LETTER

LGCN: Learnable Gabor Convolution Network for Human Gender Recognition in the Wild*

Peng CHEN[†,††], **Weijun LI**[†,††a)], ***Nonmembers***, **Linjun SUN**[†,††], ***Student Member***, **Xin NING**[†,††], **Lina YU**[†,††], ***and*** **Liping ZHANG**[†,††], ***Nonmembers***

SUMMARY Human gender recognition in the wild is a challenging task due to complex face variations, such as poses, lighting, occlusions, etc. In this letter, learnable Gabor convolutional network (LGCN), a new neural network computing framework for gender recognition was proposed. In LGCN, a learnable Gabor filter (LGF) is introduced and combined with the convolutional neural network (CNN). Specifically, the proposed framework is constructed by replacing some first layer convolutional kernels of a standard CNN with LGFs. Here, LGFs learn intrinsic parameters by using standard back propagation method, so that the values of those parameters are no longer fixed by experience as traditional methods, but can be modified by self-learning automatically. In addition, the performance of LGCN in gender recognition is further improved by applying a proposed feature combination strategy. The experimental results demonstrate that, compared to the standard CNNs with identical network architecture, our approach achieves better performance on three challenging public datasets without introducing any sacrifice in parameter size.

key words: *gender recognition, learnable Gabor convolutional neural network, learnable Gabor filter, back propagation*

1. Introduction

The existing gender recognition algorithms can be grouped into three categories: conventional hand-crafted feature-based methods, currently prevalent deep-learning-based methods and new integration-based methods. The hand-crafted feature-based methods generally use a human designed feature descriptor to extract the gender-information-related features from the image pixel space [1], [2]. Although these hand-crafted feature descriptors are sufficiently effective to extract meaningful information for gender recognition in controlled settings, their intrinsic parameters are difficult to set up. In addition, these methods have only passable performance in complex uncontrolled cases because of the limited modeling capacity. The deep-learning-based methods consider using the convolutional neural network (CNN) to extract gender information from large image sets by statistical training [3], [4]. These methods have powerful nonlinear modeling ability and can easily distinguish gender attributes in the training set when the training samples are insufficient. However, they have many parameters and can be easily overfitted when the network becomes increasingly deeper.

Unlike these two types of method, the integration-based methods attempt to combine the steerable hand-crafted features with the powerful CNN. Since gender information is highly related to facial texture features such as the angle and depth of the wrinkles and existence of beard, the bio-inspired Gabor filters are considered one of the most effective hand-crafted feature extractors. Recently, some studies [5], [6] in general feature extraction have successfully integrated Gabor filters with CNNs. [5] reduces the training complexity of CNNs by replacing certain weight kernels of a CNN with Gabor filters. The learnable convolution filters are modulated by Gabor filters in [6] to improve the robustness of CNN against image transformations. However, such excellent ideas have not been well explored in gender recognition. Though [7] fuses the human-designed Gabor filter features with original image pixels to enhance the performance of CNNs for gender recognition, it increases the depth of networks and the number of parameters. Besides, the intrinsic parameters of Gabor filters in all the methods above are fixed and not always optimal.

In this letter, a new LGF is designed for extracting specific local image patterns automatically. We then propose a framework that integrates the LGF with CNN for gender recognition in the wild. We call this framework learnable Gabor convolution network (LGCN). In our framework, partial weight kernels of a standard CNN in the first layer are replaced by LGFs. Moreover, the intrinsic parameters of LGFs can be learned automaticly using the back propagation method, which is difficult and time-consuming to manually set up. In addition, we propose a feature-combined strategy that further improves the performance of LGCN in gender recognition. The extensive experimental results show that our method consistently outperforms the state-of-the-art methods on three challenging benchmarks.

2. The Proposed Approach

2.1 Learnable Gabor Filter

A typical 2D Gabor filter is a Gaussian envelope function modulated by a sinusoidal carrier wave. It has a real and an imaginary component, which can be expressed as:

Manuscript received November 15, 2018.
Manuscript revised April 27, 2019.
Manuscript publicized June 13, 2019.

[†]The authors are with the Institute of Semiconductors, CAS, Beijing 100083, China.
[††]The authors are with the Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China.

*This work was supported by the National Nature Science Foundation of China (Grant No.61572458).

a) E-mail: wjli@semi.ac.cn (Corresponding author)

DOI: 10.1587/transinf.2018EDL8239

$$G_r(x, y; \lambda, \theta, \psi, \sigma, \gamma) = Ae^{-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}} cos(2\pi\frac{x'}{\lambda} + \psi) \quad (1)$$

$$G_i(x, y; \lambda, \theta, \psi, \sigma, \gamma) = Ae^{-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}} sin(2\pi\frac{x'}{\lambda} + \psi) \quad (2)$$

where

$$x' = xcos\theta + ysin\theta, \quad y' = -xsin\theta + ycos\theta \quad (3)$$

In these equations, A, λ, θ, ψ, γ and σ are the magnitude of the Gabor filter, wavelength of the Gabor filter function, orientation of the normal to the parallel stripes of a Gabor filter kernel, phase offset, spatial aspect ratio and standard deviation of the Gaussian envelope, respectively. x and y are the 2D world coordinates. G_r and G_i are the real and imaginary parts of the Gabor filter, respectively. By applying the chain rule, we can obtain the partial derivatives of G with respect to all parameters as follows:

$$\frac{\partial G_r}{\partial \lambda} = G_i\frac{2\pi x'}{\lambda^2}, \quad \frac{\partial G_r}{\partial \psi} = -G_i \quad (4)$$

$$\frac{\partial G_r}{\partial \theta} = \frac{G_r x' y'}{\sigma^2}(\gamma^2 - 1) - G_i\frac{2\pi y'}{\lambda} \quad (5)$$

$$\frac{\partial G_r}{\partial \sigma} = G_r\frac{x'^2 + \gamma^2 y'^2}{\sigma^3}, \quad \frac{\partial G_r}{\partial \gamma} = -G_r\frac{\gamma y'^2}{\sigma^2} \quad (6)$$

Let $K(i, j), i \in 0, 1, 2, \cdots, h-1; j \in 0, 1, 2, \cdots, w-1$ be a kernel function in the pixel space, where h and w are the height and width, respectively. In general, the height and width are restricted to positive odd numbers. By sampling in the world coordinates, we can generate the Gabor filter kernel as follows:

$$K(i, j) = G_r(s_x(i - \frac{h-1}{2}), s_y(j - \frac{w-1}{2})) \quad (7)$$

where s_x and s_y are the sampling ratios of the x and y dimension, respectively, in world coordinates. Giving input image X and generated Gabor filter kernel K, the feed-forward of the learnable Gabor filter can be written as:

$$O = X * K \quad (8)$$

In this equation, O is the convolution result of X and K. Using the standard back propagation algorithm, we can update each parameter of the Gabor filter as follows:

$$\lambda' := \lambda - \eta\sum_{i=0}^{h-1}\sum_{j=0}^{w-1}\frac{\partial O}{\partial K_{ij}}\frac{\partial K_{ij}}{\partial \lambda} \quad (9)$$

where η is the learning rate, and h and w are scalars. The process of the learnable Gabor filter is shown in Fig. 1. To prevent the parameters of Gabor filters getting away from the scope of specific physical significance, a simple clamp opreation is used to constrain the parameters in the feed-forward phase. In this study, we focus on the self-learning ability of parameter λ. The proposed method can provide reference for the other parameters adjustment. In order to evaluate the performance, other parameters other than λ in the experiment are determined empirically.

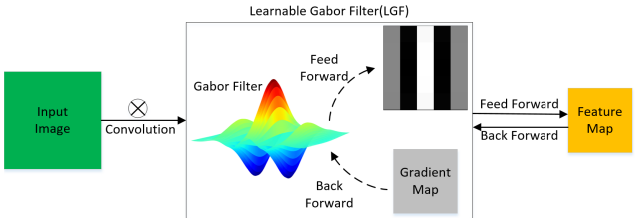

Fig. 1 Feed-forward and back-forward process of a learnable Gabor filter.

2.2 The Proposed Framework

Referenced from Levi's work [3], an Alexnet-liked network was selected as our basic network structure. The framework of the proposed method (LGCN) is shown in Fig. 2 (top). The first layer of the framework is a group of LGF modules, which are used to capture different frequency and orientation responses of the color input image. Then, the response maps are fed into the conventional CNN to extract robust and discriminant features of higher vision level for the subsequent step. At the end of the framework, the Softmax module is used to produce the final classification probability.

In detail, the proposed framework takes raw pixels of color face images as the input. The first layer of LGCN has 96 LGFs by combining the cases of twelve $\theta = 0, \frac{\pi}{12}, \frac{2\pi}{12}, \cdots, \frac{11\pi}{12}$ and eight $\psi = 0, \frac{\pi}{8}, \cdots, \frac{7\pi}{8}$. The parameters ($\sigma$ and γ) are identical for the 96 filters. We set $\sigma = 2$ and $\gamma = 0.3$ referenced from [7], whereas λ is learned from the training data. The kernel size of each Gabor filter is 5x5 with stride 1 and padding 2. Then, there are two convolution layer with 256 and 384 channels respectively. The kernel size of the second convolution layer is 5x5 with stride 1 and padding 2. The kernel size of the third convolution layer is 3x3 with stride 1 and padding 1. All the convolution layer is followed by a BactchNorm normalization layer and a ReLU non-linear unit. Behind the ReLU unit is a max-pooling layer sized 3x3 with stride 2. Finally, two fully connected layers are stacked after the pooling layer. The neurons of the fully connected layer are both 512. Dropout strategy is also adopted by us as it can limit the risk of overfitting. We set the dropout ratio as 0.5 for all networks.

2.3 Feature-Combined Strategy

As observed in [8], some of the trained filters from shallow layers are similar to Gabor filters while there are still a lot of other unknown types of patterns. Motivated by this, we propose a feature-combined strategy. We constrain part of the filters in the first layer of LGCN as LGFs. We use standard convolutional kernels to learn the remaining unknown patterns as it can fit any kind of functions. The number of LGFs ε is a hyperparameter. Here, we set ε as 24 by experience. The framework of the feature-combined LGCN is shown in Fig. 2 (bottom). We call this feature-combined

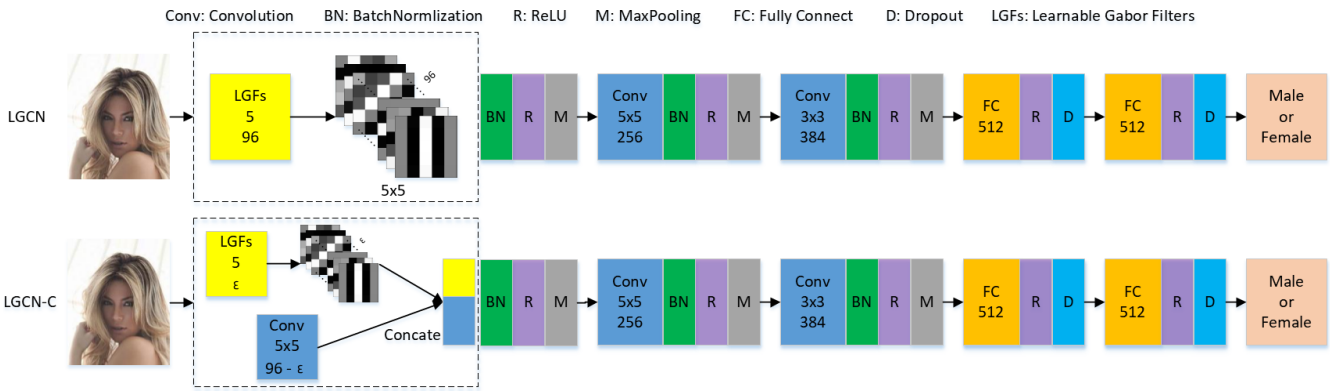

Fig. 2 Frameworks of LGCN and LGCN-C (ε is a hyperparameter which represents the number of LGFs).

framework LGCN-C. A little different from LGCN, LGCN-C will concatenate the feature maps extracted by LGFs and standard convolutional kernels along channel dimension. In addition, we reduce the number of ψ to two for convenience of calculations, i.e. $\psi = 0, \frac{\pi}{2}$, while keep other parameter setting the same as LGCN.

3. Experimental Results

The experiments were carried out in PyTorch framework on a Linux machine with Intel Xeon CPUs and Nvidia 1080Ti GPUs. We employ the SGD strategy to train our network. The initial learning rate of standard convolutional kernels and LGFs are 0.001 and 0.1, respectively, and decayed by 0.1 each 80 epochs. The total training epochs are 200. For a single image of size 227x227x3, the inference time of LGCN and LGCN-C with GPU are 9.7ms and 9.32ms, respectively. The model size are 145.094M and 145.098M, respectively.

3.1 Dataset Description

We conduct the experiments on three challenging datasets: Adience, CelebA and LFW. All these datasets can be considered a type of real-world reflection with extreme variations in head pose and lighting condition quality. We select the in-plane aligned version of Adience for our research, which were originally used in [1]. We report our results using subject-exclusive partitioning for five-fold cross validation referenced from [3]. We select the aligned and cropped version of CelebA for our research. We use the gender attribute in our experiment. The unaligned dataset of LFW is selected by us. Two protocols are performed on this dataset. The first protocol is randomly selecting the 80% images for training and the remaining images for testing, which was referenced from [2]. The other protocol is half of the images for training and half for testing, which was originally used in [9].

3.2 Effectiveness of LGF

To verify the effectiveness of LGF, we design a variant of LGCN with empirically fixed parameters of Gabor filters

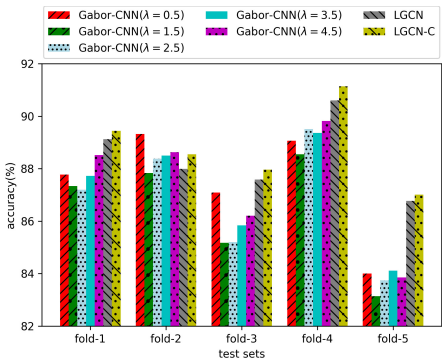

Fig. 3 Comparison of accuracy between Gabor-CNNs and LGCNs on the 5-fold experiments.

in the first layer for comparison. We call this network the Gabor convolutional neural network (Gabor-CNN). Referenced from the experimental value of $\lambda = 2.5$ in [7], we simply extend the range of λ to $0 \sim 5$. We conduct 5 groups of experiments on Adience data set by setting $\lambda = 0.5$, 1.5, 2.5, 3.5 and 4.5. Each group of experiment is measured by the five-fold cross validation. As shown in Fig. 3, Gabor-CNN with different λ values has significantly different performances, which further proves that the parameters of Gabor filters are hard to set up. For fair comparison, both the values of λ of LGCN and LGCN-C in the following experiments are initialized in the interval (0, 5). From Fig. 3, our LGCN methods outperform Gabor-CNN methods on most of the fold experiments. The reason is that LGCN can not only learn suitable λ values but also find optimal combination way, which are difficult to set up for traditional methods.

To explore other parameters of LGF, we set Gabor-CNN($\lambda = 2.5$, $\theta = 0, \frac{\pi}{12}, \frac{2\pi}{12}, \cdots, \frac{11\pi}{12}$, $\psi = 0, \frac{\pi}{8}, \cdots, \frac{7\pi}{8}$, $\sigma = 2$, $\gamma = 0.3$) as the baseline and independently learn θ(LGCN-θ), ψ(LGCN-ψ), γ(LGCN-γ) and σ(LGCN-σ), respectively. For example, if we hope to learn the parameter θ, we only update the value of θ while keep other parameters fixed as the same as the baseline. We conduct the experiments on the Adience dataset. The experimental results are reported as the following table.

Table 2 Comparison results with the state-of-the-art methods

Method	Adience*(%)	Method	CelebA(%)	Method	LFW(%)
LBP [1]	73.4 ± 0.7	LNet+ANet [9]	98.00	Kumar *et al.* [10]	85.80
FPLBP [1]	72.6 ± 0.9	MOON [11]	98.10	LNet+ANet [9]	94.00
LBP+FPLBP+Dropout 0.5[1]	76.1 ± 0.9	MCNN+AUX [12]	98.17	MCNN+AUX [12]	94.02
Best from Levi [3]	86.8 ± 1.4	DMTL [13]	98.00	Liu *et al.* [14]	95.80
CNN-ELM+Dropout 0.5[4]	87.3 ± 1.0	AFFACT [15]	98.26	Cao *et al.* [16]	96.20
CNN-ELM+Dropout 0.7[4]	88.2 ± 1.7	PaW [17]	**98.39**	PartAdaTrans [2] (80% Training)	96.80
GCN**[6]	88.1 ± 1.6	GCN**[6]	97.33	GCN**[6] (50%/80% Training)	96.82/97.80
Proposed LGCN	**88.4 ± 1.3**	Proposed LGCN	98.30	Proposed LGCN (50%/80% Training)	**97.22/97.84**
Proposed LGCN-C	**88.8 ± 1.4**	Proposed LGCN-C	**98.52**	Proposed LGCN-C (50%/80% Training)	**97.34/98.06**

* Mean ± standard error over five folds are reported as the resulting measure of performance.
** We reproduce the results of GCN method by replacing the first layer LGFs in LGCN with GoFs [6] while keep the network architecture same.

Table 1 Learning other parameters of the Gabor filter

Method	Accuracy(%)	Method	Accuracy(%)
Gabor-CNN(λ = 0.5)	87.45 ± 1.9	LGCN-λ[1]	88.44 ± 1.3
Gabor-CNN(λ = 1.5)	86.42 ± 1.9	LGCN-θ	88.41 ± 1.6
Gabor-CNN(λ = 2.5)	86.82 ± 2.0	LGCN-ψ	87.05 ± 2.1
Gabor-CNN(λ = 3.5)	87.10 ± 1.9	LGCN-γ	87.53 ± 2.0
Gabor-CNN(λ = 4.5)	87.41 ± 2.1	LGCN-σ	88.39 ± 1.4

[1] It is also referred as LGCN in this paper.

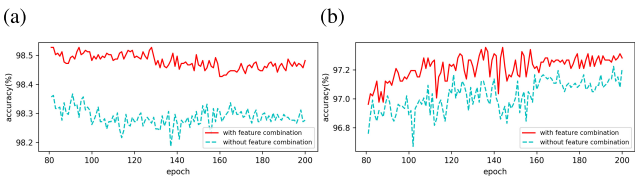

Fig. 4 Test accuracy curve: (a) CelebA and (b) LFW dataset.

As shown in Table 1, both ψ and γ have minor improvement on performance while the learning of other three parameters(λ, θ and σ) has improved the performance significantly. This is because that human face has abundant textural and directional information, which can be easily extracted by Gabor filters with suitable scale and orientation setting. However, ψ and γ are related to the phase offset and spatial aspect ratio of Gabor filter and intrinsically contributes less to this kind of pattern. Due to the best performance of LGCN-λ(LGCN), it was adopted in the following sections for comparison and feature-combined strategy exploring.

3.3 Evaluation of Proposed Feature-Combined Strategy

Figure 4 shows that the test accuracy with proposed feature-combined strategy consistenly outperforms the other method on CelebA and LFW datasets. We owe the improvement of performance to the ability of feature-combined method, which can learn more complex patterns.

3.4 Compared to the State of the Art

Table 2 reports the comparative results against the state-of-the-art methods on the Adience, CelebA and LFW datasets. It is very encouraging to see that our proposed method consistently outperforms the existing ones on the three datasets. This confirms the effectiveness of the proposed approach. Moreover, the proposed method does not introduce any sacrifice in parameter size. Compared to [3]†, the parameter size of our network is slightly reduced in two aspects: smaller kernel size and kernels with fewer parameters. Compared to [6]††, the parameter size of our network is slightly reduced due to the replacement of partial standard kernels to LGFs. As we know, the parameters of a single Gabor filter are always 5 regardless of the kernel size. Hence, the parameter size of each single standard convolutional kernel replaced by LGF is reduced to 20% in our framework.

4. Conclusions

In this letter, a new framework that integrates the proposed LGFs with CNNs is presented. The experimental results demonstrate that our method consistently outperforms the existing methods on three datasets while does not introduce any sacrifice in parameter size compared to standard CNNs with identical network architecture. The future work will focus on the joint learning of multi-parameters of Gabor filters.

†The model size of [3] is 145.100M.
††The model size of our reproduced GCN [6] is 145.121M.

References

[1] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," IEEE Trans. Inform. Forensic Secur., vol.9, no.12, pp.2170–2179, 2014.
[2] Y. Gao, Z. Li, and Y. Qiao, "Adaptive part-level model knowledge transfer for gender classification," IEEE Signal Process. Lett., vol.23, no.6, pp.888–892, 2016.
[3] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.34–42, 2015.
[4] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning CNN–ELM for age and gender classification," Neurocomputing, vol.275, pp.448–461, 2017.

[5] S.S. Sarwar, P. Panda, and K. Roy, "Gabor filter assisted energy efficient fast learning convolutional neural networks," 2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED), pp.1–6, 2017.
[6] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu, "Gabor convolutional networks," IEEE Trans. on Image Process., vol.27, no.9, pp.4357–4366, 2018.
[7] S. Hosseini, S.H. Lee, and N.I. Cho, "Feeding hand-crafted features for enhancing the performance of convolutional neural networks," arXiv, 2018.
[8] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," International Conference on Neural Information Processing Systems, pp.1097–1105, 2012.
[9] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," 2015 IEEE International Conference on Computer Vision (ICCV), pp.3730–3738, 2015.
[10] N. Kumar, A.C. Berg, P.N. Belhumeur, and S.K. Nayar, "Describable visual attributes for face verification and image search.," IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.10, pp.1962–1977, 2011.
[11] E.M. Rudd, M. Günther, and T.E. Boult, "Moon: A mixed objective optimization network for the recognition of facial attributes," European Conference on Computer Vision, vol.9909, pp.19–35, 2016.
[12] E. Hand and R. Chellappa, "Attributes for improved attributes: A multi-task network utilizing implicit and explicit relationships for facial attribute classification," 31st AAAI Conference on Artificial Intelligence, AAAI 2017, pp.4068–4074, 2017.
[13] H. Han, A.K. Jain, S. Shan, and X. Chen, "Heterogeneous face attribute estimation: A deep multi-task learning approach," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.PP, no.99, pp.1–1, 2017.
[14] H. Liu, Y. Gao, and C. Wang, "Gender identification in unconstrained scenarios using self-similarity of gradients features," 2014 IEEE International Conference on Image Processing (ICIP), pp.5911–5915, Oct. 2014.
[15] M. Günther, A. Rozsa, and T.E. Boult, "Affact: Alignment-free facial attribute classification technique," 2017 IEEE International Joint Conference on Biometrics (IJCB), pp.90–99, 2017.
[16] D. Cao, R. He, M. Zhang, Z. Sun, and T. Tan, "Real-world gender recognition using multi-order lbp and localized multi-boost learning," IEEE International Conference on Identity, Security and Behavior Analysis, pp.1–6, 2015.
[17] H. Ding, H. Zhou, S. Zhou, and R. Chellappa, "A deep cascade network for unaligned face attribute classification," 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, pp.6789–6796, 2018.