2016 IEEE 28th International Conference on Tools with Artificial Intelligence

Visual orientation inhomogeneity based convolutional neural networks

Sheng-hua Zhong
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, P.R. China
csshzhong@szu.edu.com

Jiaxin Wu
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, P.R. China
Jiaxin.wu@email.szu.edu.cn

Yingying Zhu[1,2,*]
College of Computer Science and Software Engineering
Shenzhen University
zhuyy@szu.edu.cn
[2]Department of Computer Science
The University of Texas at San Antonio
San Antonio, TX 78249, US
yingying.zhu@utsa.edu
*corresponding author

Peiqi Liu
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, P.R. China
2013150064@email.szu.cn

Jianmin Jiang
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, P.R. China
jianmin.jiang@szu.edu.cn

Yan Liu
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, P.R. China
csyliu@comp.polyu.edu.hk

Abstract—**The details of oriented visual stimuli are better resolved when they are horizontal or vertical rather than oblique. This "oblique effect" has been researched and confirmed in numerous research studies, including behavioral studies and neurophysiological and neuroimaging findings. Although the "oblique effect" has influence in many fields, little research integrated it into computational models. In this paper, we try to explore this inhomogeneity of visual orientation based on Convolutional neural networks (CNNs) in image recognition. We validate that visual orientation inhomogeneity CNNs can achieve comparable performance with higher computational efficiency on various datasets. We can also get the conclusion that, compared with the cardinal information, oblique information is indeed less useful in natural color image recognition. Through the exploration of the proposed model on image recognition, we gain more understanding of the inhomogeneity of visual orientation. It also illuminates a wide range of opportunities for integrating the inhomogeneity of visual orientation with other computational models.**

Keywords-orientation inhomogeneity; convolutional neural networks; image recognition; oblique effect; cognitive modeling

I. INTRODUCTION

The most characteristic finding of the discrimination of orientation is a preference for vertical or horizontal orientations over obliques. This preference manifests itself in adjustment and assessment of stimulus orientation, the resolution of targets, learning and discrimination of objects, and a wide assortment of other perceptual phenomena [1]. This "oblique effect" phenomenon has been documented by different research studies. As early as 1893, the experimental responses from [2] evidenced marked superiority with horizontal and vertical stimuli in the experiments of reproducing visually presented lines. Studies of visual acuity were also among the first investigations to uncover preferences for vertical and horizontal stimuli. Emsley found acuity differences among subjects asked to resolve line gratings [3]. Furthermore, several studies tried to examine the role of stimulus orientation in perceptual grouping [4][5]. Their results also suggested that the most facilitating aspect of perceptual grouping was the change from horizontal or vertical orientations to diagonal orientations [1]. Optical illusions are also subject to the oblique effect. This kind of illusion is minimal when the lines are in a horizontal or vertical position and maximal for obliques' position [6]. As well as adults, children are also susceptible to the oblique effect. Bryant tested 5-7-year-olds children on simultaneous and successive matching tasks and found successive discrimination of mirror-image obliques to be most difficult [7]. In addition to humans, species as diverse as octopuses, goldfish, rats, cats, and chimpanzees show the oblique effect to some degree [8].

Neurophysiological research has firmly established the presence of cells in the cortex of cat, monkey, and man that are selectively sensitive to orientation. Ganglion cells in rabbit, pigeon, and goldfish have been found that perform similar functions [1]. Neural processing of contours was highlighted by the classical research by Hubel and Wiesel

2375-0197/16 $31.00 © 2016 IEEE
DOI 10.1109/ICTAI.2016.76

which revealed neural units right at the entrance of visual signals into the brain that respond preferentially to lines and edges [9]. When the distribution of preferred orientation of these units was examined, there were fewer in the oblique meridians than in the vertical and horizontal [10]. Orientation differences also occur in testing the visual brain for cell connectivity [11] and with imaging techniques [12].

Although edges or contours in our visual environment obviously are distributed across the full range of orientations, it is possible that our visual system has been biased functionally and structurally by a predominance of visible information near the cardinal axes [13]. In early studies from [14][15], Fourier analysis was used to analyze the natural scenes. They found a variety of scenes have anisotropic frequency spectra, with more power near the cardinal axes. Coppola et al. reported the analysis of many digitized scenes using Sobel direction and magnitude filters [13]. Their results showed a prevalence of vertical and horizontal orientations in indoor, outdoor, and even entirely natural setting. Girshicketal et al. obtained their measurements of probability distribution over local orientation from a larger database of photographs of scenes [16]. They used a pair of rotation-invariant filters to estimate the local image gradients. After this stage, they identified strongly oriented regions, obtained dominant orientation and calculated the histograms of these values. In their results, the estimated environmental distribution indicates there existing a predominance of cardinal orientation. In 2015, we defined the environmental orientation distribution as the probability distribution over local orientation with different spatial scale [17]. A standard dataset is utilized to statistically analyze the orientation distribution on thousands of authentic images with eight semantically organized categories. We used Canny edge detector [18] to obtain the edge map of every image. The local image gradients were calculated based on the edge map. Then, the orientation histogram channels were created based on the gradient orientation values. The resulting estimated environmental distribution indicated a predominance of vertical and horizontal orientations in various urban and natural scene images. Thus, it has been suggested that the prevalence of vertical and horizontal orientations in the environment is the underlying cause of the anisotropy of orientation discriminability.

Although the “oblique effect” has an impact on psychology, little research integrated this effect into computational models. In this paper, we seek to explore this inhomogeneity of visual orientation based on CNNs in image recognition. As we known, CNNs were inspired by biological processes. In this paper, we select them as our basic model. Moreover, they have wide applications in image and video recognition.

The remainder of this paper is organized as follows. Related work on Convolutional neural networks is reviewed in Section II. Visual orientation inhomogeneity based CNNs are introduced in Section III. Section IV discusses the performance of the proposed techniques in image recognition and Section V concludes this paper.

II. Related Work on Convolutional Neural Networks

In recent years, there’s been a resurgence in the field of Artificial Intelligence. This resurgence has been powered in no small part by a new trend in AI, especially in machine learning, known as “deep Learning”. Different from shallow learning models, such as Support vector machine (SVM), deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction [19]. They exploit the property that a lot of natural signals are compositional hierarchies, that means higher-level features can be obtained by composing lower-level ones. In images, for example, edges form motifs, motifs assemble into parts, and parts form the whole objects. The similar phenomenon also happens in speech, text, or other multimedia data. Some theoretical analyses from machine learning provide support for the argument that deep models are more compact and expressive than shallow models in representing most learning functions, especially highly variable ones [20]. Thus, deep learning methods could be used to solve some optimal problems, even possible in some traditional applications [21].

However, increasing the number of hidden layers leads to two known issues: vanishing gradients and overfitting. Backpropagation, a well-known computationally efficient model for multilayer neural networks, also suffers from the problems of insufficient labeled data, high computational cost, and poor local optima when working under a deep model [22]. To reduce the difficulty, more recent researches have been devoted to investigate new learning algorithms for deep architectures, such as Deep belief networks (DBN) [23], Convolutional neural networks [24][25], Recurrent neural networks [26][27], and so on.

Convolutional neural networks (CNNs) are one type of feed-forward artificial neural networks [24][25]. Four key ideas are behind CNNs that take advantage of the properties of natural signals: local connections, shared weights, pooling and the use of many layers. The architecture of the typical CNNs can be structured as a series of stages. In their stages, the first stages are composed of two types of layers: convolutional layer and pooling layer. The role of the convolutional layer is to detect local conjunctions of features from the previous layer, and the role of the pooling layer is to merge semantically similar features into one. In the two or three stages of convolution, the non-linearity and pooling are stacked, and they are followed by more convolutional and fully-connected layers. In the last stage, the backpropagating gradients are calculated through Convolutional Neural Networks, which is as same as through a regular deep neural network, allowing all the weights to be trained. The convolutional and pooling layers in CNNs are directly inspired by the classic notions of simple cells and complex cells in visual neuroscience [28], and the overall architecture is reminiscent of the hierarchy in the visual cortex ventral pathway [29]. CNNs achieved many practical successes during the period when neural networks were out of favour and it has recently been widely adopted by the computer-

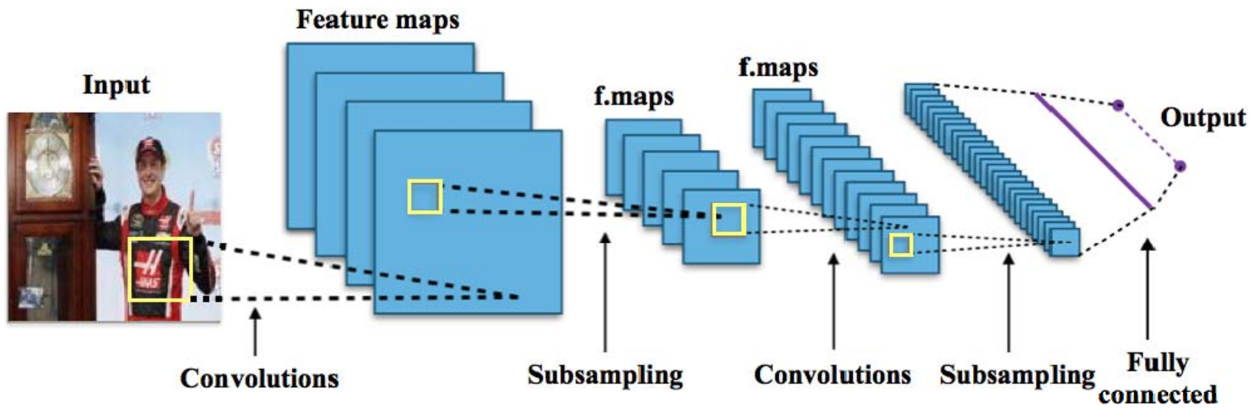

Figure 1. Typical CNNs architecture.

vision community, such as face verification [30], image segmentation [31], and object recognition [32][33]. Additionally, features learned by large networks trained on ImageNet have been shown to yield state-of-the-art performance across many standard image recognition datasets when classified with an SVM, even with no fine-tuning. A number of works have focused on understanding the representation learned by CNNs. Zeiler & Fergus have introduced a procedure to visualize what activate each unit [34]. Yosinski et al. have used transfer learning to measure how generic/specific the learned features are [35]. Unfortunately, there is no existing work try to explore the effects of the inhomogeneity of visual orientation with CNNs. In this paper, we try to investigate the performance of the orientation inhomogeneity with CNNs on image recognition task.

III. Basic Idea of Visual Orientation Inhomogeneity based CNNs

Convolutional neural networks combine three architectural ideas to ensure some degree of shift, scale, and distortion invariance: local receptive fields, shared weights, and spatial or temporal sub-sampling [25]. A typical convolutional network is shown in Fig. 1.

The convolutional layer is the core building block of convolutional neural networks. The layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field. During the forward pass, each filter is convolved across the input volume, computing the dot product between the entries of the filter and the input, and producing a 2-dimensional activation map of that filter. As a result, the network could learn filters that activate when they see some specific type of feature at some spatial position in the input. Stacking the activation maps for all filters forms the output volume of the convolution layer.

In this paper, by integrating the orientation inhomogeneity into our model, we propose a novel algorithm, Visual orientation inhomogeneity based CNNs (V-CNNs). In standard CNNs, each learnable convolutional kernel is convolved across the input volume to extract and process the information. And information from all orientations are extracted and processed equally. On the contrary, the convolution kernels in V-CNNs aim to omit the information of the oblique orientation in convolutional stages of the standard computation. In Fig. 2, we demonstrate a simple implementation of the convolution kernels with the size 5×5 in different models, including: standard CNNs, Visual orientation inhomogeneity based CNNs (V-CNNs), Double visual orientation inhomogeneity based CNNs (DV-CNNs), and Opposite visual orientation inhomogeneity based CNNs (OV-CNNs). The white color indicates the value in the corresponding position of the convolution kernels is set to be 0.

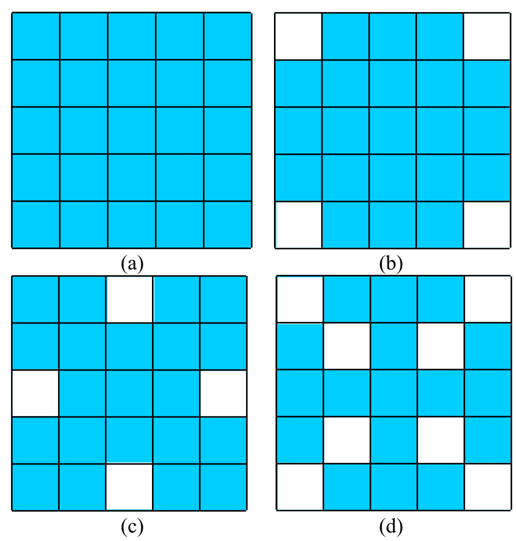

Figure 2. The convolution kernels in different models. (a). standard CNNs; (b). Visual orientation inhomogeneity based CNNs (V-CNNs); (c). Opposite visual orientation inhomogeneity based CNNs (OV-CNNs); (d). Double visual orientation inhomogeneity based CNNs (DV-CNNs). The white color indicates the value in the corresponding position of the convolution kernels is set to be 0.

In Table 1, we provide the computational complexity comparisons of different convolution kernels. In standard CNNs, if the size of the convolution kernel is 5×5, for each kernel, the number of multiplication operations per image is 25, and the number of addition operations per image is 24. In V-CNNs and DV-CNNs, both of the number of multiplication operations and the number of addition operations decrease. Although the difference of the proposed model and the standard CNNs is not large, taking into consideration the large number of epochs and the large number of images in large-scale image recognition, this difference really matters.

TABLE I. COMPUTATIONAL COMPLEXITY COMPARISONS OF DIFFERENT CONVOLUTION KERNELS WITH THE SIZE 5×5

Algorithms	CNNs	V-CNNs	DV-CNNs
Number of multiplication operations per image	25	21	17
Number of addition operations per image	24	20	16

IV. EXPERIMENTS

A. Experimental Setting

In our experiments, we try to investigate the recognition performance of the proposed V-CNNs and DV-CNNs, with the standard CNNs. And we also conduct comparisons experiments with the control model: OV-CNNs, whose cardinal feature detectors are omitted. In our experiments, CNNs, V-CNNs, DV-CNNs, and OV-CNNs are trained using MatConvNet toolbox (version 1.0-beta18) [36] on a Tesla K80 GPU. MatConvNet allows fast prototyping of typical or new CNN architectures, meanwhile, it supports efficient computation on CPU and GPU allowing training complex models on large datasets such as ImageNet ILSVRC-2012. We follow the general parameter setting in MatConvNet, such as the number of layers, the number of filters in each layer, the learning rate, and so on.

Our experiments are evaluated on three datasets, including: MNIST, CIFAR-10, and ImageNet Large-Scale Visual Recognition Challenge 2012 (ILSVRC-2012). In the first experiment, we test the visual orientation inhomogeneity on handwritten digits recognition. In the second experiment, we do the experiments on natural color image dataset. In the third experiment, we extend our comparisons on large-scale hierarchical object recognition.

B. Experiment 1: Visual Orientation Inhomogeneity on Handwritten Digits Recognition

MNIST is a standard database of handwritten digits containing a training set of 60,000 examples and a test set of 10,000 examples with 10 classes [25]. The resolution of images is 28 × 28. MNIST is a commonly used standard dataset for evaluating the performance of deep learning techniques [37]. Sample images in this dataset are shown in Fig. 3. The handwritten digits recognition is one basic but important task in computer vision. On this dataset, we try to test the effect of the inhomogeneity of visual orientation in handwritten digits recognition by evaluating the proposed V-CNNs. In this experiment, we used the MatConvNet library [36] and their reference implementation of LeNet [25].

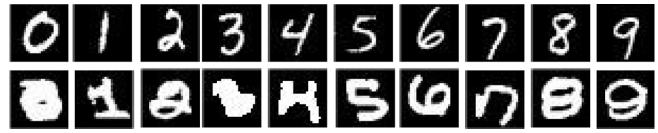

Figure 3. Sample images in MNIST handwritten digits.

The handwritten digits recognition experiments were done based on CNNs, V-CNNs, and DV-CNNs. Here, V-CNNs and DV-CNNs are proposed based on the orientation inhomogeneity in human vision. In V-CNNs, some information from the oblique orientations has been removed as Fig. 2(a). In DV-CNNs, double information from the oblique orientations has been removed as Fig. 2(d). On the contrary, classical CNNs keep extracting and processing the information from all orientation equally. The average recognition accuracies, standard deviations are given in Table 2. “Acc.” stands for average accuracy, and “Std.” stands for standard deviation. All the statistical experiments on image datasets are repeated for 5 times with randomly selected training sets and the average results are reported. The results suggest that the number of correct recognition of V-CNNs is slightly better than that of standard CNNs, although these two models are not significantly different in a paired t-test ($t(8) = 1.26$, $p = 0.242$). Moreover, if more oblique information were sacrificed in the learning procedure as DV-CNNs, the recognition performance would decrease a little.

TABLE II. PERFORMANCE COMPARISON IN HANDWRITTEN DIGITS RECOGNITION ON MNIST

Algorithms	CNNs	V-CNNs	DV-CNNs
Acc. (%)	99.09	99.11	99.07
Std. (%)	0.003	0.004	0.005

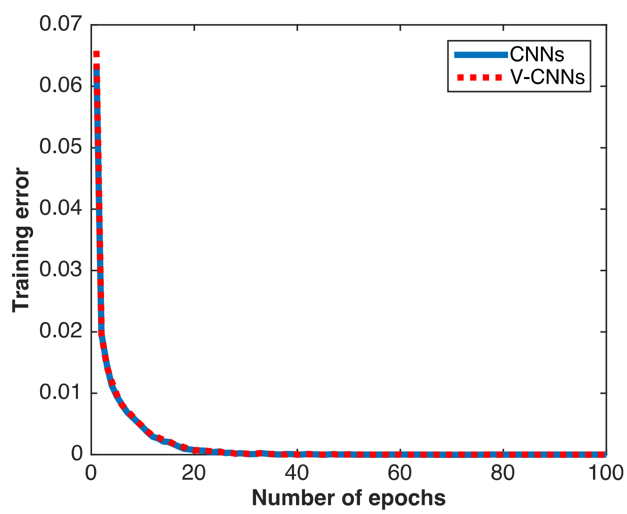

Figure 4. Convergence curve of training error of CNNs and V-CNNs.

Fig. 4 shows the convergence curve of training error of CNNs and V-CNNs. This experiment is conducted on a PC with Intel(R) Core(TM) I7-2600 3.4GHz CPU and 8.0GB

RAM. We could find that the shapes of their curves are very similar with each other. It means sacrificing the oblique information to a certain degree will not delay the learning procedure of CNNs.

C. *Experiment 2: Visual Orientation Inhomogeneity on Natural Color Image Recognition*

The CIFAR-10 dataset consists of 60,000 images with the resolution of 32×32 from 10 classes (6,000 images per class) [38]. There are 50,000 training images and 10,000 test images. This dataset includes ten categories, namely “airplane”, “automobile”, “bird”, “cat”, “deer”, “dog”, “frog”, “horse”, “ship”, and “truck”. Sample images of each category are shown in Fig. 5. The CIFAR-10 images are highly varied, and there is no standard viewpoint or scale at which the objects appear. Moreover, CIFAR-10 is also a commonly used standard dataset for evaluating the performance of deep learning algorithms [37][39]. In this experiment, we used the MatConvNet library [36] and their reference implementation of LeNet [25].

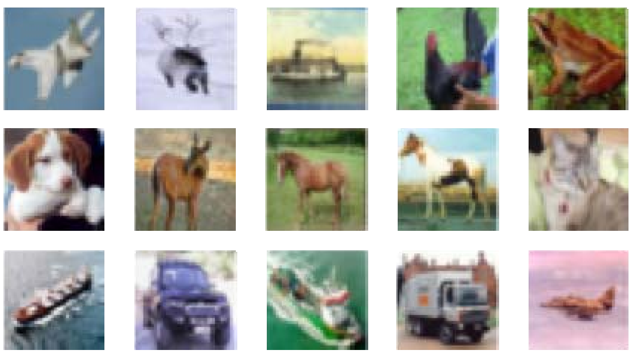

Figure 5. Sample images in CIFAR-10 dataset.

In this experiment, we explore the performance of CNNs, V-CNNs, DV-CNNs, and OV-CNNs on natural color image recognition. Here, V-CNNs and DV-CNNs are proposed based on the orientation inhomogeneity in human visual perception as Fig. 2(b) and Fig. 2(d). On the contrary, in OV-CNNs, information from the cardinal orientations has been removed as Fig. 2(c).

TABLE III. PERFORMANCE COMPARISON IN HANDWRITTEN DIGITS RECOGNITION ON CIFAR-10

Algorithms	CNNs	V-CNNs	DV-CNNs	OV-CNNs
Acc. (%)	76.49	75.32	74.78	74.66
Std. (%)	0.32	0.24	0.25	0.15

The average recognition accuracies, standard deviations are given in Table 3. “Acc.” stands for average accuracy, and “Std.” stands for standard deviation. All the statistical experiments on image datasets are repeated for 5 times with randomly selected training sets and the average results are reported. From this table, we could find standard CNNs achieve best accuracy. And the performance of V-CNNs is better than DV-CNNs and OV-CNNs. CNNs and V-CNNs are significantly different in a paired t-test ($t(8) = 6.52$, $p < 0.01$). It means in a complex task environment, such as natural color image recognition, the oblique information also plays some roles. V-CNNs and DV-CNNs are significantly different in a paired t-test ($t(8) = 3.47$, $p < 0.01$). And DV-CNNs and OV-CNNs are not significantly different in a paired t-test ($t(8) = 0.90$, $p < 0.392$). It means compared with the cardinal information, in natural color image recognition, oblique information is indeed less useful.

D. *Experiment 3: Visual Orientation Inhomogeneity on Large-Scale Hierarchical Object Recognition*

ImageNet is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories [40]. The images from ImageNet were collected from the web and labeled by human labelers using Amazon’s Mechanical Turk crowd-sourcing tool. In the experiment, we use the dataset of the ImageNet Large-Scale Visual Recognition Challenge 2012 (ILSVRC-2012). ILSVRC-2012 is a subset of ImageNet with about a million images that contained 1,000 different categories. There are roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images. Sample images in this dataset are shown in Fig. 6. What’s more, ImageNet is a well-known dataset to evaluate the capability of a deep learning algorithm [41] [42]. In this experiment, we use the MatConvNet toolbox [36] and their reference implementation of AlexNet [41].

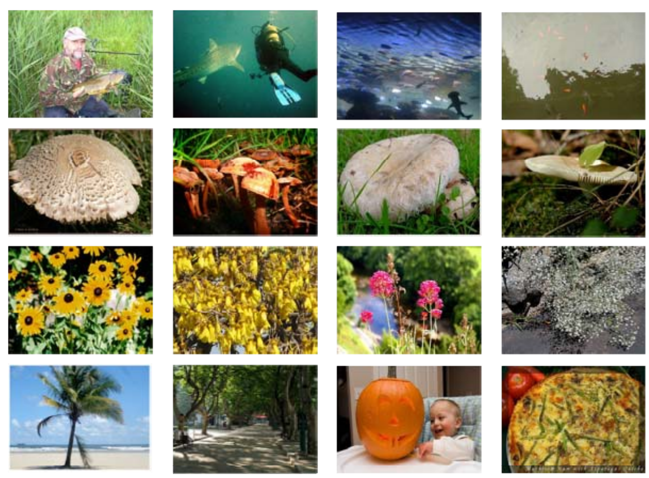

Figure 6. Sample images in ILSVRC-2012 image dataset.

We construct the proposed V-CNN model by referring to the well-known AlexNet architecture. In AlexNet, the first convolutional layer filters the 227×227×3-dimensional input image with 96 kernels of size 11×11×3 with a stride of 4 pixels. The second convolutional layer takes the output of the first convolutional layer as input and filters it with 256 kernels of size 5×5×48. The third convolutional layer has 384 kernels of size 3×3×256 connected to the outputs of the second convolutional layer. The fourth convolutional layer has 384 kernels of size 3×3×192. The fifth convolutional layer has 256 kernels of size 3×3×192. In the end, the fully-connected layer have 4096 neurons each. In AlexNet, each 11×11×3 learnable convolutional kernel is convolved across the input image to extract and process the information. And information from all orientations are extracted and processed equally. On the contrary, the convolution kernels in the proposed V-CNNs aim to only extract and process the information from cardinal orientation. Thus, the values of the corresponding position from oblique orientation in the convolution kernels are set to be 0. As the first convolution kernel in AlexNet is of size 11×11, it is almost twice the size of 5×5. Similarly as the convolution kernel shown in Fig. 2(b), some information from the oblique orientations in each corner of the kernel has been removed while they are size of 2×2 instead of 1×1. The convolution kernels in V-CNNs used for large-scale hierarchical object recognition are shown in Fig. 7.

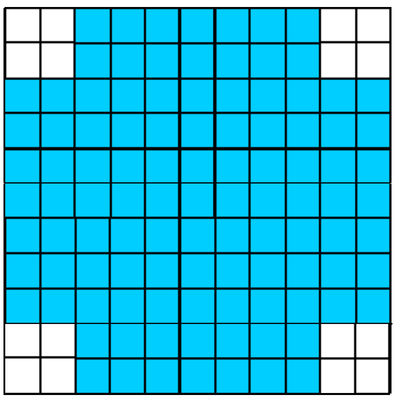

Figure 7. The convolution kernels of V-CNNs for large-scale hierarchical object recognition. The white color indicates the value in the corresponding position of the convolution kernels is set to be 0.

In this experiment, all the statistical experiments on image datasets are repeated for 5 times with randomly selected training sets and the average results across the 25 epochs are reported. The average recognition accuracy of standard CNNs is 55.40% and the value of V-CNNs is 51.43%. Moreover, the standard deviations of standard CNNs and V-CNNs are 0.21% and 0.11% respectively. Therefore, we could find in this case, the recognition performance of CNNs is better than the proposed V-CNN. These results indicate that in a large-scale recognition task, the oblique information is also important.

V. Conclusions and Future Work

Motivated by the insight of orientation inhomogeneity in human vision, we proposed novel neural networks, V-CNNs, for image recognition. In standard CNNs, the convolutional layers extract and process the information from all orientation equally. On the contrary, V-CNNs omit the information of the oblique orientation in convolutional stages of the standard computation.

In this paper, we evaluate the proposed model V-CNNs on three datasets, including: MNIST, CIFAR-10, and ImageNet Large-Scale Visual Recognition Challenge 2012 (ILSVRC-2012). In the first experiment, we test the visual orientation inhomogeneity on handwritten digits recognition. In the second experiment, we do the experiments on natural color image dataset. In the third experiment, we extend our comparisons on large-scale hierarchical object recognition. In our experiments, V-CNNs achieve comparable performance with higher computational efficiency in recognition task on MNIST and CIFAR-10. Based on the comparisons between the proposed model V-CNNs, DV-CNNs and the control model OV-CNNs, we could get the conclusion that, compared with the cardinal information, in natural color image recognition, oblique information is indeed less useful. In the recognition task on the large-scale dataset ILSVRC-2012 with hierarchical objects, the performance of CNNs is better than the proposed V-CNN, and the difference is not big. This result indicates that in a large-scale recognition task, the oblique information has some effects on the recognition performance.

We have already shown that the proposed algorithm is applicable to image recognition task in our experiments. Actually, the proposed algorithm requires smaller storage capacity and better efficiency, which makes it potentially suitable for industry application where time or space complexity is more important, such as the image search engines.

We can observe that the recognition accuracy of the proposed method is relatively lower than the original method in large-scale hierarchical object recognition task. We infer it is because there exists differences of the orientation distribution in different object category. Hence, how to improve the adaptability of the proposed method by automatically adjusting the weights of the oblique orientation according to the orientation distribution of different categories is the first future work we need to consider. Another meaningful future work is to improve the efficiency of the algorithm in order to make sure the current algorithm can be transplanted on the portable devices. Last but not least, we would like to integrate the inhomogeneity of visual orientation into other local classical and the state-of-the-art algorithms, such as ResNet [43], and apply them in other applications.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61502311, No. 61373122, No. 61602314), Natural Science Foundation of Guangdong Province (No. 2016A030310053, No. 2016A030313043), Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase), the Science and Technology Innovation Commission of Shenzhen under Grant (No. JCYJ20150324141711640, No. JCYJ20150324141711630, No. JCYJ20160422151736824, No. JCYJ20160226191842793), Strategic Emerging Industry Development Foundation of Shenzhen (No. JCYJ20130326105637578), the Shenzhen University research funding (201535), and the Tencent Rhinoceros Birds Scientific Research Foundation (2015).

REFERENCES

[1] S. Appelle, “Perception and discrimination as a function of stimulus orientation: the 'oblique effect' in man and animals,” Psychological Bulletin, 78(4), 1972, pp. 266-278.

[2] J. Jastrow, “On the judgment of angles and positions of lines,” American Journal of Psychology, 5, 1893, pp. 214-248.

[3] H. H. Emsley, “Irregular astigmatism of the eye: effect of correcting lenses,” Transaction of the Optical Society, 27(1), 1925, pp. 28-41.

[4] J. Beck, “Perceptual grouping produced by changes in orientation and shape,” Science, 154, 1966, pp. 538-540.

[5] J. Beck, “Perceptual grouping produced by line figures,” Perception & Psychophysics, 2, 1967, pp. 491-495.

[6] C.H. Judd, “The Zöllner illusion,” Psychological Monographs, 7, 1905, pp. 112-141.

[7] P. E. Bryant, “Perception and memory of the orientation of visually presented lines by children,” Nature, 244, 1969, pp. 1331-1332.

[8] I. P. Howard, “Human visual orientation,” Wiley, NY, 1982.

[9] D. H. Hubel, & T. N. Wiesel, “Receptive fields of single neurons in the cat’s striate cortex,” Journal of Physilogy, 148, 1959, pp. 574-591.

[10] B. Li, M. R. Peterson, & R. D. Freeman, “Oblique effect: a neural basis in the visual cortex,” Journal of Neurophysiology, 90, 2003, pp. 204-217.

[11] W. Li, & C. D. Gilbert, “Global contour saliency and local conlinear interactions,” Journal of Neurophysiology, 88(5), 2002, pp. 2846-2856.

[12] C. S. Furmanski, & S. A. Engel, “An oblique effect in human primary visual cortex,” Nature Neuroscience, 3, 2000, pp. 535-536.

[13] D. M. Coppola, H. R. Purves, A. N. Nccoy, & D. Purves, “The distribution of oriented contours in the real world,” Proceedings of the National Academy of Sciences of the United States of America, 1998, pp. 4002-4006.

[14] D. J. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” Journal of the Optical Society of America A Optics & Image Science, 4, 1988, pp. 2379-2394.

[15] G. J. Burton, & I. R. Moorhead, “Color and spatial structure in natural scenes,” Applied Optics, 26, 1987, pp. 157-170.

[16] A. R. Girshick, M. S. Landy, & E. P. Simoncelli, “Cardinal rules: visual orientation perception reflects knowledge of environmental statistics,” Nature Neuroscice, 14, 2011, pp. 926-932.

[17] S. H. Zhong, Y. Liu, & Q. C. Chen, “Visual orientation inhomogeneity based scale-invariant feature transform,” Expert Systems with Application, 42(13), 2015, pp. 5658-5667.

[18] J. Canny, “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6), 1986, pp. 679-698.

[19] Y. LeCun, Y. Bengio, & G. E. Hinton, “Deep learning,” Nature, 521, 2015, pp. 436-444.

[20] Y. Bengio, A. Courville, & G E. Hinton, “Representation learning: a review and new perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence,” 35 (8), 2013, pp. 1798-1828.

[21] J. Li, M.K. Qiu, Z. Ming, G. Quan, X. Qin, Z. Gu, “Online optimization for scheduling preemptable tasks on IaaS cloud systems. Journal of Parallel and Distributed Computing,” 72(5), 2012, pp. 666-677.

[22] G. E. Hinton, “Learning multiple layers of representation,” Trends in Cognitive Science, 11(10), 2007, pp. 428-4342.

[23] G. E. Hinton, & R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, 313(5786), 2006, pp. 504-507.

[24] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, & L. D. Jackel, “Handwritten digit recognition with a back-propagation network,” Proceedings of Advances in Neural Information Processing Systems, San Francisco, CA: Morgan Kaufmann Publishers Inc., 1990, pp. 396-404,

[25] Y. LeCun, L. Bottou, Y. Bengio, & P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of IEEE, 1998, pp. 2278-2324.

[26] J. Schmidhuber, “Learning complex, extended sequences using the principle of history compression,” Neural Computation, 4(2), 1992, pp. 234-242.

[27] S. Fernandez, A. Graves, & J. Schmidhuber, “Sequence labeling in structured domains with hierarchical recurrent neural networks,” Proceedings of the 20th International Joint Conference on Artificial Intelligence, San Francisco, CA: Morgan Kaufmann Publishers Inc.2007, pp. 396-404.

[28] D. H. Hubel, & T. N. Wiesel, “Receptive fields, binocular interaction, and functional architecture in the cat’s visual cortex,” Journal of Physiology, 160, 1962, pp. 106-154.

[29] D. J. Felleman, & D. C. V. Essen, D. C. V. “Distributed hierarchical processing in the primate cerebral cortex,” Cerebral Cortex, 1, 1991, pp. 1-47.

[30] Y. Taigman, M. Yang, M. Ranzato, & L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH: IEEE, 2014, pp. 1701-1708.

[31] R. Girshick, J. Donahue, T Darrell, & J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (pp. 580-587). Columbus, OH, pp. 580-587, 2014.

[32] K. He, X. Zhang, S. Ren, & J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9), 2015, pp. 1904-1916.

[33] S. Hassairi, R. Ejbali, M. Zaied, “Supervised image classification using deep convolutional wavelets network,” IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), 2015, pp. 265-271.

[34] M. D. Zeiler, & R. Fergus, “Visualizing and understanding convolutional network,” Proceedings of the 2014 European Conference on Computer Vision, Zurich, Switzerland: Springer, 2014, pp. 818-833.

[35] J. Yosinski, J. Clune, Y. Bengio, & H. Lipson, “How transferable are features in deep neural networks,” Proceedings of the

Advances in Neural Information Processing Systems, Montréal, CA, 2014, pp. 3320-3328.

[36] A. Vedaldi, & K. Lenc, “MatConvNet - Convolutional neural networks for Matlab,” arXiv preprint, arXiv: 1412.4564, 2015.

[37] D. Ciresan, U. Meier, U., & J. Schmidhuber, “Multi-column deep neural networks for image classification,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC: IEEE, 2012, pp. 3642-3649.

[38] A. Krizhevsky, & G. E. Hinton, “Learning multiple layers of features from tiny images,” Master’s thesis, Computer Science Department, University of Toronto, Ontario, 2009.

[39] D. Yu, H. Wang, P. Chen, & Z. Wei, “Mixed pooling for convolutional neural networks,” Rough Sets and Knowledge Technology, 2014, pp. 364-375.

[40] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, & F. F. Li. “ImageNet: A large-scale hierarchical image database,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL: IEEE, pp. 248-255, 2009.

[41] A. Krizhevsky, I. Sutskever, & G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Proceedings of Advances in Neural Information Processing Systems 25, Lake Tahoe, Nevada, 2012, pp. 1106-1114.

[42] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, “Going deeper with convolutions,” arXiv preprint, arXiv:1409.4842, 2014.

[43] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1-12.