Research Article

Night-Time Vehicle Detection Algorithm Based on Visual Saliency and Deep Learning

Yingfeng Cai,[1] **Xiaoqiang Sun,**[2] **Hai Wang,**[2] **Long Chen,**[1] **and Haobin Jiang**[2]

[1]*School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013, China*
[2]*Automotive Engineering Research Institute, Jiangsu University, Zhenjiang 212013, China*

Correspondence should be addressed to Hai Wang; wanghai1019@163.com

Night vision systems get more and more attention in the field of automotive active safety field. In this area, a number of researchers have proposed far-infrared sensor based night-time vehicle detection algorithm. However, existing algorithms have low performance in some indicators such as the detection rate and processing time. To solve this problem, we propose a far-infrared image vehicle detection algorithm based on visual saliency and deep learning. Firstly, most of the nonvehicle pixels will be removed with visual saliency computation. Then, vehicle candidate will be generated by using prior information such as camera parameters and vehicle size. Finally, classifier trained with deep belief networks will be applied to verify the candidates generated in last step. The proposed algorithm is tested in around 6000 images and achieves detection rate of 92.3% and processing time of 25 Hz which is better than existing methods.

1. Introduction

On average at least one person dies in a vehicle collision accident per minute globally. In addition, the accidents cause injuries to nearly ten million people every year and serious injuries to thirty percent of them. To solve this problem, Advanced Driver Assistant Systems (ADAS) including lane departure warning, forward collision warning, parking assistance systems, and night vision enhancements have became popular in research and development of automobile industry.

Vision sensor is the most popular sensors used in ADAS [1–3]. Based on this sensor, many applications can be achieved such as vehicle detection, lane detection, traffic sign recognition, and pedestrian detection. Among those applications, vehicle detection has drawn much attention of many researchers. There are three kinds of methods for object detection field which are monocular vision, stereo vision, and motion based methods. In monocular vision based vehicle detection, Sivaraman models it as a two-class problem and uses Haar feature and cascaded Adaboost algorithm to recognize image patch [4]. In stereo vision based method, Hermes proposes a vehicle detection method based

on density map clustering [5]. Besides, motion based method has also been used for overtaking vehicle detection [6].

Most of the existing vehicle detection systems and methods mainly focus on vehicle detection on daytime light conditions with traditional optical sensor. However the statistics show that more than half of the traffic accidents cause fatalities to occur at night. There are some researchers who have put effort into night-time vehicle detection with visible spectrum optical sensors; the detection results are often affected by many factors such as low illumination, light reflection on rainy days, and the camera exposure time [7]. Recently, another kind of optical sensor named as far-infrared (FIR) camera is becoming more and more frequently used in automotive field for night vision application. The biggest benefit for FIR is that it does not require any illumination source and relies purely on heat signatures from the environment to produce a grey-scale image. For night-time vehicle detection tasks, FIR camera can detect the heat signal generated by heat sources such as vehicle tire, engine, and exhaust pipe. A group of sample images were taken from visible spectrum optical camera and FIR camera is shown in Figure 1. It can be found that the target vehicle is very hard to

(a)   (b)

FIGURE 1: Night views using visible spectra (a) and FIR spectrum cameras (b).

be identified in visible spectrum optical camera due to the big light spot while the target vehicle is easy to be found in FIR image.

Nowadays, many automotive companies can provide FIR image to drivers for night-time driving, but they all lack the ability to interpret image automatically and the driver needs to view the image and understand it all by themselves. There are also not many reports that have been presented in the field of FIR image based vehicle detection. In 2000, Dr. Andreone from University of Parma developed a demo car with FIR camera. In this system, the function of night-time vehicle detection relies on the size of the image area of high brightness, shape, and texture information [8]. Machine learning based framework then became popular in object detection tasks. In 2010, Besbes et al. introduce it to vehicle detections in which a SURF feature and SVM classifier based method are proposed [9]. Similarly, this kind of machine learning based framework is also used in pedestrian detection in FIR image [10, 11]. Feature selection and classifier design are two critical components for object detection. The most common two-class classification method now is handcrafted feature and shallow model based method such as HOG feature and SVM classifier. However, the abilities of those hand-crafted features and complex function descriptions of shallow models are both insufficient for vehicle classification in complex traffic environments in FIR image. Besides, the vehicle detection processing time is also not able to satisfy ADAS application. Luckily, a new machine learning method named as deep learning is proposed by scientists recently. The concept of deep leaning is to convert features from lower layer to higher layer to automatically learn more compact features and improve classification or predicting problems. On the other hand, different from those shallow models, deep learning is with the ability of learning multiple layers of representation and abstraction that helps to make better understanding of image data. As a novel branch of machine learning, deep learning technology draws much concern in industry and academic field recently. The recent research of our work which uses vehicle contour based deep learning approach performs a significant improvement compared with traditional shallow model based method [12]. Meanwhile, it is found that human drivers can identify target such as pedestrians in road image or video very fast. The reason is that human has the ability of visual saliency which means that the object with strong visual significance can be found fast. So,
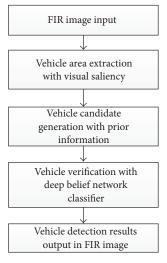

FIGURE 2: Flow of the proposed algorithm.

inspired by human attention mechanism, the saliency model can be used to decrease processing time.

Based on the analysis above, integrating deep leaning method and visual saliency, this work proposed a new vehicle detection algorithm in FIR image. Firstly, most of the nonvehicle pixels will be removed with visual saliency computation. Then, vehicle candidate will be generated by using prior information such as camera parameters and vehicle size. Finally, classifier trained with deep belief networks will be applied to verify the candidates generated in last step. The overall flow of the proposed method is shown in Figure 2.

Compared with our former work in [12], there are two significant improvements. Firstly, a visual saliency based vehicle candidate generation approach is proposed. With this framework, the vehicle candidate can be generated fast and accurately with less processing time. Secondly, the sub images are used to train the deep model in this work instead of the contour used in our former work. Since subimages contain much more rich information than contour, the training performance has a significant improvement.

2. Vehicle Candidate Generation

For vehicle detection tasks in image, they usually contain two steps which are vehicle candidate generation (VCG) and

FIGURE 3: Saliency extraction for a FIR image.

vehicle candidate verification (VCV) [13]. In VCG, all the image areas that are with small probability to be vehicles will be selected. In this step, prior knowledge of vehicles such as horizontal/vertical edges, symmetry, color, shadow, and texture may often be used. In VCV, the image areas selected in VCG will be further verified to eliminate that belonging to nonvehicles. In this step, a two-class classification framework is often used and a classifier that can distinguish vehicles from nonvehicles will be trained from a set of training images. In our work, we will also follow this two-step framework.

FIR image reflects the information of temperature in objects. Conversely, it cannot show the detailed information in visible domain such as texture, color, and shadows. Due to this, the existing vehicle detection algorithms used in visible spectrum camera are not suitable for far-infrared camera. With observation, it is found that there are many highlighted areas of vehicles in FIR image due to many hot parts in vehicles such as engine. Human can easily focus on interesting region of an image fast while traditional machine vision system needs to process whole image traversal. This concept is used to calculate the level of saliency of area and is named as visual saliency. For this application, in FIR image light vehicle area is easy to be viewed by observer and with high visual saliency. With this reason, the vehicle region will be extracted with visual salience firstly.

According to the human eye perceives significance features; researchers proposed saliency measurement operator in pixel level. One of the most typical saliency measurement operator is iNVT feature which is proposed by Itti et al. [14]. Based on that, Montabone proposed another VSF feature which improves the detailed information on extraction ability of iNVT [15]. In this work, the saliency value will be calculated with the VSF feature.

The brief introduction of VSF feature is shown as follows.

Set the coordinate of one pixel in a grey-scale image as (x, y) and its value is $i(x, y)$. The center and surrounding operator of this point are defined as

$$\text{center}(x, y) = i(x, y)$$
$$\text{surround}(x, y, \varsigma)$$
$$= \frac{\text{rectSum}(x - \varsigma, y - \varsigma, x + \varsigma, y + \varsigma) - i(x, y)}{(2\varsigma + 1)^2 - 1},$$

(1)

in which ς is multiscale factor $\varsigma \in \{1, 2, 4, 8, \ldots\}$ and rectSum means integral image.

FIGURE 4: Vehicle candidate generation in a FIR image.

After that, calculating saliency value Int_ς in each scale of each point

$$\text{VSF}_\varsigma(x, y)$$
$$= \max\{\text{center}(x, y) - \text{surround}(x, y, \varsigma), 0\}.$$

(2)

Then all the saliency value VSF_ς in different scale will be added together to get the final saliency value:

$$\text{VSF} = \sum_\varsigma \text{VSF}_\varsigma.$$

(3)

A saliency map calculated with VSF of a FIR road image which contain a vehicle is shown in Figure 3. It can be seen that vehicle and part of the sky area are given higher saliency value while the vast majority of nonvehicular pixels such as road are with low saliency value and have been rejected in this step.

When part of nonvehicle area is removed, the vehicle candidates will then be generated with the method of article [16]. In this method, the parameters of the camera will be used to maintain road vanish point and road area. Then the sliding windows to select vehicle candidates will be generated according to prior information such as vehicle width. Processing results of Figure 3 are shown in Figure 4. From this figure, it can be found that only the image patch contact with road surface is concerned as vehicle candidate and most light area belonging to sky is further rejected. With this step, the processing pressure of VCV is further reduced.

3. Vehicle Candidate Verification

In last section, the areas which may contain vehicles are extracted with the visual saliency calculation method we proposed. Then the vehicle candidates are further generated

with prior information such as road surface and vehicle width. In this section, deep learning based method will be used to verify the vehicle candidates generated in last step.

There are many types of deep architectures such as deep belief networks (DBN) and deep convolution neural networks (DCNN) which use DBN as a typical deep learning structure and is used in many tasks such as MNIST classification, 3D object recognition, and voice recognition. In our work, DBN is applied and a classifier is trained for vehicle candidate verification tasks.

3.1. Deep Belief Network (DBN) for Vehicle Candidate Verification.

In this subsection, the overall architecture of the DBN classifier for vehicle candidate verification will be firstly introduced.

Let X be the set of training samples which contain vehicle images and nonvehicle images which are generated manually by our group. X consists of K samples which are shown in the following:

$$X = [x_1, x_2, \ldots, x_k, \ldots, x_K]. \tag{4}$$

In X, X_k is a training sample and all samples are resized to $P \times Q$. Y represents the labels corresponding to X, which can be written as

$$Y = [y_1, y_2, \ldots, y_k, \ldots, y_K]. \tag{5}$$

In Y, y_k is the label vector of X_k. If X_k belongs to a vehicle, $y_k = (1, 0)$. Otherwise, $y_k = (0, 1)$.

The purpose of the vehicle candidate verification task is to learn the mapping function from the training data X to the label data Y based on a given training set. With this trained mapping function, unknown images can be classified as either vehicle or nonvehicle.

In this task, a DBN architecture is applied to address this problem. Figure 5 shows the overall architecture of the DBN. It is a fully interconnected deep belief network including one visible input layer V^1, N hidden layers $H^1, \ldots, H^N$, and one visible label layer La at the top. The visible input layer V^1 maintains a $P \times Q$ neural number which is equal to the dimensions of the training feature, that is, the pixel values of the training samples. On the top, the La layer just has two states which can be either $(1, 0)$ or $(0, 1)$.

The learning process of the DBN has two main steps. In the first step, the parameters of the two adjacent layers will be refined with the greedy-wise reconstruction method. This step will be repeated until the parameters of all the hidden layers are fixed. This first step is also called the pretraining process. In the second step, the whole pretrained DBN will be fine-tuned with the La layer information based on back propagation. This second step can be considered to be the supervised training step.

3.2. Pretraining Method.

In this subsection, the training method of the whole DBN for vehicle candidate verification will be presented.

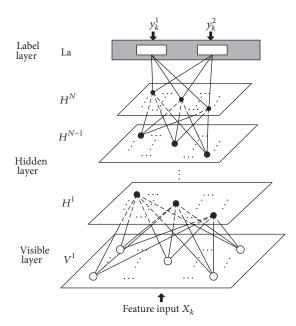
Figure 5: Proposed DBN for vehicle candidate verification.

Assume that the size of the upper layer is $P \times Q$. In this step, the parameters of the two adjacent layers will be refined using the greedy-wise reconstruction method proposed by Hinton et al. [17]. The visible input layer V^1 and the first hidden layer H^1 are taken here as an example; other adjacent layers use the same pretraining method.

The visible input layer V^1 and the first hidden layer H^1 contract a Restricted Boltzmann Machine (RBM). $I \times J$ is the neural number in V^1 and $P \times Q$ is that of H^1. The energy of state (v^1, h^1) in this RBM is

$$E(v^1, h^1, \theta^1) = -(v^1 A h^1 + b^1 v^1 + c^1 h^1)$$

$$= -\sum_{i=1, j=1}^{i \leq M, j \leq M} \sum_{p=1, q=1}^{p \leq P, q \leq P} v^1_{ij} A^1_{ij, pq} h^1_{pq} \tag{6}$$

$$- \sum_{i=1, j=1}^{i \leq M, j \leq M} b^1_{ij} v^1_{ij} - \sum_{p=1, q=1}^{p \leq P, q \leq P} c^1_{pq} h^1_{pq},$$

where $\theta^1 = (A^1, b^1, c^1)$ are the parameters between the visible input layer V^1 and the first hidden layer H^1, $A^1_{ij, pq}$ are the symmetric weights from input neural (i, j) in V^1 to the hidden neural (p, q) in H^1, and b^1_{ij} and c^1_{pq} are (i, j)th and (p, q)th bias of V^1 and H^1. So the RBM has the following joint distribution and Z is the normalization parameter:

$$P(v^1, h^1; \theta^1) = \frac{1}{Z} e^{-E(v^1, h^1; \theta^1)} = \frac{e^{-E(v^1, h^1; \theta^1)}}{\sum_{v^1} \sum_{h^1} e^{-E(v^1, h^1; \theta^1)}}. \tag{7}$$

Then, the conditional distribution over the visible input state v^1 in layer V^1 and the hidden state h^1 in H^1 can be given by the logistic function:

$$p\left(x_i^1\right) = p\left(h^1 \mid v^1\right) = \prod_{p,q} p\left(h_{pq}^1 \mid v^1\right),$$

$$p\left(h_{pq}^1 \mid v^1\right) = \sigma\left(\sum_{i=1,j=1}^{i \le M, j \le M} v_{ij}^1 A_{ij,pq}^1 + c_{pq}^1\right), \tag{8}$$

where $\sigma(x) = 1/(1 + \exp(-x))$.

The unsupervised training target function is

$$L = -\sum_{i=1}^{|D_{\text{unlab}}|} \log p\left(x_i^1\right), \tag{9}$$

in which $D_{\text{unlab}} = \{(x_i^1)\}_{i=1}^{|D_{\text{unlab}}|}$ is all unlabeled samples in lth layer.

For this target function, it can be trained with Contrastive Divergence (CD) algorithm proposed by Wood and Hinton [18].

3.3. Global Fine-Tuning. In this subsection, a traditional back propagation algorithm will be used to fine-tune the parameters $\theta = [A, b, c]$ using the information of the label layer La.

Since the pretraining process has already identified strong initial parameters, the backpropagation step is just used to finely adjust the parameters so that local optimum parameters $\theta^* = [A^*, b^*, c^*]$ can be found. At this stage, the learning objective is to minimize the classification error $[-\sum_t y_t \log \widehat{y}_t]$, where y_t and $\widehat{y}_t$ are the real label and the output label of data x_t in layer N.

Till this step, the deep classifier for vehicle detection can be maintained.

4. Experiment and Analysis

4.1. DBN Based Vehicle Verification Effect. The proposed DBN based vehicle verification method is trained on our image dataset captured by a SAT NV628 FIR camera as shown in Figure 6. The numbers of positive samples and negative samples for training are 5000 and 10000, respectively. Besides, another 5733 images are selected as testing samples.

By using the proposed method for vehicle verification in FIR image, four different architectures of 2D-DBN are applied. They all contain one visible layer and one label layer, but with two, three, four, and five hidden layers, respectively. In training, the critical parameters of the proposed 2D-DBN in experiments are set to $\alpha = 0.4$ and $\vartheta = 0.75$ and the image samples for training are all resized to 24×24 pixels which is equal to the size of training samples.

The detection of the four 2D-DBN architectures is compared with two shallow models (SVM and Adaboost) and the results are shown in Table 1. It is observed that the 2D-DBN with four hidden layers maintains the highest detection rate in the test set. It is also seen that deep architecture performs much better than the existing shallow models.

Figure 6: SAT NV628 FIR camera.

Table 1: Vehicle detection rates with different DBN structures and two shallow models.

Classifier types	Correct detection	Correct rate
2D-DBN (two hidden layers)	5057/5733	88.21%
2D-DBN (three hidden layers)	5250/5733	91.57%
2D-DBN (four hidden layers)	*5337/5733*	*93.09%*
2D-DBN (five hidden layers)	5291/5733	92.29%
HOG + SVM	4976/5733	86.80%
Haar + Adaboost	4792/5733	83.59%

Table 2: Overall vehicle detection result comparison.

Method	Detection rate (%)	False detection rate (per frame)	Time
Andreone et al. [8]	85.8	1.2	15 Hz
Besbes et al. [9]	91.4	0.9	13 Hz
Wang et al. [12]	91.8	0.7	18 Hz
Our framework	*92.3*	*0.3*	*25 Hz*

4.2. System Overall Effect. All the methods described below are tested using the same image dataset containing 5000 images captured by our group. Some of the vehicle candidate generation effects are shown in left column of Figure 7. Based on the vehicle candidate generation results, the DBN based vehicle candidate verification method is further applied. Some of the vehicle candidate verification results are shown in Figure 7. The left column shows identified vehicle candidates marked in red and the right column shows the verified vehicles marked with a blue rectangle.

The experimental platform of the experiment is HP work stations, the main parameters are Intel Core 2 Duo 2.67 G processor and 8 G memory and the operating system is Windows 7 while the programming software is Microsoft Visual 2008 + OPENCV2.3. The overall vehicle detection effects are shown in Table 2, as well as some state-of-the-art vehicle detection effects.

From the results shown in Table 2, it is seen that the proposed vehicle detection framework exhibits the lowest false detection rate while achieving the highest detection rate, which is 0.6% higher than that of our former method. Besides,
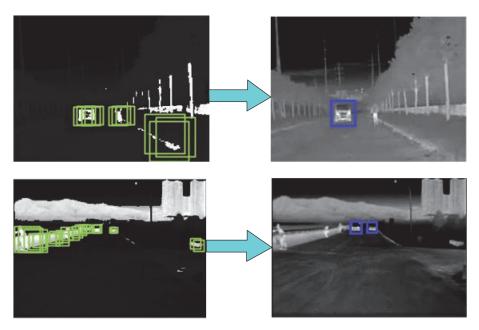
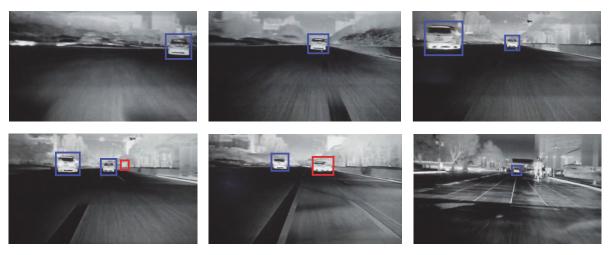FIGURE 7: Vehicle candidate verification results.

FIGURE 8: Vehicle sensing results in a continuous video.

importantly, the processing speed of proposed method is 25 Hz/frame which satisfies the real-time requirement of ADAS system.

Figure 8 shows a group of vehicle detection results in a continuous video. The blue rectangles represent correct vehicle detection and the red rectangles represent missed detection or false detection. It can be found from the vehicle detection results that most vehicles are correctly detected and pedestrians and bicycles are not falsely detected by our method while also a few bright blocks are misidentified as vehicles.

5. Conclusion

This work proposed a vehicle detection algorithm in FIR image based on visual saliency and deep learning. Firstly, most of the nonvehicle pixels will be removed with visual saliency computation. Then, vehicle candidate will be generated by using prior information such as camera parameters and vehicle size. Finally, classifier trained with deep belief networks will be applied to verify the candidates. Experiment demonstrates that this method achieves the highest vehicle detection rate compared with existing state-of-the-art methods and the processing time is below 40 ms per frame which satisfies requirements for real-time applications.

Competing Interests

The authors declare that they have no conflict of interests.
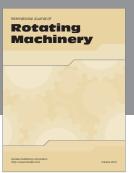
Acknowledgments

References

[1] B. Dorj and D. J. Lee, "A precise lane detection algorithm based on top view image transformation and least-square approaches," *Journal of Sensors*, vol. 2016, Article ID 4058093, 13 pages, 2016.

[2] R. Baran, A. Glowacz, and A. Matiolanski, "The efficient real- and non-real-time make and model recognition of cars," *Multimedia Tools and Applications*, vol. 74, no. 12, pp. 4269–4288, 2015.

[3] P. Hurney, P. Waldron, F. Morgan, E. Jones, and M. Glavin, "Review of pedestrian detection techniques in automotive far-infrared video," *IET Intelligent Transport Systems*, vol. 9, no. 8, pp. 824–832, 2015.

[4] J. Lu, F. Meng-Yin, W. Mei-Ling et al., "Vehicle detection based on vision and millimeter wave radar," *Journal of Infrared and Millimeter Waves*, vol. 33, no. 5, pp. 465–471, 2014.

[5] C. Hermes, J. Einhaus, M. Hahn, C. Wöhler, and F. Kummert, "Vehicle tracking and motion prediction in complex urban scenarios," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '10)*, pp. 26–33, June 2010.

[6] A. Ramirez, E. Ohn-Bar, and M. Trivedi, "Integrating motion and appearance for overtaking vehicle detection," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 96–101, IEEE, Ypsilanti, Mich, USA, June 2014.

[7] J. Arróspide, L. Salgado, M. Nieto, and F. Jaureguizar, "On-board robust vehicle detection and tracking using adaptive quality evaluation," in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP '08)*, pp. 2008–2011, San Diego, Calif, USA, October 2008.

[8] L. Andreone, P. C. Antonello, M. Bertozzi, A. Broggi, A. Fascioli, and D. Ranzato, "Vehicle detection and localization in infra-red images," in *Proceedings of the 5th IEEE International Conference on Intelligent Transportation Systems (ITSC '02)*, pp. 141–146, Singapore, September 2002.

[9] B. Besbes, A. Apatean, A. Rogozan, and A. Bensrhair, "Combining SURF-based local and global features for road obstacle recognition in far infrared images," in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC '10)*, pp. 1869–1874, IEEE, Madeira Island, Portugal, September 2010.

[10] P. Hurney, P. Waldron, F. Morgan, E. Jones, and M. Glavin, "Night-time pedestrian classification with histograms of oriented gradients-local binary patterns vectors," *IET Intelligent Transport Systems*, vol. 9, no. 1, pp. 75–85, 2015.

[11] Q. Liu, J. Zhuang, and S. Kong, "Detection of pedestrians at night time using learning-based method and head validation," in *Proceedings of the IEEE International Conference on Imaging Systems and Techniques (IST '12)*, pp. 398–402, Manchester, UK, July 2012.

[12] H. Wang, Y. Cai, X. Chen, and L. Chen, "Night-time vehicle sensing in far infrared image with deep learning," *Journal of Sensors*, vol. 2016, Article ID 3403451, 8 pages, 2016.

[13] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 694–711, 2006.

[14] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
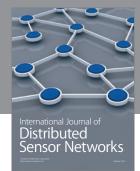
[15] S. Montabone and A. Soto, "Human detection using a mobile platform and novel features derived from a visual saliency mechanism," *Image and Vision Computing*, vol. 28, no. 3, pp. 391–402, 2010.

[16] S. J. Noh, D. Y. Shim, and M. G. Jeon, "Adaptive sliding-window strategy for vehicle detection in highway environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 323–335, 2016.

[17] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

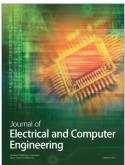[18] F. Wood and G. E. Hinton, "Training products of experts˜by minimizing contrastive divergence," Tech. Rep., Brown University, 2012.