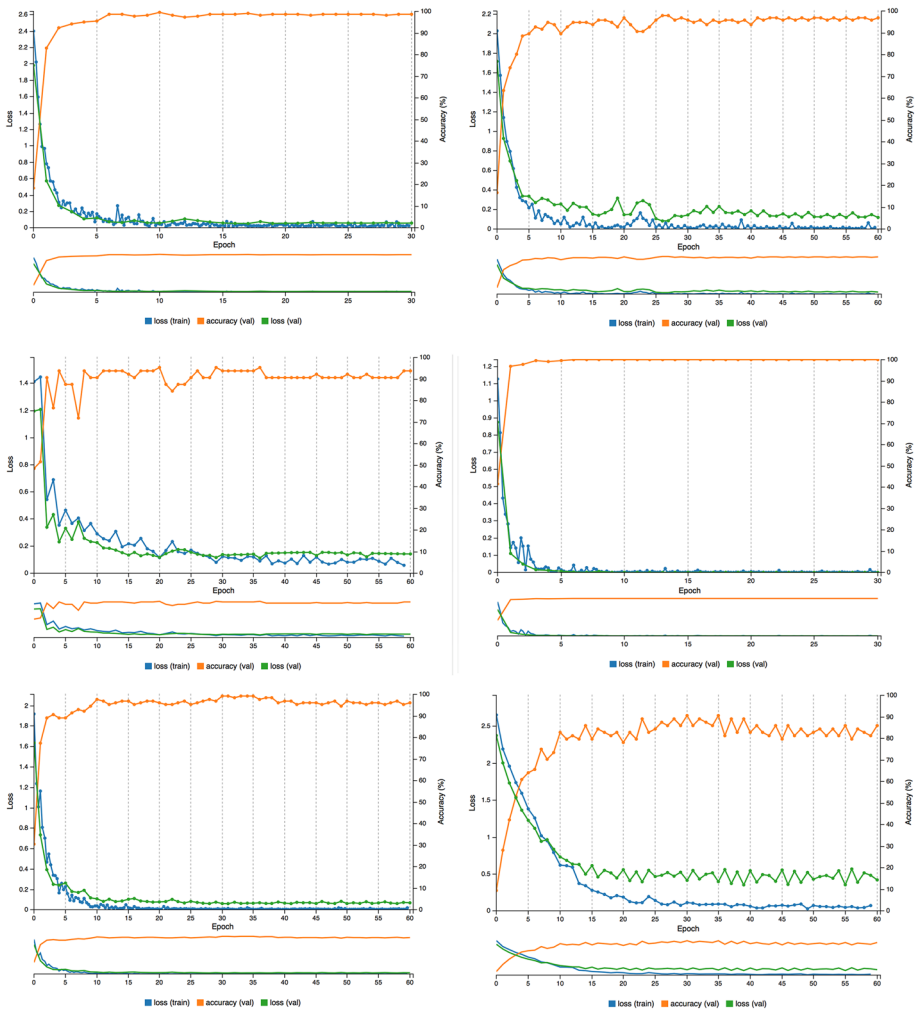

Fig. 12 The washing machine classification accuracy is 96.85% (*top-left*); the stand-mounted air-conditioner accuracy is 96.85% (*top-right*). The centerlines are arranged in the order of wall-mounted air-conditioners (93.75%), robotic vacuum cleaners (99.9%). The *lines* at the *bottom* are arranged in order of vacuum cleaners (96.1%) and microwaves (85.9%), for all of which we extract only the object area and use the image for classification

Future work on this topic could include the exploration of extracting meaningful features not only from visual images but also based on material and atmosphere, especially in the field of fashion, to generate a model with enhanced performance.

Authors' contributions
KMK and EYC designed the study, developed the study and the methodology, collected the data, performed the analysis, and wrote the manuscript together. Both authors read and approved the final manuscript.

Acknowledgements
This work was supported by a 2-Year Research Grant of Pusan National University.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
Not applicable.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Funding
Not applicable.

Publisher's Note

Received: 25 May 2017 Accepted: 29 August 2017
Published online: 19 November 2017

References

1. Lowe DG (1990) Object recognition from local scale-invariant features. In: The proceedings of the seventh IEEE international conference on, vol 2. IEEE, New York, pp 1150–1157
2. Lindeberg Ty (1994) Scale-space theory: a basic tool for analysing structures at different scales. J Appl Stat 21(2):224–270
3. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: Computer vision–ECCV 2006. pp 404–417
4. Calonder M, Lepetit V, Strecha C, Fua P (2010) BRIEF—binary robust independent elementary features. In: Computer vision–ECCV 2010. pp 778–792
5. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB—an efficient alternative to SIFT or SURF. In: 2011 IEEE international conference on computer vision (ICCV). IEEE, New York, pp 2564–2571
6. Donoser M, Horst B (2006) Efficient maximally stable extremal region (MSER) tracking. In: Proc. of IEEE computer society conference on computer vision and pattern recognition (CVPR). pp 553–560
7. Forssen P-E, Lowe DG (2007) Shape descriptors for maximally stable extremal regions. IEEE 11th international conference on computer vision, 2007. ICCV 2007. pp 1–8
8. Courty N, Marchand E (2003) Visual perception based on salient features. In: Proceedings. International conference on 2003 IEEE/RSJ, vol 1. IEEE, New York, pp 1024–1029
9. Uijlings JRR, van de Sande KE, Gevers T, Smeulders AWM (2013) Selective search for object recognition. Int J Comput Vis 104(2):154–171
10. Maninis KK, Pont-Tuset J, Arbeláez P, Van Gool L (2016) Deep retinal image understanding. In: International conference on medical image computing and computer assisted intervention. Springer International Publishing, Berlin, pp 140–148
11. Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A (2016) Deep neural networks predict hierarchical spatio-temporal cortical dynamics of human visual object recognition. arXiv preprint arXiv:1601.02970
12. Zufferey D, Gisler C, Khaled OA, Hennebert J (2012) Machine learning approaches for electric appliance classification. In: 2012 11th international conference on information science, signal processing and their applications (ISSPA). IEEE, New York, pp 740–745
13. Gajowniczek K, Ząbkowski T (2015) Data mining techniques for detecting household characteristics based on smart meter data. Energies 8:7407–7427
14. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 8(6):679–698
15. Fitzgibbon AW, Fisher RB (1995) A buyer's guide to conic fitting. In: Proc. 5th British machine vision conference, Birmingham, pp 513–522
16. Thomas W, Klein JC (2001) Segmentation of color fundus images of the human retina—detection of the optic disc and the vascular tree using morphological techniques. In: International symposium on medical data analysis. Springer, Berlin, pp 282–287
17. Sari S, Asahrori SE, Roslan H, Ibrahim N (2015) Gabor edge detection method based on bilateral filter and otsu threshold for noisy ultrasound image. In: Proceedings of recent advances in mathematical and computational methods. pp 88–95
18. Koo KM, Cha EY (2013) A novel container ISO-code recognition method using texture clustering with a spatial structure window. Int J Softw Eng Appl 7(3):51–61
19. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105