Neurocomputing 378 (2020) 387–398

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Deep Gabor convolution network for person re-identification

Yuan Yuan, Jian'an Zhang, Qi Wang*

School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China

ARTICLE INFO

Article history:
Received 12 March 2019
Revised 16 September 2019
Accepted 22 October 2019
Available online 31 October 2019

Communicated by Dr. Zhang Zhaoxiang

Keywords:
Person re-identification
Gabor convolution
Resnet-50
Gabor filter

ABSTRACT

Person re-identification is an import problem in computer vision fields and more and more deep neural network models have been developed for representation learning in this task due to their good performance. However, compared with hand-crafted feature representations, deep learned features cannot not be interpreted easily. To meet this demand, motivated by the Gabor filters' good interpretability and the deep neural network models' reliable learning ability, we propose a new convolution module for deep neural networks based on Gabor function (Gabor convolution). Compared with classical convolution module, every parameter in the proposed Gabor convolution kernel has a specific meaning while classical one has not. The Gabor convolution module has a good texture representation ability and is effective when it is embedded in the low layers of a network. Besides, in order to make the proposed Gabor module meaningful, a new loss function designed for this module is proposed as a regularizer of total loss function. By embedding the Gabor convolution module to the Resnet-50 network, we show that it has a good representation learning ability for person re-identification. And experiments on three widely used person re-identification datasets show favorable results compared with the state-of-the-arts.

1. Introduction

Person re-identification addresses the problem of matching persons across non-overlapping camera networks, which has attracted many researchers recent years. It can be regarded as a retrieval problem as well as a classification problem. Person re-identification is a challenge problem as there are various changes for person appearance, such as different illumination condition, different viewpoint and pose changes, etc.

As many computer vision tasks do, the first step for person re-identification is to extract feature representations for person images. Traditionally, a hand-crafted descriptor will be extracted such as color histogram of different color spaces (e.g. RGB, HSV, YCrCb, Lab), texture histogram (e.g. LBP, SILTP, Gabor filters) and combination of them (e.g. ELF [1], SDALF [2], LOMO [3], GOG [4]),enhanced LOMO [5]. Recently, with the success of deep neural networks (DNN) in computer vision fields, more and more works focus on representation learning and feature representation will be learned through DNN models, such as [6–26]. A hand-crafted feature representation is direct and can be interpreted easily, but it is less discriminative than deep features. While a deep feature representation often has a more discriminative ability, but it cannot be interpreted as easy as hand-crafted features. It is a consistent demand that one DNN module can be interpreted as hand-crafted features do.

In this paper, to meet the demand for explaining what DNN modules learn, motivated by the Gabor filters' good interpretability and the DNN models' reliable learning ability, we propose a new convolution module for deep neural networks based on Gabor function, which has a good interpretability and compatibility with deep neural network models. Gabor filters are generated from Gabor function and have been extensively used in computer vision tasks as they show impressive ability to model texture information for images. Traditionally, for the usage of Gabor filters, as shown in Fig. 1(a), we will first generate a group filters based on Gabor function with a group of predefined parameters, and then convolute them with an image to get a series of feature maps and at last histograms are computed on these feature maps.

As we can expect, in order to apply Gabor filters, we have to manually select proper values of parameters of Gabor function which is a cumbersome task. Besides, hand-designed parameters only cover a very small range of parameter space which will be suboptimal for certain inputs. Motivated by the learning ability of deep neural networks, it is natural to come out that we can learn the parameters of Gabor function through a deep neural network model instead of manually designing for solving above drawbacks of Gabor filters. In order to make Gabor filters be compatible with deep neural network models, we design Gabor filter as a special convolution module that all the convolution kernels are generated from Gabor function with learnable parameters. As shown in

* Corresponding author.
E-mail addresses: y.yuan1.ieee@gmail.com (Y. Yuan), zhangjianan09@gmail.com (J. Zhang), crabwq@gmail.com, crabwq_elsevier@163.com (Q. Wang).

https://doi.org/10.1016/j.neucom.2019.10.083
0925-2312/

Fig. 1(b), the generated Gabor filter is embedded into a neural network as a convolution layer and all the parameters are learned when the network is trained. In this pipeline, feature representations are acquired through the output of a certain layer. Next, we will refer the proposed convolution module as *Gabor convolution*.

Different from general convolution module, where all the elements of the convolution kernel are randomly generated and there are no relationships between them, the Gabor convolution kernel is generated from Gabor function and each element in the kernel is related to each other. As every parameter of Gabor convolution has a specific meaning and we can interpret what this module learns easily.

Note that every parameter of the Gabor convolution module has a specific range. In order to make the parameters of Gabor convolution legal when it is trained in a DNN model, we have to constrain the range of each parameter. To achieve this purpose, we propose a new regularizer loss function designed for the Gabor convolution module by taking advantage of the hinge function. Experiments show that with the proposed regularizer loss, the Gabor convolution module can learn legal parameters and improves the performance of person re-identification.

The main contributions of this paper can be summarized as

- A new convolution module is proposed for the DNN based on Gabor function and compared with traditional convolution module, and the proposed Gabor Convolution is more suit for low-level and show admi effect.
- A new regularizer loss function designed for the Gabor Convolution module is proposed by taking advantage of the hinge function.
- Performance of person re-identification is improved by embedding Gabor convolution module to ResNet-50 and extensive experiments validate the effectiveness of the proposed Gabor convolution module.

In the next section, we will review the related works. And then we will present the proposed Gabor convolution module in Section 3. Section 4 presents an extensive comparison with state-of-the-art algorithms, and we analyze each component of our method. Section 5 concludes the paper and discusses the future works.

2. Related works

In this section, we will review two types of works that are related to our work, (1) Gabor Filter related works in person re-identification, (2) deep neural networks based models for person re-identification.

2.1. Gabor filter in person re-identification

Two-dimensional Gabor functions were first proposed by Daugman [27] to model the spatial summation properties (of the receptive fields) of simple cells in the visual cortex. They are widely used in image processing, computer vision, neuroscience and psychophysics. In computer vision fields, Gabor filters can be interpreted as a kind of texture feature. Gabor filters have been widely used in person re-identification task. For feature based methods, ELF [1] takes 8 Gabor filters for feature extraction as a kind of texture information and combines color information for person re-identification. Ma et al. [28] uses the feature map convoluted with Gabor filters for pixel feature extraction and a covariance-based descriptor name BiCov is developed. Liu et al. [29] proves that Gabor feature is an important feature representation for person re-identification via a plenty of experiments. Ma et al. [30] extends [28] and improves the performance. Besides, Gabor features are used as part of feature representation for person re-identification in [31].

Note that, all the above works are based on hand-designed Gabor filters. Different from all these works, the proposed Gabor convolution can learn the parameters through a deep network model automatically.

2.2. Person re-identification based on DNN

Deep neural network models have received great success in many computer vision fields [32–35] and person re-identification is not an exception. More and more works focus on this kind of model for solving person re-identification.

Generally, deep neural networks can be regarded as feature extractors for person re-identification and the whole training process is often treated as embedding learning or representation learning. As many embedding learning works do, there are lots of works for person re-identification focus on designing of more effective loss functions such as [6–11]. Triplet loss was first proposed by Schroff et al. [36] used for face verification and has been widely used in person re-identification. Cheng et al. [6] adds a new item to triplet loss to constrain the distances of positive pairs be less than a predefined threshold. Hermans et al. [7] improves the triplet loss by integrating hardest positive and hardest negative into the loss function. Zhou et al. [8] develops a P2S loss function consisted of a pairwise term, a triplet term and a regularizer term, which combines the margin function and symmetric triplet function of the distance of point to set. Chen et al. [9] proposed a new quadruple-based loss function by taking the distances of two negative samples into consideration. Yu et al. [11] develops a HAP2S loss function that adopts an adaptive hard ming scheme integrating adaptive weights. Apart from triplet-based loss function, plenty of works use contrastive loss function to train models such as [15,37]. Xiao et al. [10] also develops a new OIM loss function which is based on contrastive loss function and specific batches of images.

Besides, taking the structure of pedestrian images into consideration, many researchers devote to designing special networks applied to person re-identification such as [12,13,15–19]. Ahmed et al. [12] proposed a new layer that computes cross-input neighborhood difference that captures the local relationship between the two input images and a layer of patch summary features. Wu et al. [13] extends [12] and achieved higher performance. Varior et al. [15] designs a matching gate architecture composed of feature summarization, similarity computation and filtering modules, which is used to compare the feature maps of two input images. Lin et al. [16] takes use of the consistent information of the whole network and proposes a similarity measure module with constraints and optimized the module through the deep neural network. Li et al. [17] proposes MACAN model that integrates spatial transformer net into the network to localize the body parts and combine feature from different parts for person re-identification. Zhao et al. [18] also focuses on body part localization and design a part-aligned representation extractor with a sub-deep neural network by input feature maps of images. Wu et al. [19] integrates SIFT feature using fishier encoding into a deep neural network for representation learning for person re-identification. Some gait-based and clustering methods have also proposed in this community such as [38–41].

Furthermore, many attention-based models have developed to learn more discriminative feature for person re-identification such as [20–22]. Liu et al. [22] proposes a comparative attention network (CAN) that combines a LSTM structure with a sub-convolutional neural network to examine the importance of multiple highly discriminative regions. Liu et al. [20] proposes a new attention-based network named HydraPlus-Net (HPnet) that

combines the outputs of different layers of inception net and such a structure can enrich the final feature representation. Si et al. [21] designs a dual attention block which is composed of one transformation layer and one attention layer, and based on such a module a new distance function is derived for aggregating features.

Recently, plenty of works try to improve the performance of person re-identification with auxiliary information such as pose skeleton and semantic segmentation of human. Representative works include [23–26]. With the human pose skeleton information, Zhao et al. [23] proposes a Region Proposal Network(RPN) to extract the body regions and embed the RPN to person re-identification networks. Zheng et al. [24] introduces a PoseBox structure which is generated though pose estimation followed by affine transformations to align pedestrian to a standard pose. Su et al. [25] also uses pose skeleton to extract body parts and design a feature embedding sub-net to alleviate the pose variations and learn robust feature representations. Kalayeh et al. [26] integrates human semantic parsing instead of body parts into person re-identification and achieves state-of-the-art performance.

Different from above works, the proposed method designs a new convolution module based on Gabor function, which is a more basic structure than some specific structures proposed for person re-identification. Although it is a basic structure, is shows an effective representation learning ability.

3. Methodology

3.1. Revisit to Gabor filter and parameters standardization

A two-dimensional Gabor function is defined as a sinusoidal wave multiplied by a Gaussian function in a complex number form

$$g_{\Theta}(x,y) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda}+\phi\right)\right), \tag{1}$$

where

$$\begin{aligned} x' &= x\cos(\theta)+y\sin(\theta)\\ y' &= -x\sin(\theta)+y\cos(\theta), \end{aligned} \tag{2}$$

i is imaginary unit and $\Theta = \{\lambda, \theta, \phi, \sigma, \gamma\}$ is the parameter. Note that Eq. (1) can be expressed in a real number form, where the real part is

$$g^{real}_{\Theta}(x,y) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\phi\right), \tag{3}$$

and the imaginary part is

$$g^{img}_{\Theta}(x,y) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda}+\phi\right). \tag{4}$$

Both of the two parts can be used for convolution and we will see in the experimental part they have different convolution effects.

There are five parameters in Gabor function (1) and each of them has a specific meaning. And every parameter can take values in a specific range. We will describe the meanings and specify the range of each parameter in a standard way as follows.

- *Orientation* θ
  Parameter θ specifies the orientation of the Gabor filter generated by the Gabor function. Valid values are real number between 0 and 2π. Fig. 2 (a) shows the Gabor filters with the values of θ be 0, $\frac{\pi}{8}$, $\frac{\pi}{4}$, $\frac{3\pi}{8}$, $\frac{\pi}{2}$, $\frac{5\pi}{8}$, $\frac{3\pi}{4}$, $\frac{7\pi}{8}$. It can be found the bands of Gabor filters rotate with the change of orientation parameter.
- *Wavelength* λ
  Parameter λ represents the wavelength of the Gabor filter and its values is specified in pixels. Generally, valid values of λ are real number equal or greater than 2. Take into consideration that when $\lambda = 2$ and $\phi = \frac{\pi}{2}$, the Gabor function will be zero, so we set $\lambda > 2$. Besides, in practice, in order to prevent the occurence of undesired effects at the image borders, the wavelength value should be smaller than one fifth of the input image size. We set $C_\lambda = \min(I_{width}, I_{height})/5$, where I_{width} and I_{height} corresponds input image width and height and C_λ is the upper bound of λ. Hence we get the range of λ (2, C_λ]. Fig. 2(b) shows the Gabor filters with the values of λ changed from 2 to 16 with step 2. With the increase of λ, the bandwidth increases.
- *Phase offset* ϕ
  Parameter ϕ is the phase offset in the argument of the sine or cosine factor in the Gabor function.Its valid values are real numbers from $-\pi$ to π. The values 0 and π correspond to center-symmetric 'center-on' and 'center-off' functions, respectively, while $-\pi/2$ and $\pi/2$ correspond to anti-symmetric functions. All other cases correspond to asymmetric functions. Fig. 2(c) shows the Gabor filters with the values of ϕ ranged from 0 to $\frac{7\pi}{8}$. This parameter shows the phase offset.
- *Aspect ration* γ
  Parameter γ specifies the ellipticity of the support of the Gabor function. When $\gamma = 1$, the support is circular and when $\gamma < 1$ the support is elongated in orientation of the parallel stripes of the function. Generally, it takes real values that are greater than 0 and less or equals than 1, so its range is (0,1]. Fig. 2(d) shows the Gabor filters with the values of γ be 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.8, 1.0. We can find that with the increase of aspect ratio γ, the length of band decreases.
- *Generate* σ *from bandwidth* b
  Parameter σ is the standard deviation of the Gaussian factor of the Gabor function. Since λ and σ are not independent, Kruizinga adn Petkov [42] argues that σ cannot be specified directly and can only be generated through the bandwidth b, where it satisfies $\frac{\sigma}{\lambda} = \frac{1}{\pi}\sqrt{\frac{\log 2}{2}}\frac{2^b+1}{2^b-1} \triangleq \hat{b}$. Following [42], instead of specifying the σ directly, we generate it from the bandwidth b and treat $\hat{b}$ as a parameter of Gabor filter. It is required that b is greater than 0 and hence we can infer $\hat{b} > \frac{1}{\pi}\sqrt{\frac{\log(2)}{2}} \triangleq C_{\hat{b}}$. So the range of $\hat{b}$ is $(C_{\hat{b}}, \infty]$. Fig. 2(e) shows the Gabor filters with the values of σ be 0.1,0.2,0.4,0.5,0.7,0.8,1,2. It is obvious the value of σ influences the numbers of bands.

The parametrization used in Eq. (1) follows references [42–47], where further details can be found.

3.2. Gabor convolution

Traditionally, we have to design a chunk of Gabor filters with different parameter values in order to apply them. However, this kind of pipeline adopts only a few specific parameters which will be suboptimal for certain inputs. Besides, manually selecting proper values of these parameters of Gabor function is a cumbersome task. For these reasons, we pursue a machine learning approach that can learn these parameters in a discriminative fashion from training data. Motivated by the deep convolution network's learning ability, we propose the Gabor convolution filter module that can adaptively learn parameters through a deep neural network instead of handcrafted designed.

Normally, from the Gabor function shown in (3) and (4), we can create a set of Gabor convolution kernels with different parameters. Generally for a kernel of size $2\times k+1$, we set $x = [-k, -k+1, \ldots, k-1, k]$ and $y = [-k, -k+1, \ldots, k-1, k]$, and then the

Gabor convolution can be expressed in the following form

$$G_{\mathbf{\Theta}} = \begin{bmatrix} g_{\mathbf{\Theta}}^{(-k,-k)} & g_{\mathbf{\Theta}}^{(-k,-k+1)} & \cdots & g_{\mathbf{\Theta}}^{(-k,k)} \\ g_{\mathbf{\Theta}}^{(-k+1,-k)} & g_{\mathbf{\Theta}}^{(-k+1,-k+1)} & \cdots & g_{\mathbf{\Theta}}^{(-k+1,k)} \\ \vdots & \vdots & \ddots & \vdots \\ g_{\mathbf{\Theta}}^{(k,-k)} & g_{\mathbf{\Theta}}^{(k,-k+1)} & \cdots & g_{\mathbf{\Theta}}^{(k,k)} \end{bmatrix}, \tag{5}$$

where $g_{\Theta}^{x,y} = g_{\mathbf{\Theta}}(x, y)$.

Different from classical convolution, all elements in the proposed Gabor convolution kernels are generated from a meaningful Gabor function. While elements in a classical convolution kernels are randomly generated and there are no relationships between them. This is the crucial difference. Additionally, it is notable that the proposed Gabor convolution only has 5 parameters and a classical convolution with kernel size k will have k^2 parameters. This can reduce the size of a network to some extent especially when the k is large. In addition, all the parameters learned from Gabor convolution have its own meanings while a classical convolution cannot be explained one by one.

As a module, the Gabor convolution can be embedded into any network as classical convolution does. Training the Gabor convolutional layers with the 5 parameters can be done using a stochastic gradient descent method as usual.

3.3. Regularization loss of Gabor parameters

It is noted that all the parameters in Gabor function have a range of values. Compared with general convolution kernel which parameters are not limited in a range, we have to impose some constraints in order to make the Gabor kernel meaningful. One of the common ways is to enforce a regularizer for each parameter in the final loss function because it encodes our prior knowledge by penalizing solutions that do not satisfy the desired values. In such a way, an invalid learned parameter will be avoided. For this problem, a good regularizer should limit the parameters in a range while penalize them when they are out of range. A direct intuition is that we can use the indicator function. However, the indicator function is not differential. Instead, we propose to use a margin function as regularizer for each parameters as follows,

$$\mathcal{L}(x) = \max\{0, \mu_1 - x\} + \max\{0, x - \mu_2\}, \tag{6}$$

where μ_1 and μ_2 are lower and upper bound of x. Eq. (6) implies that when $x \in [\mu_1, \mu 2]$ the regularizer will be zero while when x is out of the range it will be penalized.

According to Eq. (6), for the parameter θ, we have

$$\mathcal{L}(\theta) = \max\{0, -\theta\} + \max\{0, \theta - 2\pi\}, \tag{7}$$

as it ranges from 0 to 2π. Similarly, we can get the regularizer of θ as

$$\mathcal{L}(\phi) = \max\{0, -\pi - \phi\} + \max\{0, \phi - \pi\}, \tag{8}$$

which restricts the parameter in $[-\pi, \pi]$.

For the parameter λ, which is different from the parameter θ where the range is a close set, it takes value from $(2, C_\lambda]$, so a margin m_λ is add to the margin loss in order to avoid close set boundary value. We have

$$\mathcal{L}(\lambda) = \max\{0, 2 + m_\lambda - \lambda\} + \max\{0, \lambda - C_\lambda\}. \tag{9}$$

In implementation, we can set m_λ be a small positive value, e.g. $1e-3$. For the parameter γ, similar to parameter λ, we set a margin m_γ and we have

$$\mathcal{L}(\gamma) = \max\{0, m_\gamma - \gamma\} + \max\{0, \gamma - 1\}, \tag{10}$$

which restricts the parameter in an open set.

At last, instead of using the σ for optimized, we turn to $\hat{b}$ and regularize it as

$$\mathcal{L}(\hat{b}) = \max\{0, C_{\hat{b}} - \hat{b}\}. \tag{11}$$

This is because only lower bound is constrained.

In total, for the Gabor filter parameters, we can impose the regularizer as

$$\mathcal{L}_{\hat{\mathbf{\Theta}}} = \mathcal{L}(\theta) + \mathcal{L}(\lambda) + \mathcal{L}(\phi) + \mathcal{L}(\gamma) + \mathcal{L}(\hat{b}), \tag{12}$$

where $\hat{\mathbf{\Theta}} = \{\lambda, \theta, \phi, \hat{b}, \gamma\}$ and $\mathcal{L}_{\hat{\mathbf{\Theta}}}$ stands for the total regularizer for the Gabor convolution. It is noted that there is no trade-off parameter between every parameter's regularizer because we argue that all the parameters share the same importance.

As most of the related works do, a triplet-like loss function will be used to train the whole network. Similar to Hermans et al. [7], we utilize the hard-mining triplet loss function to train our deep neural network model. And by introducing Eq. (12) as a regularizer, the final loss function is shown as follows

$$\mathcal{L}_{total} = \mathcal{L}_{tri} + \mu \mathcal{L}_{\hat{\mathbf{\Theta}}}, \tag{13}$$

where μ is a trade-off parameter that represents how importance of the regularizer contributes in the total loss function, and $\mathcal{L}_{tri}$ is the hard-mining triplet loss function with the formulation as

$$\mathcal{L}_{tri} = \frac{1}{N_s} \sum_{a=1}^{N_s} [\max_{y_a = y_b} d(\mathbf{f}_a, \mathbf{f}_\mathbf{b}) - \min_{y_n \neq y_a} d(\mathbf{f}_a, \mathbf{f}_n) + \tau]_+, \tag{14}$$

where $[x]_+$ is the abbreviation of function $max(0, x)$ and ($\mathbf{f}_a$, $\mathbf{f}_\mathbf{b}$, $\mathbf{f}_n$) is the input triplet tuple.

3.4. The gradient of the Gabor convolution module

In this sub-section we will give a brief derivation for the gradient of the Gabor convolution. Different from classical convolution module, the Gabor convolution module need to calculate gradients for only the parameter $\mathbf{\Theta}$. Let $\mathcal{L}_{total}$ be the total loss function, and $\mathbf{X}_i^l, l = 1, 2, \ldots, N$ be i-th feature maps in the l layer. So, according to the chain rule, the gradient of Gabor convolution (given it in the first layer) can be formulated as

$$\begin{aligned} \frac{\partial \mathcal{L}_{total}}{\partial \mathbf{\Theta}} &= \sum_{i=1}^{z_l} \frac{\partial \mathcal{L}_{total}}{\partial \mathbf{X}_i^1} \frac{\partial \mathbf{X}_i^1}{\partial g_{\mathbf{\Theta}}} \frac{\partial g_{\mathbf{\Theta}}}{\partial \mathbf{\Theta}} \\ \frac{\partial \mathcal{L}_{total}}{\partial \mathbf{X}_i^1} &= \sum_{k=1}^{z_{l+1}} \frac{\partial \mathcal{L}_{total}}{\partial \mathbf{X}_k^{1+1}} \frac{\partial \mathbf{X}_k^{1+1}}{\partial \mathbf{X}_i^1} \\ where\, l &= 1, 2, \cdots, N \end{aligned} \tag{15}$$

where z_l represents the number of feature maps in the l-th layer, and $g_{\mathbf{\Theta}}$ is the Gabor kernel shown in Eq. (5). $\sum_k^{z_{l+1}} \frac{\partial \mathbf{X}_i^{1+1}}{\partial \mathbf{X}_i^1}, l = 1, 2, \ldots, N$ represents the gradient of l-th layer with regard to $l+1$-th layer and $\frac{\partial \mathbf{X}_i^{1+1}}{\partial \mathbf{X}_i^1}$ implies the gradient of $g_{\mathbf{\Theta}}$. So according to Eq. (15), we can get the gradient with regard to feature maps and $g_{\mathbf{\Theta}}$ from back-propagation process layer by layer. In order to get the gradient of the Gabor convolution module $\frac{\partial \mathcal{L}_{total}}{\partial \mathbf{\Theta}}$, we need only to show how to compute $\frac{\partial g_{\mathbf{\Theta}}}{\partial \mathbf{\Theta}}$. The gradient of $\mathbf{\Theta}$ can be formulated as

$$\frac{\partial g_{\mathbf{\Theta}}}{\partial \mathbf{\Theta}} = \left[\frac{\partial g_{\mathbf{\Theta}}}{\partial \lambda}, \frac{\partial g_{\mathbf{\Theta}}}{\partial \theta}, \frac{\partial g_{\mathbf{\Theta}}}{\partial \phi}, \frac{\partial g_{\mathbf{\Theta}}}{\partial \hat{b}}, \frac{\partial g_{\mathbf{\Theta}}}{\partial \gamma} \right]^T, \tag{16}$$

where each element represents the gradient with regards to corresponding parameter. Next, we need to calculate every parameter's gradient which is easy to derive. Take the λ and g_{real} in Eq. (3) for example, we have

$$\frac{\partial g_{\mathbf{\Theta}}}{\partial \lambda} = \sum_{x,y} \frac{\partial g_{\mathbf{\Theta}}(x, y)}{\partial \lambda}.$$

and

$$\frac{\partial g_{\mathbf{\Theta}}(x,y)}{\partial \lambda} = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda}+\phi\right)\frac{2\pi x'}{\lambda^2}.$$

Similarly all other 4 parameters can be acquired in the same way.

The above formulas show how to calculate the gradient manually. In real experiments, we will adopt high-level deep learning packages such as PyTorch to calculate the gradients automatically.

4. Experiments

4.1. Evaluation and datasets

We conduct experiments on three widely used person re-identification datasets, namely, Market1501 [48], DukeMTMC-REID [49] and CUHK03 [50]. Besides, in order to validate the effectiveness of the proposed Gabor convolution module, we also perform a series of experiments on two image classification datasets MNIST and Cifar10 as this task is similar to person re-identification as far as feature representation learning. For the person re-identification task, each dataset is divided into training set and testing set. After training process executed on training set, an evaluation process is performed on testing set. Specifically, for each image in testing set, a deep feature representation is extracted based on the output of higher layer of trained model (e.g. FC layer), and next we split the testing set into gallery set and probe set, finally a matching process is executed between the probe set and gallery set. All the experiments are performed in a single-query setting. The performance is evaluated by the commonly used CMC top-k accuracy and rank-1, rank-5, rank-10 and rank-20 are reported as usual. Meanwhile, as [48] suggested, we also report the mAP on all three datasets. For the image classification task, the feature extraction process is the same as person re-identification does and we reported the accuracy evaluated on testing set after the training process is finished.

Next, we will give a detailed introduction to the datasets mentioned above. For the person re-identification task, three datasets are described as follows.

Market1501: The Market-1501 dataset [48] is proposed by Zheng et al. in 2015. It is a large person re-identificatio dataset that contains 1501 identities and 32,668 pedestrian images in total. All the pedestrian images are automatically detected by DPM detector from six videos captured by six cameras with different resolutions. This dataset provide a split of training set and testing set, where 751 identities with 12,936 images are used for training and the remaining 751 identities with 19,732 images (2793 distractor images are included) are used for testing. As the pedestrian images are acquired by DPM detector, this dataset has incorrect detections of pedestrian and occlusion also exists, which make it quite challenging.

DukeMTMC-REID: The DukeMTMC-REID dataset [49] is proposed by Ristani et al. in 2016. This dataset contains 1404 identities and 36,411 images in total. All the pedestrian images are hand-drawn from eight 85-min high-resolution videos captured by eight different cameras. Similar as Market1501, a official split of training set and testing set is provided, where 702 identities with 16,522 images consist of training set and the remaining 702 identities with 19,889 images (2228 are used for probe and 17,661 are used for gallery)compose testing set. This dataset undergoes high inter-class similarity and certain occlusion, which make it very challenging.

CUHK03: The CUHK03 dataset [50] is proposed by Li et al. in 2014. This dataset consists of 1467 identities with total 14,097 pedestrian images. It has five pairs of camera views and each identity only appears in tow disjoint camera views on CUHK campus. There are 4.8 images on average in each view. This dataset has two kinds of versions: labeled and detected, where images in the labeled version are human annotated from original videos while images in the detected version are automatically generated from DPM detector. It provides a split of training and testing set, where 1267 identities are used for training and 100 identities are used for validation and the last 100 identities are used for testing.

For the task of classification, MNIST and CIFAR-10 are used to evaluate the effectiveness of the proposed Gabor convolution module. MNIST dataset is a handwritten digits dataset and has widely used to test machine learning algorithms. It has a training set of 60,000 examples and a test set of 10,000 examples. CIFAR-10 dataset is a widely used basic dataset used in computer vision task. It consists of 50k training images and 10k testing images in 10 classes.

4.2. Implement detail

Network architechture: As the SPReid [26] method claims, by employing a simple yet effective training strategy, standard popular deep convolutional architectures with no modification can outperform current state-of-the-art. Without loss of generality, we use ResNet-50 [51] as backbone for representing learning. As lots of peer works do, for the person re-identification task, we adopt ResNet-50 as a backbone network, and replace the first convolution layer with the combination of the proposed Gabor convolution layer and a classical convolution layer. From experiments, we also yield the same conclusion.

Settings: All the network models used in this work are implemented by PyTorch and run on a computer configured with NIDIA Tesla K80 GPU cards. For all the experiment, training data is randomly divided into mini-batches with batch size of 32. Forward propagation is performed on each mini-batch and the loss is computed. After then, back propagation is executed to compute the gradients and weights are updated with Adam optimizer. We set the initial learning rate to $3e^{-4}$ for Market-1501 dataset and $5e^{-5}$ for DukeMTMC-REID dataset. We use a momentum of 0.9 and weight decay 5×10^{-3}. For the person re-identification, followed by authors in [11,26], we utilize the pretrained model on ImageNet to initialize the weights. The epoch for the three person re-identification datasets is set to 60 and 5 for the two classification dataset.

4.3. Effectiveness of Gabor convolution module

We first will validate the effectiveness of the proposed Gabor convolution module in image classification task on Cifar10 and MNIST dataset. The reasons for choosing image classification datasets for evaluation lie in three aspects. First, image classification is the most basic task in computer vision field and lots of important networks are first evaluated on image classification task such as Alexnet, VGG net and Resnet. Second, image classification and person re-identification are similar in terms of feature representation learning and the module will be still effective for person re-identification if it works for image classification. Third, it is easy and time-saving to train on the two image classification datasets and from these evaluations we can efficiently get insight into how the person re-identification networks can take advantage of the Gabor convolution module.

The purpose of this experimental setup is to prove that Gabor convolution is as effective as traditional convolutions and show its superiority. So we will only conduct experiments focusing on the structure itself and do not compare with other state-of-the-arts results in these datasets. For this purpose, we design a simple convolution net with only two layers as shown in Fig. 3.

Effective of the Gabor convolution: In order to evaluate the effectiveness of Gabor convolution, we combine different convolution modules in the designed two-layer network. Specifically, we can generate three convolution modules Greal, Gimg and Conv, where

Table 1
Different structure of DNN, 'conv' stands for classical convolution kernel, 'greal' represents the Gabor kernel generated by real part of Gabor function in Eq. (3) and 'gimg' implies the kernel generated by imaginary part of Gabor function Eq. (4).

Structure/dataset	MNIST	CIFAR10
Conv+Conv	98.74	58.67
Conv+Greal	97.97	54.33
Conv+Gimg	98.15	56.83
Greal+Conv	98.79	58.40
Greal+Greal	80.02	55.41
Greal+Gimg	80.11	54.89
Gimg+Conv	**98.82**	**61.26**
Gimg+Greal	95.12	56.12
Gimg+Gimg	96.60	53.27

Greal represents the Gabor kernel generated by real part of Gabor function in Eq. (3) and Gimg implies the kernel generated by imaginary part of Gabor function Eq. (4), while the Conv is the classical convolution module. And hence we can get eight different network models as shown in Table 1. The Greal + Conv means that the first layer is Greal and the second layer is Conv for the two convolution network, and all other combinations has the same meaning.

From Table 1 we can draw the following conclusions. First, we can find the Gabor convolution is still effective in Deep Convolution Networks with self-learned parameters. As we can find from 1, the combination of 'Greal' and 'Gimg' works worst for MNIST dataset, but still get 80% accuracy. Second, the Gabor convolution generated by imaginary part of Gabor function is more effective than the real part. As we can see, the combinations with 'Greal' module work worse than the combinations with 'Gimg' module. Traditionally, an imaginary Gabor convolution tends to find image edges while a real Gabor convolution tends to smooth an image. We argue that this property is also useful in deep neural network structure. Third, a network with all Gabor convolution modules performs pool than the network combined classical convolution module and Gabor convolution module. For instance, the combination of 'Greal'+'Gimg' performs poorer than 'Greal+Conv' and 'Conv+Gimg'. This is easy to explain because a Gabor convolution usually corresponds to texture features which are effective in low layers of a network. Fourth, the combination of 'Gimg+Conv' performs better than 'Conv+Conv' and is most effective structure in all these combinations on both datasets. This implies that we can use 'Gimg+Conv' to replace a classical convolution module in low layers of a network to achieve high performance.

Effective of the Gabor function: In order to verify the effectiveness of Gabor function we design a series experiments that use different part of Gabor function to generate convolution kernels. Eqs. (3) and (4) imply that the Gabor convolution is decided by three functions $\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)$, $\sin\left(2\pi\frac{x'}{\lambda}+\phi\right)$, and $\cos\left(2\pi\frac{x'}{\lambda}+\phi\right)$. So instead of using Gabor function to generate convolution kernels, we generate convolution kernels using these three functions and we name them 'ellipse-part', 'sin-part' and 'cos-part', respectively. Besides, we also test kernels generate by tangent function $\tan\left(2\pi\frac{x'}{\lambda}+\phi\right)$. Note that all above kernels are generated by specific function, where elements in convolution kernel are related with each other. In order to test how a regular function influence the performance, as a baseline, we also implement a random generated convolution with only 5 elements randomly located in a kernel while all the remainders are zeros, and we name this module as 'random'. We use 'Gimg+Conv' as baseline. Table 2 shows the result on MNIST and Cifar-10 dataset with the net structure as shown in Fig. 3.

Table 2
Different part of Gabor function.

Structure/dataset	MNIST	CIFAR-10
ellipse-part	89.01	32.97
sin-part	98.58	59.71
cos-part	98.57	60.08
tan	11.02	10.00
random	76.79	27.10
baseline	**98.82**	**60.54**

Table 3
Gonv2 structure with different loss functions. "CE" stands for Cross Entropy loss and "HM" represents Hard-Mine triplet loss. "GLoss" is the regularizer.

Settings	rank-1	rank5	rank10	rank20	mAP
ResNet-50 + CE	83.2	93.3	95.6	97.0	66.2
ResNet-50 + HM	85.7	94.3	96.2	97.6	69.5
struct. 1 + CE	79.8	91.4	94.5	96.2	60.4
struct. 1 + HM	81.2	92.6	95.1	96.8	62.8
struct. 2 + CE	84.0	93.6	95.6	97.2	66.7
struct. 2 + HM	87.0	94.5	96.2	97.8	71.7
struct. 1 + HM +gloss	83.8	93.6	96.0	97.9	66.8
struct. 2 + HM +gloss	**88.1**	**95.1**	**96.8**	**98.0**	**72.3**

It can be seen from Table 2 'ellipse-part' performs poorer than 'sin-part' and 'cos-part' convolution kernel. This demonstrates that in Gabor convolution the sin/cos part contributes more information than the ellipse part. Besides, we can find 'tan' generated convolution kernel perform poorest as we find that it cannot converge when training. This implies that not all functions can generate good convolution kernels. In addition, as we expected, a random convolution kernel performs poor because the elements in a kernel are randomly located and there are no relationship between elements. And we can find that all these generated convolution kernels have lower performance than the baseline. These experiments indicate that Gabor function is effective to generate a convolution kernel and not all other functions are as effective as the Gabor function.

4.4. How person re-id network benefit from Gabor filter

In this section, we will show that how person re-identification networks can benefit from Gabor filter. We will test two invariants of ResNet-50 [51] network by embedding the Gabor convolution module. In fact, other models can be also used as backbones. ResNet-50 model is the most used backbone network for person re-identification so we also test the performance of our Gabor convolution based on this network. ResNet-50 consists of one convolution module followed by four bottleneck modules.

Fig. 4 shows the two variant structures. First, we will use Gabor filter to replace a low-level convolution filter in a network and we name this structure as "struct.1". Second, a residual block combined with Gabor filter and classical convolution filter is designed as shown in Fig. 3 and we name this structure as "struct.2". Note that the Gabor convolution shows effectiveness in low-level of a network as proved in Section 4.3, so we replace only first few layers with the proposed structures in classical networks.

Table 3 shows the results of different settings evaluated on Market-1501 dataset. The networks include ResNet-50 and its two invariants 'struct.1' and 'struct.2'. We evaluate two kinds of loss functions, namely cross entropy (CE) and hard-mining triplet loss (HM). Besides, we also evaluate the proposed regularizer and we name it 'gloss' abbreviated for Gabor loss. So with the combination of the networks and loss functions, we get eight settings as shown in Table 3.

Gabor convolution embedded network vs Baseline: Table 3 gives the results of different structures of the network combined with

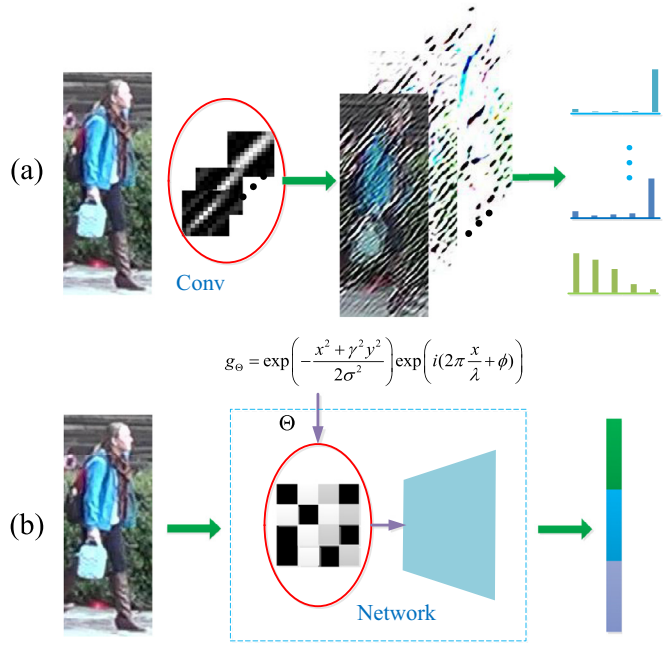

Fig. 1. The usage of Gabor filters: (a) Traditional way, Gabor filters are generated through a group of predefined parameters. (b) The proposed pipeline, Gabor filters are learned through a DNN model.

cross entropy loss and hard-mining triplet loss. From the first six rows, we can find that, 'struct.1' degrades the performance of person re-identification compared with the baseline ResNet-50 network. For instance, it reduces 3.4 rank-1 performance with cross-entropy loss and 4.5 rank-1 performance with hard-mining triplet loss. This demonstrates that it is not effective with only one Gabor convolution module instead of one classical convolution as shown in Fig. 4(a). As we can expect, the Gabor convolution module has a good texture representation ability as traditional Gabor filters. So it may focus more on texture information than other information such as color. This property may lead to the reduction of the performance of 'struct.1'. However, the feature maps after Gabor convolution also include rich information such as color which can be viewed in Fig. 6. So we designed a 'struct.2' for information capturing. Compared with baseline ResNet-50 network, 'struct.2' is more effective. It improves 0.8 rank-1 performance with cross entropy loss and 1.3 rank-1 performance with hard-mining triplet loss. This indicates that the combination of Gabor convolution and classical convolution is more effective than only one classical convolution in the low layers of a network. We argue that this is because the Gabor convolution module and classical convolution module are complementary to each other. Gabor convolution focuses on modeling texture information while classical convolution may not be proficient at, and classical convolution module has a general feature representation ability and it can enhance this ability combined with Gabor convolution module.

With/without gloss: We also evaluate the effectiveness of the designed regularizer for Gabor convolution as Eq. (12). Compared with the setting without regularizer loss, it can be easily found that it improves the performance of person re-identification with regularizer loss. For instance, with the regularizer loss, 2.6 rank-1 performance is improved when 'struct.1' is trained with hard-mining triplet loss, and 1.1 rank-1 performance is improved when 'truct.2' is trained with hard-mining triplet loss. This indicates that it is effective to regularize the Gabor convolution with the proposed 'gloss'. As the regularizer constrains the range of each parameter in Gabor convolution and parameters will tend to effective when trained with the regularizer loss.

Table 4
Comparison of our method's performance on the Market-1501 dataset.

Models	rank-1	rank-5	rank-10	mAP(%)
PersonNet [13]	37.21	–	–	18.57
SSDAL+XQDA [57]	39.4	–	–	19.6
MST-CNN [58]	45.1	70.1	78.4	–
Hybrid [19]	48.15	–	–	29.94
HistLoss [59]	59.47	80.73	86.94	–
CAN(VGG-16) [22]	60.3	–	–	35.90
Gated [15]	65.88	–	–	39.55
MR B-CNN [60]	66.36	85.01	90.17	41.17
P2S [8]	70.72	–	–	45.5
CADL [16]	73.84	–	–	47.11
SpindleNet [23]	76.9	91.5	95.6	–
PIE [24]	79.33	90.76	94.41	55.95
Resnet50(I+V) [61]	79.51	–	–	59.87
MSCAN [17]	80.31	–	–	57.53
Part-Aligned [18]	81.0	92.0	94.7	63.4
SVDNet [62]	82.3	–	–	66.07
DSR [63]	83.58	–	–	64.25
LSRO [64]	83.97	–	–	66.07
PDC [25]	84.14	92.73	94.92	63.14
HAP2S_E [11]	84.20	–	–	69.76
HAP2S_P [11]	84.59	–	–	69.43
JLML [56]	85.1	–	–	65.5
TriNet [7]	86.67	93.38	–	81.07
DML [52]	87.73	–	–	68.83
PSE [53]	87.7	–	–	69.0
CamStyle [54]	88.12	–	–	68.72
AWTL [55]	89.46	–	–	75.67
Baseline	85.7	94.3	96.2	69.5
Gconv(ours)	88.1	95.1	96.8	72.3

4.5. Comparison with state-of-the-arts

In this section, we will show the result of our method compared with existing published state-of-the-arts methods on Markert-1501, DukeMTMC-REID and CUHK03 (both labeled and detected are included). Besides, brief analyses about the advantages and disadvantages are given.

Evaluation on Market-1501: Table 4 shows the comparison results of our method to the state-of-the-arts on Market-1501 dataset. We collect 26 state-of-the-art deep models that have been evaluate on this dataset for person re-identification, including most recently proposed DML [52], PSE [53], CamStyle [54] and AWTL [55]. Some traditional hand-designed features and metric learning methods are not listed by consideration the different pipeline of our method. We achieve 88.1 rank-1 performance and 72.3 mAP performance on this dataset. From Table 4, we can find that our method performs better than most of these state-of-the-art models (23 models in total from Table 4). Such as it improves 3.0 rank-1 than JLML [56] and 3.96 rank-1 performance than PDC [25]. This shows the effectiveness of Gabor convolution module for representation learning for person re-identification. We have to note that the proposed method does not achieve the best performance than the most state-of-the-art method AWTL [55] As we have claimed before, we only use ResNet-50 network for evaluation and take no other kinds of prior information into consideration. However, all the state-of-the-arts methods achieve the best performances by utilizing certain prior information or some special structure beside the baseline structure. For example, AWTL [55] adopts certain attention structures for improving performance. *But based on the proposed Gabor structure, it improves the baseline method 3.6 rank1 performance and 2.8 mAP, which shows the effectiveness of the proposed method.*

Evaluation on DukeMTMC-REID: Table 5 shows the comparison results of our method to the state-of-the-arts on DukeMTMC-REID dataset. As it is a recently proposed dataset, methods evaluated on

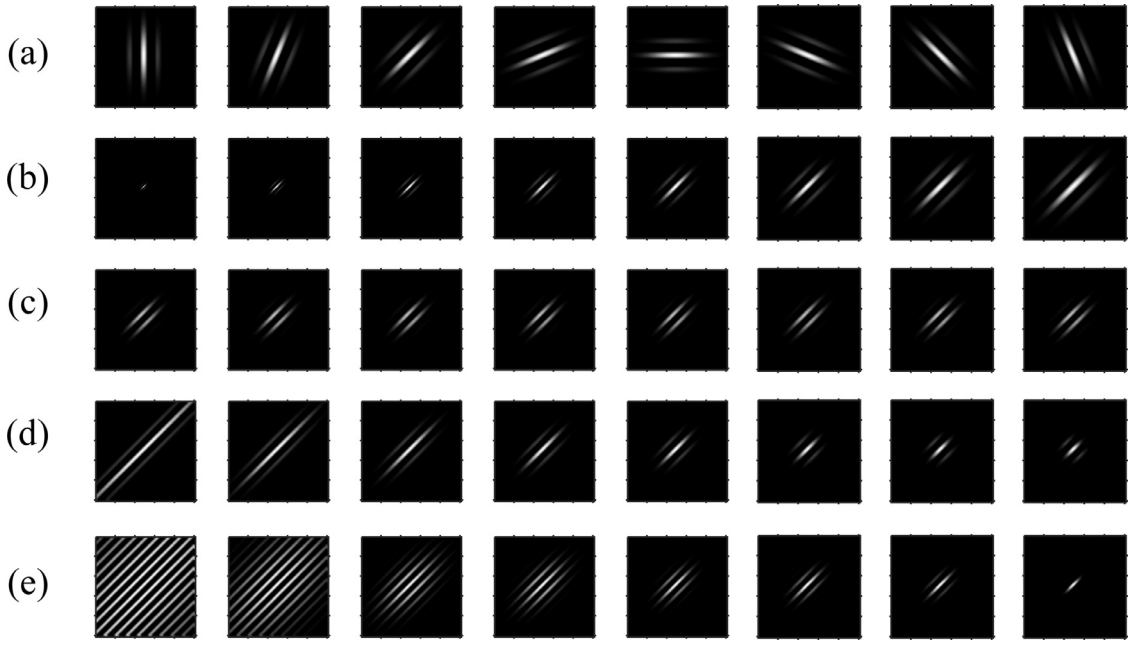

Fig. 2. Gabor filter images with different parameters: (a) Different orientation θ, (b) Different wavelength λ, (c) Different phase offset ϕ, (d) Different aspect ratio γ and (e) Different bandwidth b.

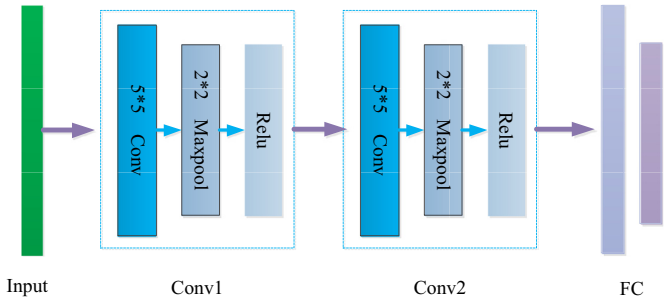

Fig. 3. The model used to evaluate the effectiveness of Gabor convolution module on image classification datasets MNIST and CIFAR-10. Conv1 and Conv2 can be either Gabor convolution (real part or image part) or classical convolution.

Table 5
Comparison of our method's performance on the DukeMTMC-ReID.

Models	rank-1	rank-5	rank-10	mAP(%)
CamStyle [54]	75.27	–	–	53.48
HAP2S_P [11]	75.94	–	–	60.64
HAP2S_E [11]	76.08	–	–	59.58
SVDNet [62]	76.7	–	–	56.8
PSE [53]	79.8	–	–	62.0
AWTL [55]	79.80	–	–	63.40
Baseline	75.7	87.4	90.6	58.6
Gconv(ours)	77.3	88.1	91.3	61.7

this dataset are not as much as Market-1501. We list 7 state-of-the-art methods in Table 5. Our method achieves 77.3 rank-1 accuracy and 61.7 mAP performance. Similar to the results on Market-1501, the proposed methods achieve better performance than CamStyle [54], HAP2S_P [11]and SVDNet [62], while it degrades the performance than PSE [53], AWTL [55].

As have discussed on Market-1501 evaluation section, the reasons for our method's lower performance than the best one are same. The backbone model is undistinguished and some prior information are not took into consideration. Note that, we only evaluate the performance using a fundamental structure Resnet-50, which no other structures are used such as AWTL. It improves the baseline 3.1 mAP performance and 1.6 rank-1 performance as well, which shows the effectiveness of the proposed Gabor structure.

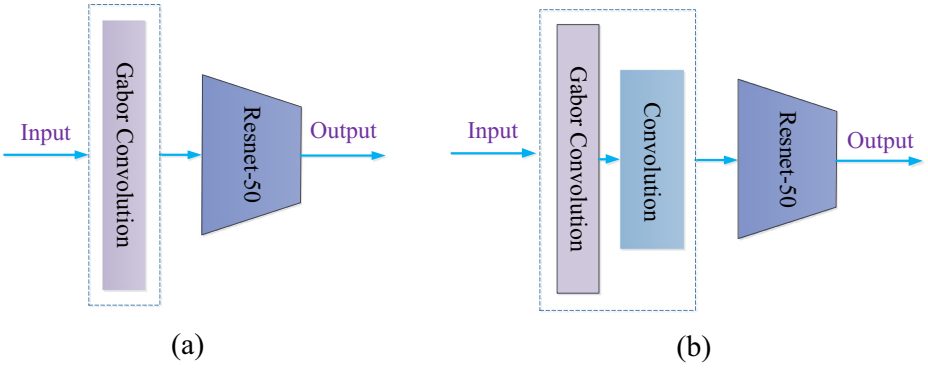

Fig. 4. Two variants of Resnet-50 network used for person re-identification. (a) Replace the first convolution module with Gabor convolution module, (b) Replace the first convolution module with the combination of a Gabor covolution module and a convolution module.

Table 6
Comparison of our method's performance on the CUHK03 dataset with labeled setting to state-of-the-arts.

Models	rank-1	rank-5	rank-10	rank-20	mAP(%)
Re-ranking+IDE [65]	61.6	–	–	–	67.6
NFST+Fusion [66]	62.55	90.05	94.80	98.10	–
Hybrid [19]	63.23	89.95	92.73	97.55	–
PersonNet [13]	64.80	89.40	94,92	98.20	–
HistLoss [59]	65.77	92.85	97.62	99.43	–
MR B-CNN [60]	69.7	93.37	98.91	–	–
MSCAN [17]	74.21	94.33	97.54	99.25	–
SSM [67]	76.6	94.6	98.0	–	–
CAN(VGG-16) [22]	77.6	95.2	99.3	100	–
WARCA [68]	78.38	94.55	–	–	–
JLML [56]	83.2	98.0	99.4	–	–
Part-Aligned [18]	85.4	97.6	99.4	99.9	–
SpindleNet [23]	88.5	97.8	98.6	99.2	–
Baseline	83.8	97.9	99.4	99.8	90.0
Gconv(ours)	85.9	98.3	99.3	99.7	91.5

Table 7
Comparison of our method's performance on the CUHK03 dataset with detected setting to state-of-the-arts.

Models	rank-1	rank-5	rank-10	rank-20	mAP(%)
NFST+Fusion [66]	54.70	84.75	94.80	95.20	–
S-LSTM [37]	57.3	80.1	88.3	–	–
Re-ranking+IDE [65]	58.5	–	–	–	64.7
MR B-CNN [60]	63.67	89.15	94.66	–	–
PIE [24]	67.10	92.20	96.60	98.10	71.32
MSCAN [17]	67.99	91.04	95.36	97.83	–
Gated [15]	68.1	88.1	94.6	–	58.84
CAN(VGG-16) [22]	69.2	88.5	94.1	97.8	–
SSM [67]	72.7	92.4	96.1	–	–
PDC [25]	78.29	94.83	97.15	98.43	
JLML [56]	80.6	96.9	98.7	–	–
Part-Aligned [18]	81.6	97.3	98.4	99.5	–
SVDNet [62]	81.8	–	–	–	84.8
Resnet50(I+V) [61]	83.4	97.1	98.7	–	86.4
LSRO [64]	84.6	97.6	98.9	–	87.4
CRAFT [69]	87.5	94.7	98.7	99.5	–
Baseline	81.9	96.9	98.1	98.9	88.5
Gconv(ours)	83.1	96.4	98.0	98.9	89.0

Evaluation on CUHK03: We also evaluate our method on CUHK03 dataset and both labeled and detected settings are used. Tables 6 and 7 show the comparison results of our method to the state-of-the-arts on CUHK03 labeled and CUHK03 detected dataset, respectively. As some methods only are evaluated on one of the settings, so some methods are not listed in both tables. We list 12 state-of-the-arts deep learning based methods on CUHK03 labeled dataset and 16 methods on CUHK03 detected dataset. We achieve 85.9 rank-1 performance and 91.5 mAP performance on CUHK03 labeled dataset, and rank-1 and mAP on CUHK03 detected dataset. On both settings, the proposed methods achieve better performance than most of the state-of-the-art methods. Of course, due to there are no special module for modeling prior information such as pose and human parsing, the proposed method does not achieve the best performance. We improve the baseline 2.1 rank-1 performance and 1.5 mAP performance on labeled CUHK03 dataset and 2.2 rank1-1 performance and 0.5 mAP performance on detected CUHK03 dataset.

Note that, *we only use a fundamental structure rather than carefully designed structure for person re-identification to verify the effectiveness of the proposed Gabor convolution*. So it cannot perform better than the state-of-the-arts. However, we argue the Gabor structure also shows its superiority as follows. First, compared with attention based methods that need a careful train process, the proposed Gabor convolution takes no more efforts than classical convolution. Second, the structure of Gabor convolution is simple and can be interpreted easily than attention based methods.

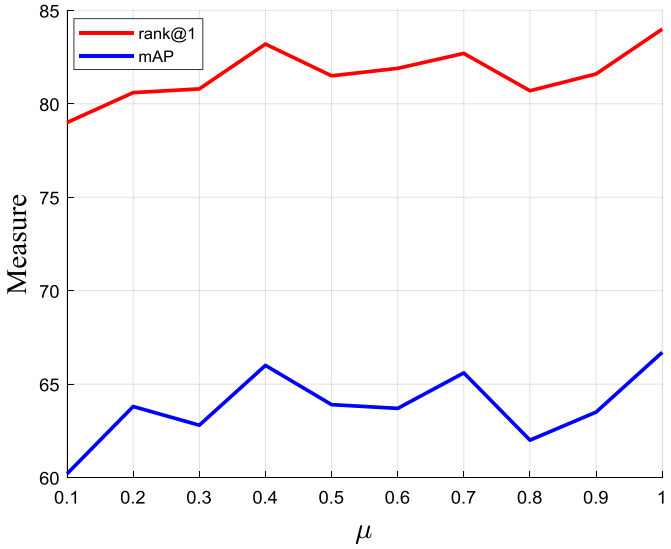

Fig. 5. The rank1 values and mAP values with the change of μ.

4.6. Parameter experiments

In this subsection, we will evaluate how the hyper-parameters influence the performance of our model. Note that there are many hyper-parameters in regularizer loss function such as C_λ in Eq. (9) and $C_{\hat{b}}$ in Eq. (11), m_λ in Eq. (9) and m_γ in Eq. (10). But these hyper-parameters are not important as they only constrain the range of the parameters of Gabor function. In other word, these hyper-parameters are trivial as they are used to constrain the Gabor convolution's parameters to a set of priori values. But when relaxing the priori values, they will not influence much to the Gabor convolution's parameters as the module's parameters are optimally learned. Some relaxation of these hyper-parameter influences little performance and we will not give an evaluation to them.

Besides, the number of Gabor convolution filters is another parameter. In experiments, we set it as the same as the Resnet-50's first convolution layer. This is because (1) it will not change the baseline structure of Resnet-50, (2) the influence of this parameter will not be prominent with the deeper of the network goes as have discussed in previous pipelines.

However, the hyper-parameter μ in Eq. (13) is very import because it influences the value of loss function. So we will evaluate how the trade-off parameter μ based on 'struct.1' network. We take the value of μ ranged from 0.1 to 1.0 and report the rank1 value and mAPs. Fig. 5 shows the results.

From Fig. 5, we can find that when $\mu = 1.0$ we can get the largest rank1 value as well as mAP value. It can be found the hyper-parameter μ is important as the rank-1 values range from 79 to 84 which is a large range for the performance of person re-identification. On the other hand, when the values of μ range from 0.4 to 1.0, the changes of rank-1 value are relatively small. Generally, a larger μ gets a larger measure but with some certain changes. Without loss the generality, we set $\mu = 1$ in the whole setting.

4.7. Visualization of the learned Gabor kernels

After we have learned the parameters of the Gabor convolution through a deep neural network, a question arises naturally that what we have learned and how the kernels look like. In this

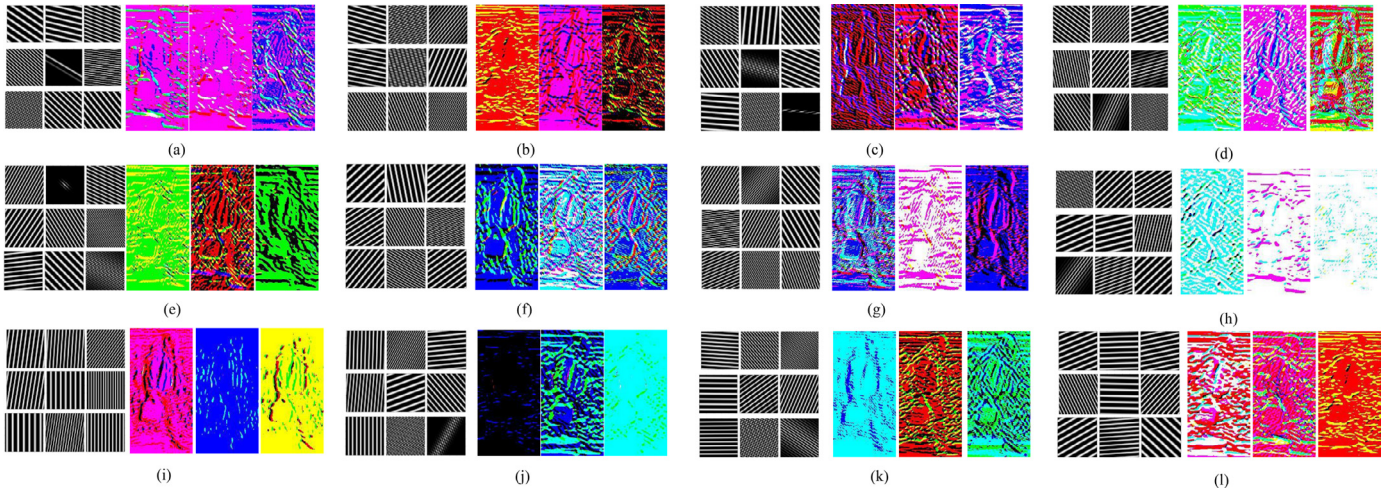

Fig. 6. The learned Gabor convolution kernels and corresponding feature maps. (a)–(l) are 12 group of maps of kernels according to the orientations of three channels of each kernel. The left part of each group represents three Gabor convolution kernels (from top to bottom) and the right part corresponds feature maps of the left Gabor filters (from left to right). The image used to get feature maps is the same as the person image shown in Fig. 1.

section, we will visualize the learned kernels and corresponding feature maps.

According to the learned parameters $\hat{\mathbf{\Theta}} = \{\lambda, \theta, \phi, \hat{b}, \gamma\}$, we generate a chunk of Gabor convolution kernels of size 50×50 based on Gabor function. Note that every three continuous kernels correspond to three channels of a RGB image. So we treat every three continuous kernels as a whole part and get a three-channels feature map by convoluting the three kernels with every channel of a RGB image. Next, we will refer to the Gabor convolution kernel as three-channels one for simplicity.

As shown in Fig. 6, we divide the learned kernels into 12 groups according to the orientation of three channels of each kernel, and the left part of each group represents three Gabor convolution kernels (from top to bottom) and the right part corresponds feature maps of the left Gabor filters (from left to right). The image used to get feature maps is the same as the person image shown in Fig. 1. In group (a)–(d) all the bands in the first channel of the Gabor convolution kernel lean to the left while in group (e)–(h) all the bands in the first channel lean to the right. In fact, all the groups (a)–(h) are combinations of left-leaning and right-leaning of bands in three channels. For example, the bands of three channels in group (f) lean to the right, the left and the right respectively. Nevertheless, maps in group (i)–(l) are miscellaneous and less regularity. In group (i) all the bands of the three channels are nearly vertical while in group (j) only the bands in the first channel are nearly vertical. In group (k) the bands in the first channel are horizontal and in group (l) the bands in the second channel are horizontal. In fact, there are a few other kinds of Gabor kernels but they have no distinct properties as analyzed in the next paragraphs and hence we do not show them in the figure.

From Fig. 6, we can observe that all the feature maps are rich of texture information. For instance, features maps in (g) show texture information along to the edges of the pedestrian while feature maps in group (l) show horizontal texture information. From this observation, we can draw that the learned Gabor convolution kernels have the ability of capturing diverse texture information automatically as expected. Furthermore, focusing on each group, we can observe that the feature maps in the same group have similar texture information. For example, in group (f), all the feature maps have similar contour texture information and look like each other, while in group (i) though they are dissimilar to each other in color, they share similar texture information that can be regarded as shape texture without background. This implies that the learned Gabor convolution kernels in the same group have a close relationship with each other and are complementary each other to model the diverse texture information.

Compared with traditional Gabor filters, the proposed Gabor convolution has three advantages. First, Gabor convolution can go through more parameters' space as it learns the parameters from the a large predefined range. Second, Gabor Convolution can get optimal parameters compared with the hand-designed parameters. Third, Gabor convolution can be learned to be adaptive to input data while hand-designed parameters cannot be. From Fig. 6 we can see there are various parameters of different shapes, which proves the large parameter space. Besides, after the Gabor convolution, the feature maps are of various texture information and are adaptive to pedestrian images, while head designed parameters have limited texture information. This also shows the optimal and adaptive property of Gabor convolution.

All the above observations and conclusions demonstrate that the proposed Gabor convolution module can have an effective ability to interpret as a texture extractor, which classical convolution module has not.

5. Conclusion and future works

In this paper, we propose a new convolution module for deep neural network models for person re-identification name Gabor convolution. The proposed Gabor convolution module show good interpretability for deep neural networks models as well as superior performance in low-level layers of a network. Apart from the designation of the new module, a new regularizer loss function based on hinge function is proposed to constrain the range of each parameter of the Gabor convolution kernels. Additionally, we evaluate how the Gabor convolution can be used for representation learning for person re-identification and find that the combination of Gabor convolution and classical convolution can achieve improved performance.

As far as application, we argue that the proposed Gabor convolution module can be applied to general computer vision tasks by embedding it into the low layers of a deep neural network. Especially for some tasks that texture information is dominant among all the information, such as texture classification.

For future works, we will improve some limitations of the proposed method. As we show in the experimental section, only the baseline network ResNet-50 is used for evaluation and it do not

achieve the best performance though it is comparable with state-of-the-arts. Besides, only two structures are designed for representation learning for person re-identification and some prior information are not used as attention-based models. So in future works, we will improve the performance of person re-identification in the following two aspects. The first one is to evaluate the performances of other kinds of networks by embedding Gabor convolution module. The second on is to design new structures based on Gabor convolution and prior information such as pose and human parsing.

Declaration of Competing Interest

None.

References

[1] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: Proceedings of European Conference on Computer Vision, Springer, 2008, pp. 262–275.

[2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2360–2367.

[3] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015, pp. 2197–2206, doi:10.1109/CVPR.2015.7298832.

[4] T. Matsukawa, T. Okabe, E. Suzuki, Y. Sato, Hierarchical gaussian descriptor for person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016, pp. 1363–1372.

[5] H. Dong, L. Ping, Z. Shan, C. Liu, J. Yi, S. Gong, Person re-identification by enhanced local maximal occurrence representation and generalized similarity metric learning, Neurocomputing 307 (2018) 25–37.

[6] D. Cheng, Y. Gong, S. Zhou, J. Wang, N. Zheng, Person re-identification by multi-channel parts-based cnn with improved triplet loss function, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1335–1344.

[7] A. Hermans, L. Beyer, B. Leibe, In Defense of the Triplet Loss for Person Re-identification, arXiv preprint arXiv:1703.07737 (2017).

[8] S. Zhou, J. Wang, J. Wang, Y. Gong, N. Zheng, Point to set similarity based deep feature learning for person reidentification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 6, 2017, pp. 3741–3750.

[9] W. Chen, X. Chen, J. Zhang, K. Huang, Beyond triplet loss: a deep quadruplet network for person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2, 2017, pp. 403–412.

[10] T. Xiao, S. Li, B. Wang, L. Lin, X. Wang, Joint detection and identification feature learning for person search, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3376–3385.

[11] R. Yu, Z. Dou, S. Bai, Z. Zhang, Y. Xu, X. Bai, Hard-aware point-to-set deep metric for person re-identification, in: Proceedings of European Conference on Computer Vision, 2018, pp. 1–17.

[12] E. Ahmed, M. Jones, T.K. Marks, An improved deep learning architecture for person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3908–3916.

[13] L. Wu, C. Shen, A.v. d. Hengel, Personnet: Person Re-identification with Deep Convolutional Neural Networks, arXiv preprint arXiv:1601.07255 (2016).

[14] C. Zhao, Y. Chen, X. Wang, W.K. Wong, D. Miao, J. Lei, C. Zhao, Y. Chen, X. Wang, W.K. Wong, Kernelized random kiss metric learning for person re-identification, Neurocomputing 275 (2017) 403–417.

[15] R.R. Varior, M. Haloi, G. Wang, Gated siamese convolutional neural network architecture for human re-identification, in: Proceedings of European Conference on Computer Vision, Springer, 2016, pp. 791–808.

[16] J. Lin, L. Ren, J. Lu, J. Feng, J. Zhou, Consistent-aware deep learning for person re-identification in a camera network, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 6, 2017, pp. 5771–5780.

[17] D. Li, X. Chen, Z. Zhang, K. Huang, Learning deep context-aware features over body and latent parts for person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 384–393.

[18] L. Zhao, X. Li, Y. Zhuang, J. Wang, Deeply-learned part-aligned representations for person re-identification., in: Proceedings of IEEE International Conference on Computer Vision, 2017, pp. 3239–3248.

[19] L. Wu, C. Shen, A. van den Hengel, Deep linear discriminant analysis on fisher networks: a hybrid architecture for person re-identification, Pattern Recognit. 65 (2017) 238–250.

[20] X. Liu, H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, X. Wang, Hydraplus-net: attentive deep features for pedestrian analysis, in: Proceedings of IEEE International Conference on Computer Vision, 2017, pp. 350–359.

[21] J. Si, H. Zhang, C. Li, J. Kuen, X. Kong, A.C. Kot, G. Wang, Dual attention matching network for context-aware feature sequence based person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5363–5372.

[22] H. Liu, J. Feng, M. Qi, J. Jiang, S. Yan, End-to-end comparative attention networks for person re-identification, IEEE Trans. Image Process. 26 (7) (2017) 3492–3506.

[23] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, X. Tang, Spindle net: person re-identification with human body region guided feature decomposition and fusion, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1077–1085.

[24] L. Zheng, Y. Huang, H. Lu, Y. Yang, Pose Invariant Embedding for Deep Person Re-identification, arXiv preprint arXiv:1701.07732 (2017).

[25] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, Q. Tian, Pose-driven deep convolutional model for person re-identification, in: Proceedings of IEEE International Conference on Computer Vision, IEEE, 2017, pp. 3980–3989.

[26] M.M. Kalayeh, E. Basaran, M. Gökmen, M.E. Kamasak, M. Shah, Human semantic parsing for person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1062–1071.

[27] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, J. Opt. Soc. Am. A 2 (7) (1985) 1160–1169.

[28] B. Ma, Y. Su, F. Jurie, Bicov: a novel image representation for person re-identification and face verification, in: Proceedings of British Machive Vision Conference, 2012, pp. 11–pages.

[29] C. Liu, S. Gong, C.C. Loy, X. Lin, Person re-identification: what features are important? in: Proceedings of European Conference on Computer Vision, Springer, 2012, pp. 391–401.

[30] B. Ma, Y. Su, F. Jurie, Covariance descriptor based on bio-inspired features for person re-identification and face verification, Image Vis. Comput. 32 (6–7) (2014) 379–390.

[31] W. Li, X. Wang, Locally aligned feature transforms across views, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3594–3601.

[32] Q. Wang, J. Gao, X. Li, Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes, IEEE Trans. Image Process. 28 (9) (2019) 4376–4386.

[33] Q. Wang, W. Jia, X. Li, Robust hierarchical deep learning for vehicular management, IEEE Trans. Circuits Syst. Video Technol. 19 (5) (2019) 4148–4156.

[34] Q. Wang, M. Chen, F. Nie, X. Li, Detecting coherent groups in crowd scenes by multiview clustering, IEEE Trans. Pattern Anal. Mach. Intell. doi:10.1109/TPAMI.2018.2875002.

[35] J. Yu, X. Yang, F. Gao, D. Tao, Deep multimodal distance metric learning using click constraints for image ranking, IEEE Trans. Cybern. 47 (12) (2017) 4014–4024.

[36] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: a unified embedding for face recognition and clustering, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.

[37] R.R. Varior, B. Shuai, J. Lu, D. Xu, G. Wang, A siamese long short-term memory architecture for human re-identification, in: Proceedings of European Conference on Computer Vision, Springer, 2016, pp. 135–153.

[38] Y. Huang, J. Xu, Q. Wu, Z. Zheng, Z. Zhang, J. Zhang, Multi-pseudo regularized label for generated data in person re-identification, IEEE Trans. Image Process. 28 (3) (2019) 1391–1403.

[39] Z. Zhang, J. Chen, Q. Wu, L. Shao, Gii representation-based cross-view gait recognition by discriminative projection with list-wise constraints, IEEE Trans. Cybern. 48 (10) (2017) 2395–2947.

[40] X. Peng, J. Feng, S. Xiao, W. Yau, J.T. Zhou, S. Yang, Structured autoencoders for subspace clustering, IEEE Trans. Image Process. 27 (10) (2018) 5076–5086.

[41] Z. Huang, H. Zhu, J.T. Zhou, X. Peng, Multiple marginal fisher analysis, IEEE Trans. Ind. Electron. 66 (12) (2019) 9798–9807.

[42] P. Kruizinga, N. Petkov, Nonlinear operator for oriented texture, IEEE Trans. Image Process. 8 (10) (1999) 1395–1407.

[43] N. Petkov, Biologically motivated computationally intensive approaches to image pattern recognition, Futur. Gener. Comput. Syst. 11 (4–5) (1995) 451–465.

[44] N. Petkov, P. Kruizinga, Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells, Biol. Cybern. 76 (2) (1997) 83–96.

[45] N.P. S.E. Grigorescu, P. Kruizinga, Comparison of texture features based on gabor filters, IEEE Trans. Image Process. 11 (10) (2002) 1160–1167.

[46] N. Petkov, M.A. Westenberg, Suppression of contour perception by band-limited noise and its relation to non-classical receptive field inhibition, Biol. Cybern. 88 (10) (2003) 236–246.

[47] C. Grigorescu, N. Petkov, M.A. Westenberg, Contour detection based on non-classical receptive field inhibition, IEEE Trans. Image Process. 12 (7) (2003) 729–739.

[48] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: a benchmark, in: Proceedings of IEEE International Conference on Computer Vision, 2015, pp. 1116–1124.

[49] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: Proceedings of European Conference on Computer Vision, Springer, 2016, pp. 17–35.

[50] W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: deep filter pairing neural network for person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 152–159.

[51] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[52] Y. Zhang, T. Xiang, T.M. Hospedales, H. Lu, Deep mutual learning, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4320–4328.
[53] M.S. Sarfraz, A. Schumann, A. Eberle, R. Stiefelhagen, A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1–10.
[54] Z. Zhong, L. Zheng, Z. Zheng, S. Li, Y. Yang, Camera style adaptation for person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5157–5166.
[55] E. Ristani, C. Tomasi, Features for multi-target multi-camera tracking and re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1–11.
[56] W. Li, X. Zhu, S. Gong, Person re-identification by deep joint learning of multi-loss classification, in: Proceedings of International Joint Conferences on Artifical Intelligence, 2017.
[57] C. Su, S. Zhang, J. Xing, W. Gao, Q. Tian, Deep attributes driven multi-camera person re-identification, in: Proceedings of European Conference on Computer Vision, Springer, 2016, pp. 475–491.
[58] J. Liu, Z.-J. Zha, Q. Tian, D. Liu, T. Yao, Q. Ling, T. Mei, Multi-scale triplet cnn for person re-identification, in: Proceedings of ACM Proc. Multimedia Conf., ACM, 2016, pp. 192–196.
[59] E. Ustinova, V. Lempitsky, Learning deep embeddings with histogram loss, in: Proceedings of Advances in Neural Information Processing Systems, 2016, pp. 4170–4178.
[60] E. Ustinova, Y. Ganin, V. Lempitsky, Multi-region bilinear convolutional neural networks for person re-identification, in: Proceedings of Advanced Video and Signal Based Surveillance, IEEE, 2017, pp. 1–6.
[61] Z. Zheng, L. Zheng, Y. Yang, A discriminatively learned cnn embedding for person reidentification, ACM Trans. Multi. Comput. Commun. Appli. 14 (1) (2017) 13.
[62] Y. Sun, L. Zheng, W. Deng, S. Wang, Svdnet for pedestrian retrieval, in: Proceedings of International Conference on Computer Vision, 2017, pp. 3800–3808.
[63] L. He, J. Liang, H. Li, Z. Sun, Deep spatial feature reconstruction for partial person re-identification: alignment-free approach, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7073–7082.
[64] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro, in: Proceedings of IEEE International Conference on Computer Vision, 2017, pp. 3754–3762.
[65] Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2017, pp. 3652–3661.
[66] L. Zhang, T. Xiang, S. Gong, Learning a discriminative null space for person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1239–1248.
[67] S. Bai, X. Bai, Q. Tian, Scalable person re-identification on supervised smoothed manifold, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 6, 2017, pp. 2530–2539.
[68] C. Jose, F. Fleuret, Scalable metric learning via weighted approximate rank component analysis, in: Proceedings of European Conference on Computer Vision, Springer, 2016, pp. 875–890.
[69] Y.-C. Chen, X. Zhu, W.-S. Zheng, J.-H. Lai, Person re-identification by camera correlation aware feature augmentation, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2) (2018) 392–408.

Yuan Yuan (M'05-SM'09) is currently a Full Professor with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. She has authored or coauthored over 150 papers, including about 100 in reputable journals, such as the IEEE TRANSACTIONS AND PATTERN RECOGNITION, and also conference papers in CVPR, BMVC, ICIP, and ICASSP. Her current research interests include visual information processing and image/video content analysis.

Jian'an Zhang received the B.E. degree in information and computing science from the Ocean University of China, Qing Dao, China, 2015. He is currently persuing the Ph.D. degree with the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xian. His research interests include computer vision and pattern recognition.

Qi Wang (M'15-SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science, with the Unmanned System Research Institute, and with the Center for OPTical IMageryAnalysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.