Pattern Recognition Letters 138 (2020) 424–430

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Weighted-capsule routing via a fuzzy gaussian model

Ouafa Amira[a], Shuang Xu[a], Fang Du[b], Jiangshe Zhang[a,*], Chunxia Zhang[a], Rafik Hamza[c]

[a] *School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an Shaanxi 710049, China*
[b] *School of Mathematics and Statistics, Chang'an university, Xi'an Shaanxi 710064, China*
[c] *Big Data Analytics Laboratory, National Institute of Information and Communications Technology, Japan*

ARTICLE INFO

Article history:
Received 9 January 2020
Revised 4 August 2020
Accepted 9 August 2020
Available online 10 August 2020

Keywords:
Capsule network
Deep learning
Image classification
Weighted capsule fuzzy gaussian model
Pose loss

ABSTRACT

Capsule network (CapsNet) is a novel architecture that takes into account the hierarchical pose relationships between object parts, which had achieved desirable results on image classification. EM-Routing (EM-R) used in CapsNet is the process of assigning child capsules (parts) to each parent capsule (objects) based on a level of agreement, which is similar to the fuzzy clustering process. However, CapsNet still struggles with backgrounds and the presence of noise. In this paper, a new routing algorithm based on a weighted capsule fuzzy gaussian model (WCFGM-R) and a pose loss function are proposed. The proposed algorithm aims to prohibit atypical child capsules from contaminating the parent capsules by incorporating the activations of capsules in a lower layer as weights that play the role of precision. The pose loss provides the best inter-class separation and improves the ability of pattern classification. Indeed, the experimental analyses demonstrate that CapsNet with WCFGM-R outperforms the CapsNet with EM-R in which it shows excellent results on three datasets (MNIST-bg-img, MNIST-bg-rnd, and CIFAR10).

1. Introduction

A convolutional neural network (CNN) [1] is commonly used in analyzing visual imagery and has achieved state of the art on the image classification task. Although CNNs can efficiently detect useful features, they can not explore the spatial relationships between features. The lack of relevant information and the transition invariance caused by the pooling operation affect the classification results.

Capsule network (CapsNet) is proposed by Sabour et al. [2], and extended by Hinton et al. [3] to overcome CNN's impediments. The main idea is to learn the part-whole relationships (e.g., perspective, size, scale, orientation) between the observed objects or object parts to achieve the translation equivariance. Due to the equivariance property, capsules can recognize an object with specific variations (rotated, scaled, etc.) without the need to train a model with these variations.

The routing by agreement process is an interaction that occurs between each two adjacent capsule layers, which is an information selection mechanism. This mechanism attempts to ensure that the outputs of lower layer capsules are sent to the appropriate higher layer capsules. In [3], the Expectation Maximization algorithm has been proposed as a routing technique (EM-R), in which it assigns the output of each capsule in the lower layer to the appropriate capsule in the layer above. Despite the outperformance of CapsNet with EM-R, it still struggles from the presence of noise and backgrounds, and it may misclassify many examples from different classes into the same class. Firstly, the responsibilities used in EM-R indicate which capsule's pose is the closest to the child capsule but not any of them is close at all. Typically, an outlying child capsule which is far from any parent capsule's pose will have a small responsibility while it may still be assigned to the closest parent capsule. This issue may influence the classification results, especially in case of data that contains backgrounds and noise. Secondly, if the inputs have some similarities in the descriptive characteristics, CapsNet may extract similar poses of different classes, which can confuse the classifier.

To solve the issues mentioned above, a routing algorithm based on a weighted capsule fuzzy Gaussian model (WCFGM-R) and a pose loss function are proposed in this paper. We sum up the main contributions of this paper as follows.

(a) The activations of the child capsules are used as weights that show the importance of them. These weights play the role of precisions. Hence, each child capsule with a very small activation will be treated as noise and will have a little influence on the estimation of the means (poses) and variances of the parent capsules.

[*] Corresponding author.
E-mail address: jszhang@mail.xjtu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.patrec.2020.08.009
0167-8655/© 2020 Elsevier B.V. All rights reserved.

(b) To provide better inter-class separation characteristics, we propose a pose loss function which can be combined with the original loss function to achieve higher classification accuracies.
(c) Several experiments are provided to test the performance of CapsNet with WCFGM-R. The results show that the CapsNet with WCFGM-R outperforms the CapsNet with EM-R in terms of accuracy.

The rest of the paper is organized as follows. Section 2 introduces related work about the capsule network and its application in different fields. Section 3 introduces the proposed routing algorithm. Section 4 presents the proposed pose loss. Section 5 introduces a comparison study between capsule network with the proposed routing and that one with EM-R. Section 6 concludes this paper.

2. Related work

Hinton et al. [4] proposed a new concept called capsules that provides a simple way to recognize the wholes by recognizing their parts while simultaneously preserving the spatial information. The capsule's output represents the different properties of the same entity. Sabour et al. [2] proposed the so-called dynamic routing used between every two adjacent layers for capsules assignment as the first implementation of the CapsNet.

Hinton et al. [3] extended a new version of capsules, where the capsule's output is a matrix instead of a vector to learn the relationship between the entity and the pose. Moreover, the Expectation-Maximization (EM) algorithm has been used as a new routing technique to fit a mixture of Gaussians. Zhang et al. [5] proposed two fast routing methods within the framework of weighted kernel density estimation, which provide efficient solutions with much higher capacity for large scale vision tasks. Xiang et al. [6] proposed a multi-scale capsule network MS-Capsnet, which shows an efficiency for feature representations in image classification. Lately, Rebeiro et al. [7] proposed a new capsule routing algorithm, which is derived by the variational Bayes for fitting a mixture of transforming Gaussians, and it treats some of the inherent weaknesses of EM-R.

Recently, researchers have introduced CapsNet into many fields. Parnian et al. employed [8] CapsNets to classify the brain tumor. A convolutional deconvolutional capsule network has been proposed by Rodney et al. [9], in which it extends capsule networks to object segmentation problem. Jaiswal et al. proposed the Capsule-GAN [10], where the capsule networks replace the standard convolutional neural networks and plays the role of discriminators. Du et al. [11] proposed a capsule-based hybrid neural network model for sentiment classification of short text. Zhu et al. [12] proposed a novel capsule network with an inception block and a regression branch for bearing fault diagnosis. Singh et al. proposed a Direct-CapsNet [13], to address a very low resolution digit and face recognition. In [14], a traffic classification method based on capsule network has been proposed to form an efficient classification mechanism that separates the traffic.

Although there are attempts to improve the CapsNet with different aspects and tasks, only a few algorithms have been proposed to improve the original EM routing [3].

3. Proposed routing (WCFGM)

Fuzzy c-means (FCM) is a well known non-parametric clustering method [15]. It has been extensively applied in a variety of substantive areas [16–19]. Since there are a lot of similar characteristics between FCM and the Gaussian mixture models (GMMs), many authors have strived to combine the FCM algorithm with GMMs [20–22].

In this section, a new clustering or routing algorithm is proposed for capsule networks. In what follows, capsule(s) in the lower layer is named Childcaps, and capsule(s) in the higher layer is named Parentcaps. Before wading to the proposed routing method, some basic information of capsule networks is reviewed.

Each Capsule Network (CapsNet) consists of ρ layers of capsules ($\rho \ \geq 1$), each layer contains a set of capsules denoted by Ω_{layer}, and each two adjacent layers are connected by a routing procedure. Each Childcaps has a pose matrix $M_j \in R^{4 \times 4}$, and an activation probability, α_j. The pose matrix of instantiation parameters attempts to learn how to encode the relationship of an entity to the viewer. Each Childcaps gives a vote $V_{i|j}$ for the pose of Parentcaps by multiplying its pose matrix M_j with a trainable viewpoint invariant transformation matrix W_{ij}: $V_{i/j} = W_{ij}M_j$. In the following discussions, V_j is used instead of $V_{i|j}$ for simplicity. In EM-R [3], the pose matrix of a Parentcaps M_i is computed by a weighted mean of the votes it received from all Childcaps, where the weights or routing coefficients are the posterior probabilities of each Parentcaps given the occurrence of the Childcaps. In [3], the routing coefficients are tuned for fitting a mixture of Gaussians and can be interpreted as the responsibility that each Parentcaps takes for explaining Childcaps j. These routing coefficients are considered as a measure of agreement between V_j and M_i.

The proposed method treats the following optimization problem:

$$\underbrace{\min}_{(S,\mu)} \left\{ L_{wcfgm}(S,\mu;V) = \sum_{i=1}^{l_h}\sum_{j=1}^{l_l} S_{ij}D_{ij} - \sum_{i=1}^{l_h}\sum_{j=1}^{l_l} S_{ij}\log\gamma_i \right.$$
$$\left. +\theta\sum_{i=1}^{l_h}\sum_{j=1}^{l_l} S_{ij}\log\frac{S_{ij}}{P(i|V_j)} \right\}, \tag{1}$$

where

$$D_{ij} = -\log\mathcal{N}\left(V_j, \mu_i, \frac{(\sigma_i)^2}{\alpha_j}\right) \tag{2}$$

with the constraint

$$\sum_{i=1}^{l_h} S_{ij} = 1, \quad 1 \leq j \leq l_l, \tag{3}$$

where l_l and l_h are the number of capsules in the lower layer and higher layer respectively, S_{ij} is the fuzzy membership degree of the vote of the Childcaps j in the Parentcaps i, D_{ij} is the negative log-likelihood function, $\mathcal{N}(V_j, \mu_i, \frac{(\sigma_i)^2}{\alpha_j})$ is the density function of the vectorized vote V_j with mean μ_i (a vectorized version of i's pose matrix M_i) and variance $(\sigma_i)^2$, α_j is the activation of jth Childcaps or its importance that plays the role of precision and it is different for each one, γ_i is the activation of the ith Parentcaps (considered as a mixing proportion of the ith Gaussian or cluster), and θ is a positive regularization parameter. $P(i|V_j)$ is a posterior probability of V_j to the ith Parentcaps, and by assuming that it follows the Gaussian mixture model, i.e.

$$\mathrm{P}(i|V_j) = \frac{\Pi_i\mathcal{N}\left(V_j|\mu_i, \frac{(\sigma_i)^2}{\alpha_j}\right)}{\sum_{l=1}^{l_h}\Pi_l\mathcal{N}\left(V_j|\mu_l, \frac{(\sigma_l)^2}{\alpha_j}\right)}, \tag{4}$$

here Π_i is the mixture weight, i.e., prior probability of the ith Parentcaps that satisfies $\Pi_i \geq 0 \ \forall i$ and $\sum_{i=1}^{l_h}\Pi_i = 1$. As in [3], in later discussions the activations of Parentcaps γ_i will replace the mixture weights Π_i, $\forall i$.

The objective function in Eq. (1) contains three terms:

1. The first term is the negative log-likelihood multiplied by the fuzzy membership degree, minimizing the objective func-

tion would maximize the log-likelihood function. The activations of Childcaps has been included in the framework of fuzzy Gaussian model as weights α_j, $1 \leq j \leq l_l$ that play the role of precisions. Motivated by [23], the weights are incorporated in a statistical model as the power of likelihood function: $\mathcal{N}\left(V_j, \mu_i, \sigma_i^2\right)^{\alpha_j} \propto \mathcal{N}\left(V_j, \mu_i, \frac{\sigma_i^2}{\alpha_j}\right)$. Moreover, these weights indicate the importance and the relevance of the Childcaps, which allow each of them to be potentially treated differently. The purpose here is to reduce the effect of atypical Childcaps (noise) on estimating the pose of Parentcaps.

2. The second term is the entropy of the activation of Parentcaps γ_i over average of fuzzy memberships S_{ij}. This term gives more interaction between the fuzzy membership degrees and the activations of capsules in higher layer. Minimizing the objective function would minimize this entropy term.
3. The latter one is the Kullback-Leiblure divergence of S_{ij} and $P(i|V_j)$. This term is a regularization of the objective function, where it would equate the fuzzy membership degrees with the posterior probabilities. The parameter θ controls the fuzziness of the partition.

Solving the optimization problem in Eq. (1) leads to the following updating rules:

$$S_{ij} = \frac{\mathrm{P}(i|V_j)\mathcal{N}(V_j|\mu_i, \frac{(\sigma_i)^2}{\alpha_j})^{\frac{1}{\theta}}\gamma_i^{\frac{1}{\theta}}}{\sum_{l=1}^{l_h}\mathrm{P}(l|V_j)\mathcal{N}(V_j|\mu_l, \frac{(\sigma_l)^2}{\alpha_j})^{\frac{1}{\theta}}\gamma_l^{\frac{1}{\theta}}}, \tag{5}$$

According to Eq. (4), and by replacing Π_i with γ_i, we can obtain

$$S_{ij} = \frac{\mathcal{N}(V_j|\mu_i, \frac{(\sigma_i)^2}{\alpha_j})^{1+\frac{1}{\theta}}\gamma_i^{1+\frac{1}{\theta}}}{\sum_{l=1}^{l_h}\mathcal{N}(V_j|\mu_l, \frac{(\sigma_l)^2}{\alpha_j})^{1+\frac{1}{\theta}}\gamma_l^{1+\frac{1}{\theta}}}, \tag{6}$$

$$\mu_i = \frac{\sum_{j=1}^{l_l}\alpha_j S_{ij} V_j}{\sum_{j=1}^{l_l}\alpha_j S_{ij}}, \tag{7}$$

$$\sigma_i^2 = \frac{\sum_{j=1}^{l_l}\alpha_j S_{ij}(V_j - \mu_i)^2}{\sum_{j=1}^{l_l} S_{ij}}, \tag{8}$$

In the proposed routing (Algorithm 1), the above formulas include each dimension h of the vote V_j from the Childcaps j to the Parentcaps i (V_j^h). For each Parentcaps i, the sum of the fuzzy membership degrees over Childcaps ($S_i = \sum_j S_{ij}$) represents the amount of capsules assigned to this Parentcaps or Gaussian i. To compute each Parentcaps activation γ_i, a logistic non-linearity function is applied, that is,

Algorithm 1 Weighted capsule fuzzy gaussian model routing.

```
1. Procedure: WCFGM(α, V)
2. Initialize ∀ $j \in \Omega_{l_l}$, $i \in \Omega_{l_h}$ : $S_{ij} \leftarrow 1/|\Omega_{l_h}|$, fix the parameter $\theta$.
3. for $k$ iterations do
    1. ∀ $i \in \Omega_{l_h}$ : Update($\alpha$, S, V, $i$)
    2. ∀ $j \in \Omega_{l_l}$ : Calculate-S($\mu$, $\sigma$, $\gamma$, V, $j$)
   Output $\gamma$, $M$
5. Procedure: Update($\alpha$, S, V, $i$)
    1. ∀ $h$: Compute the mean for the $h$th dimension of the $i$th gaussian $\mu_i^h$ by using Eq. (7).
    2. ∀ $h$: Compute the variance for the $h$th dimension of the $i$th gaussian $(\sigma_i^2)^h$ by using Eq. (8).
    3. Compute the activation probability of the $i$th Parentcaps $\gamma_i$ by using Eq. (9).
6. Procedure: Calculate-S($\mu$, $\sigma$, $\gamma$, V, $j$)
    1. ∀ $i \in \Omega_{l_h}$: Compute the following probability density function
       $\mathcal{N}\left(V_j^h|\mu_i^h, \frac{(\sigma_i^h)^2}{\alpha_j}\right) = \frac{1}{\sqrt{\prod_h^H 2\pi \frac{(\sigma_i^h)^2}{\alpha_j}}} \exp\left(-\sum_h^H \frac{(V_j^h-\mu_i^h)^2}{2(\sigma_i^h)^2}\right)^{\alpha_j}$
    2. ∀ $i \in \Omega_{l_h}$: Recalculate the fuzzy membership degree of the $j$th Childcaps in the $i$th Parentcaps $S_{ij}$ by using Eq. (6).
```

$$\gamma_i = sigmoid(\lambda(\beta_a - \sum_h(\beta_u + \log(\sigma_i^h))S_i)), \tag{9}$$

where the sum here is over all dimensions of the cost of explaining each dimension, h, of the vote V_j. β_a and β_u are discriminatively learnable parameters, where -β_a represents a fixed cost for coding the mean and the variance of the Parentcaps, and -β_u represents a fixed cost for describing the poses of all Childcaps that are assigned to the Parentcaps (the reader is referred to [3] for more details). λ is an inverse temperature parameter that increases in each iteration with a fixed schedule.

4. The pose loss

Developing an effective loss function is an interesting way to improve the ability of pattern classification. Intuitively, providing the best inter-class separation characteristics is the key. For this purpose, a pose loss is proposed as follows:

$$L_p = \sum_{i \neq t} \|\mu_t - \mu_i\|^2, \tag{10}$$

where μ_t is the pose of the target class, and μ_i is the pose of the ith class. The formulation effectively characterizes the inter-class variations. Therefore, the loss function used in this paper is given by the following:

$$Loss = GL - \xi L_p, \tag{11}$$

where L_p is the introduced pose loss, ξ is a parameter for scaling down the pose loss, and GL is a given loss function. In this paper, the spread loss (SL) and the cross-entropy loss (CEL) are used and they are given as follows:

$$SL_i = (\max(0, m - (\gamma_t - \gamma_i))^2), \quad SL = \sum_{i \neq t} SL_i, \tag{12}$$

$$CEL = -\sum_i L_i \log \gamma_i, \tag{13}$$

where m is the margin,γ_t is the activation of the target class, γ_i is the activation of the ith class capsule, and L_i is the true probability for that class. Hence, minimizing the loss function defined in Eq. (11) would simultaneously minimize the SL (or CEL) and maximize the inter-class variations.

5. Experimental results

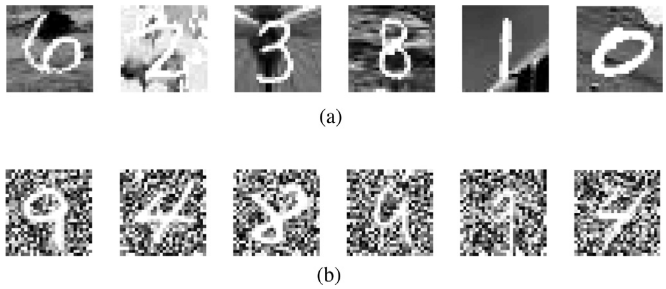
In this section, to compare the performance of CapsNet with WCFGM-R and that one with EM-R, the following experiments have been conducted.

5.1. Implementation details

5.1.1. Datasets

In the experiments, the following datasets are employed.

1. MNIST-back-image (bg-img) [24]: MNIST is composed of 0-–9 handwritten digital images, including 60,000 training images and 10,000 test images, each image is a gray image with 28 × 28 pixels. MNIST-bg-img is a variant of the MNIST dataset. It is made more complicated by adding a background image to the original binary image. This dataset contains 12,000 training


(a)

(b)

Fig. 1. A visualization of some instances of (a) MNIST-bg-img, and (b) MNIST-bg-rnd.

images and 50,000 test images. Fig. 1(a) shows some instances of MNIST-bg-img.

2. MNIST-back-rand (bg-rnd) [24]: MNIST-bg-rnd is also a variant of the MNIST dataset. It is made more complicated by adding a random background to the original binary image. This dataset contains 12,000 training images and 50,000 test images. Fig. 1(b) shows some instances of MNIST-bg-rnd.
3. CIFAR10 [25]: The CIFAR-10 dataset consists of 32×32 colored images in 10 different classes, with 6000 images per class. 50,000 images are used for training and 10,000 images for testing. Fig. 2 shows some samples from CIFAR10 dataset.

5.1.2. Network architecture and settings

The capsule network architecture with WCFGM-R framework used in this paper is given in Fig. 3. The model starts with a 5×5 convolutional layer with 32 channels (i.e., $A = 32$), a stride of 2, and a ReLU activation function. Then, it is followed by 8 (i.e., $B = 8$) primary capsule types with a stride of 1 and a sigmoid activation function. The two following capsule layers are 3×3 convolutional ($K = 3$), each with 16 capsule types (i.e., $C = D = 16$) and strides of 2 and 1, respectively. Finally, the last convolutional capsules layer is connected to the class capsules layer which has one 16-dimensional capsule per class (i.e., E classes). The coordinate addition is implemented (see [3]) by adding the scaled coordinate of the center of the receptive field to the vote matrices.

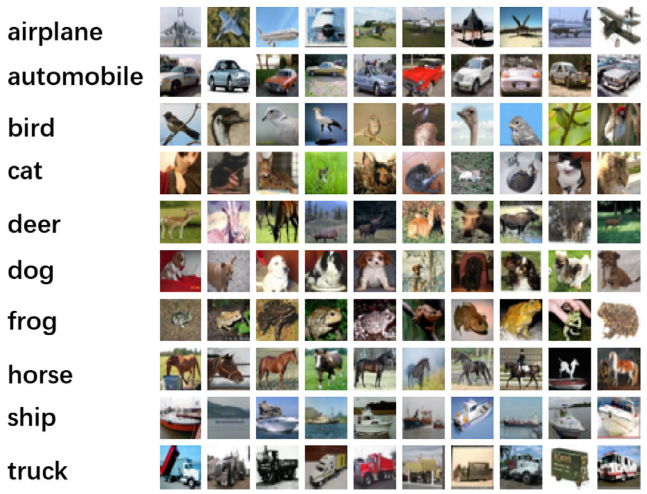


Fig. 2. A visualization of some examples in CIFAR10.

ADAM [26] optimizer and an exponential decay learning scheme with a learning rate 0.001, and a decay rate 0.8 are used. We establish the batch size to 32, the number of routing iterations to 2, the number of epochs to 50, and θ to 10, for all datasets. The parameter ξ for scaling down the pose loss has been set to 0.05 for MNIST-bg-img and MNIST-bg-rnd datasets, and 0.005 for CIFAR10 dataset. The same network architecture and settings are used for both CapsNet with WCFGM-R and CapsNet with EM-R. By following [3], the spread loss is started with a small margin of 0.1 and linearly increased to 0.9 during the training.

5.2. Performance comparison

In this subsection, we employed three datasets: MNIST-bg-img, MNIST-bg-rnd, and CIFAR10 to conduct experiments. We aim to an-

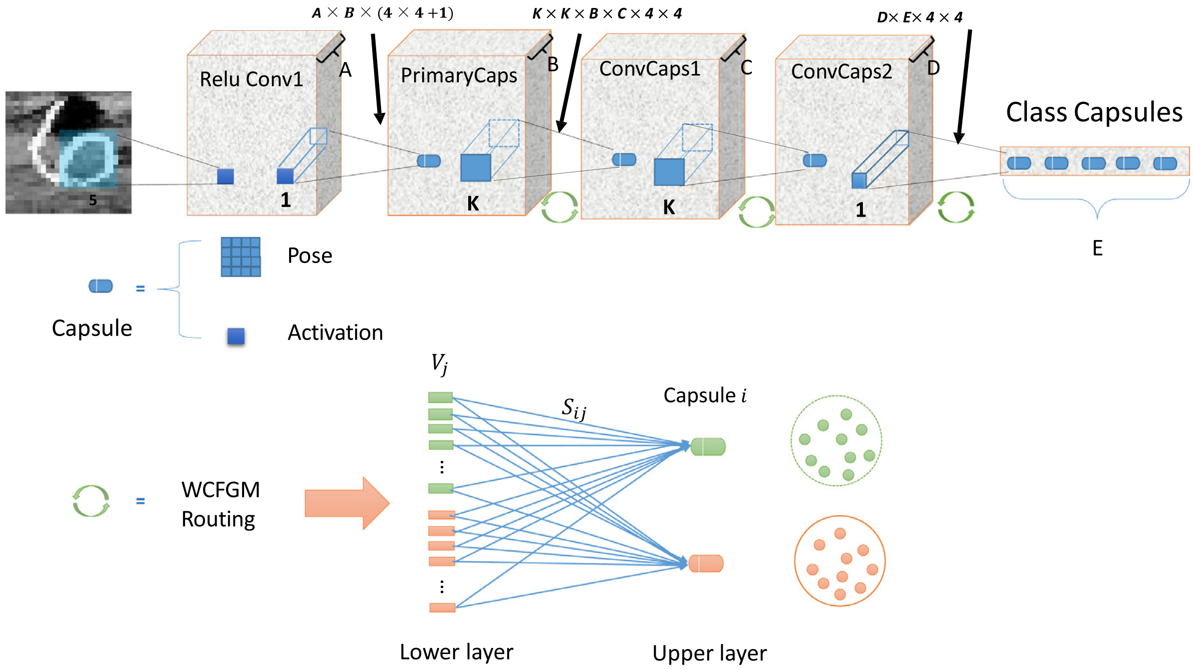
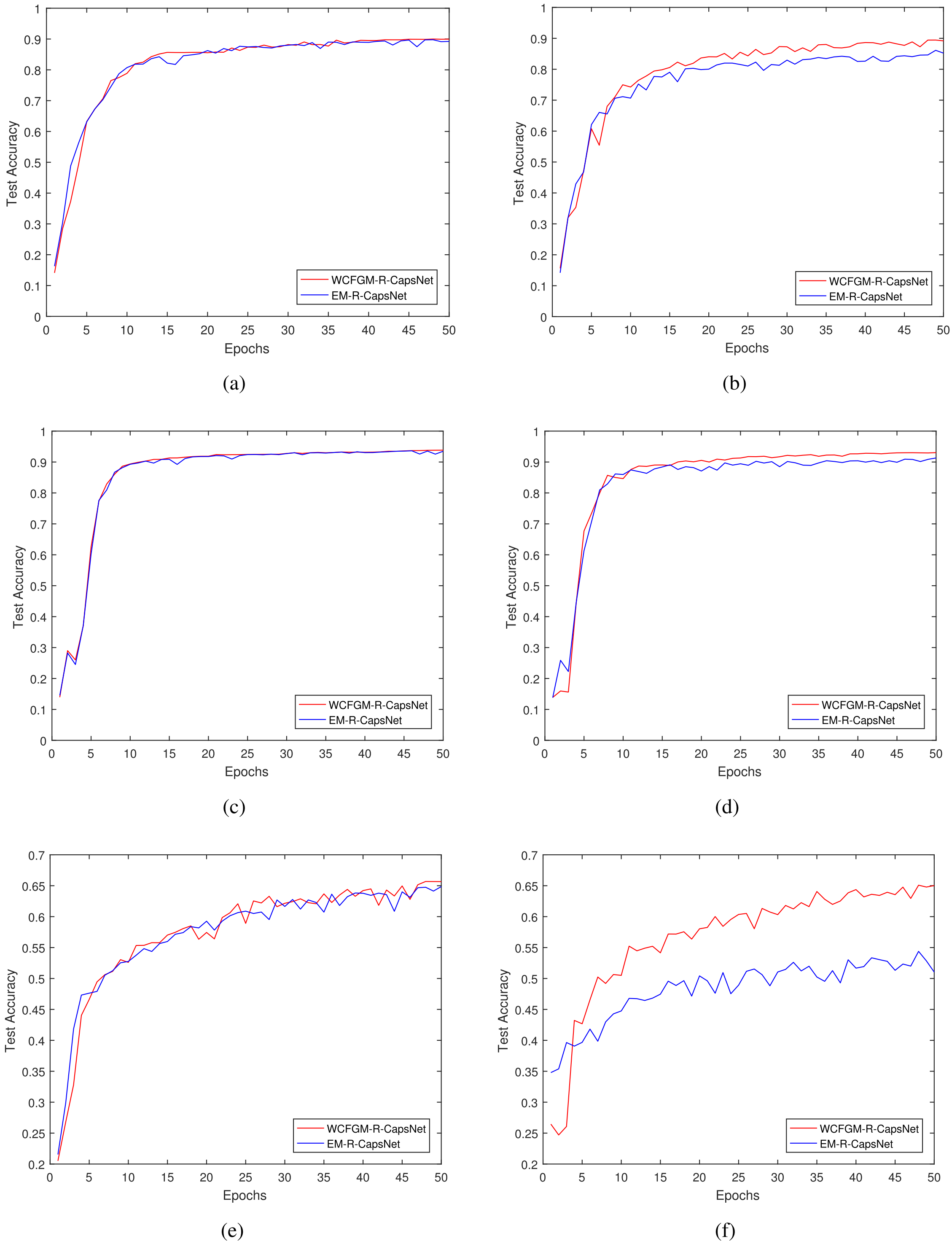


Fig. 3. Capsule Network with WCFGM-R framework.



(a)

(b)

(c)

(d)

(e)

(f)

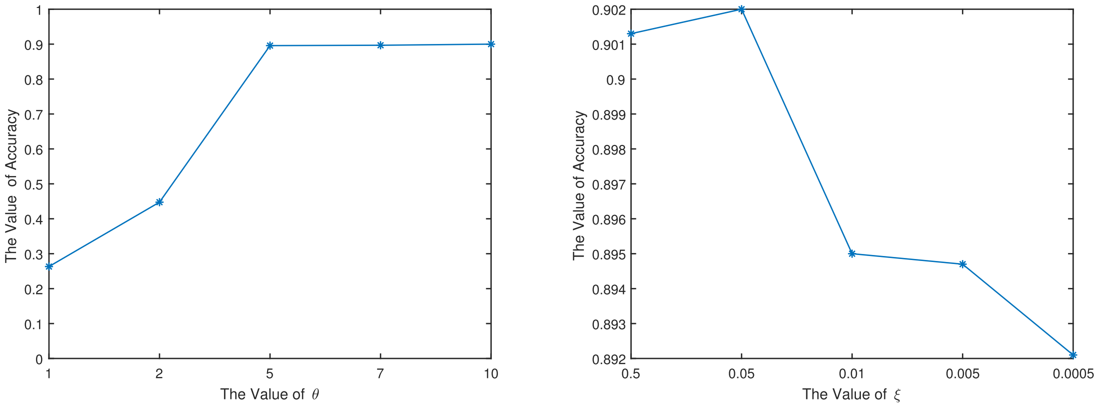
Fig. 4. The change of test accuracies versus epochs for the two models on (a) MNIST-bg-img with spread loss, (b) MNIST-bg-img with cross entropy loss, (c) MNIST-bg-rnd with spread loss, (d) MNIST-bg-rnd with cross entropy loss, (e) CIFAR10 with spread loss, (f) CIFAR10 with cross entropy loss. CapsNet with WCFGM-R has better performance than that one with EM-R.

alyze the performance of WCFGM-R by embedding noise with different images from different datasets.

Fig. 4 shows the test prediction accuracy curves of CapsNet with EM-R and CapsNet with WCFGM-R on the three datasets, with varying loss functions: the spread loss (left side) and the cross entropy loss (right side). The experimental results show that the CapsNet with WCFGM-R outperforms that one with EM-R on the three datasets especially when the cross entropy loss is used, in which it achieves greater improvement as it can be seen from Fig. 4(b), (d) & (f). The performance of CapsNet with EM-R dramatically changed if the cross entropy loss is employed, which implies that the original CapsNet is sensitive to loss functions. In the contrast, CapsNet with WCFGM-R is more robust to loss functions. The results reveal that the proposed routing is robust against noise compared to the EM-R. The proposed routing algorithm has excellent robustness due to the fact that the incorporation of the activities of Childcaps

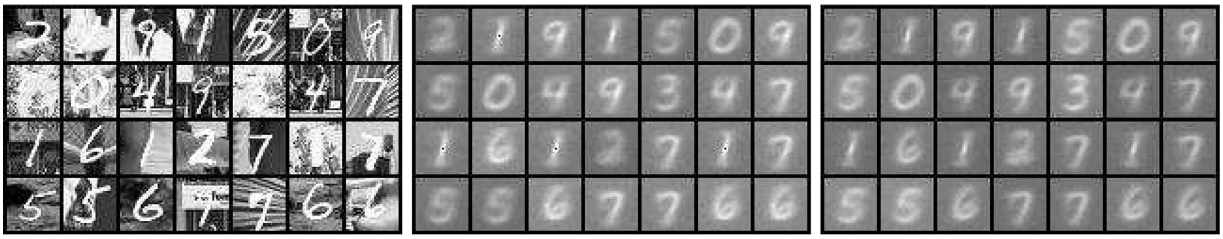
Table 1
Comparison of the best test accuracy (%) across three datasets with different loss functions.

Method	Loss	MNIST-bg-img	MNIST-bg-rnd	CIFAR10
CapsNet + EM-R	Spread	89.77	93.62	64.86
	CrossEnt	86.12	91.31	54.39
CapsNet + WCFGM-R	Spread	90	93.82	65.69
	Spread + L_p	90.2	93.94	65.94
	CrossEnt	89.29	93.03	65.13
	CrossEnt + L_p	89.44	93.69	65.85



(a) (b)

Fig. 5. The analysis of the influence of parameters θ and ξ in WCFGM-R on MNIST-bg-img dataset: (a) the accuracy of WCFGM-R in terms of θ when $\xi = 0$; (b) the accuracy of WCFGM-R in terms of ξ when $\theta = 10$.



(a) (b) (c)

Fig. 6. Visualization of reconstructed images on MNIST-bg-img. (a) Original images, (b) Images reconstructed by CapsNet with EM-R, (c) Images reconstructed by CapsNet with WCFGM-R.

as weights that play the role of precisions in the objective function, reducing the effect of noise on the estimation of poses.

Table 1 reports the main comparative results of the classification task between CapsNet with WCFGM-R and that one with EM-R across three datasets and with varying loss functions. This comparison is based on the best test accuracies. The performance of CapsNet with WCFGM-R is better than the one with EM-R on the three datasets. CapsNet with spread loss shows a better performance than that one with cross entropy loss for both routings. By using the spread loss, the best test accuracy of CapsNet with WCFGM-R is higher than CapsNet with EM-R by 0.23 %, 0.2 % and 0.83% on MNIST-bg-img, MNIST-bg-rnd, and CIFAR10, respectively. Moreover, by using the cross entropy loss, the best test accuracy of CapsNet with WCFGM-R is higher than that one with EM-R by 3.17 %, 1.72 % and 10.74 % on MNIST-bg-img, MNIST-bg-rnd, and CIFAR10, respectively. Furthermore, incorporating the pose loss to the spread loss and the cross entropy loss increases the accuracies by around 0.2% and 0.7%, respectively. The results emphasize that incorporating the pose loss would maximize the inter-class separation. As a result, the classification ability of CapsNet is also enhanced.

Since the performance of WCFGM-R relies on its associated parameters, we carried out some experiments with MNIST-bg-img to investigate their influence and the results are displayed in Fig. 5. From Eq. (1), we can see that θ is a regularization parameter which represents the degree of fuzziness of the proposed model, and from Eq. (11) ξ is a parameter for scaling down the pose loss. Fig. 5(a) plots the accuracy of WCFGM-R as a function of θ when $\xi = 0$ (i.e., the pose loss is not used). In Fig. 5(b), the accuracy of WCFGM-R versus ξ is depicted by setting $\theta = 10$ (one of the optimal parameter). From these results, we can see that when θ is larger than 5, the accuracy is optimal. But the experimental results become worse when θ is smaller than 5. This means that the model needs a much fuzzier partition to achieve better results. While, ξ is a crucial parameter, in which the pose loss should not dominate the original loss during the training. As we can see, the optimal value of ξ is 0.05, meanwhile smaller values will decrease the accuracies.

To compare the reconstruction performance of our method and the CapsNet with EM-R, Fig. 6 exhibits some original images (Fig. 6(a)) and the reconstructed images (Fig. 6(b), (c)). By following [2], we train a multi-layer perceptron to reconstruct MNIST from MNIST-bg-img. By comparison, it is found that the images reconstructed by the two capsule networks are very similar. The backgrounds are homogenized and the writing styles of the digits are also visually close. We also find that the digit 1 generated by CapsNet with EM-R tends to lose the center pixels. It is an interesting phenomenon which is deserved to investigate the reason in the future. In summary, our routing algorithm can improve the accuracy in the presence of noisy backgrounds without sacrificing the reconstruction ability.

6. Conclusion

In this work, a novel weighted capsule fuzzy gaussian model routing (WCFGM-R) and a pose loss function are introduced. The proposed WCFGM-R is employed to prohibit the effect of noise on the classification task. A pose loss function provides the best interclass separation which improves the ability of pattern classification. Experimental analyses show that the proposed CapsNet with WCFGM-R improves the accuracy in the presence of noisy backgrounds without sacrificing the reconstruction ability compared with CapsNet with EM-R. In future work, we plan to learn the input background and other non-discriminant information by further improving CapsNet.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Editor and the reviewers for their useful suggestions which have helped to improve the paper substantially. This work is supported by the National Natural Science Foundation of China under Grant Nos. 61976174, 11671317.

References

[1] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proc. 25th Int. Conf. Neural Inf. Process. Syst, 2012, pp. 1097–1105.
[2] S. Sabour, N. Frosst, G.E. Hinton, Dynamic routing between capsules, in: Proc. 31st Int. Conf. Neural Inf. Process. Syst, 2017, pp. 3856–3866.
[3] G.E. Hinton, S. Sabour, N. Frosst, Matrix capsules with EM routing, in: Proc. Int. Conf. Learn. Representations, 2018.
[4] G.E. Hinton, A. Krizhevsky, S.D. Wang, Transforming autoencoders, in: Proc. Int. Conf. Artif. Neural Netw, 2011, pp. 44–51.
[5] S. Zhang, Q. Zhou, X. Wu, Fast dynamic routing based on weighted kernel density estimation, in: International Symposium on Artificial Intelligence and Robotics, Springer, 2018, pp. 301–309.
[6] C. Xiang, et al., MS-CapsNet: a novel multi-scale capsule network, in: IEEE Signal Processing Letters, 2018, pp. 1850–1854.
[7] F.D.S. Ribeiro, G. Leontidis, S. Kollias, Capsule routing via variational bayes, 2019. arXiv:190511455.
[8] P. Afshar, A. Mohammadi, K.N. Plataniotis, Brain tumor type classification via capsule networks, in: 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 3129–3133.
[9] R. LaLonde, U. Bagci, Capsules for object segmentation, 2018. [Online] Available: arXiv:1804.04241.
[10] A. Jaiswal, W. ABDAlmageed, Y. Wu, P. Natarajan, CapsuleGAN: generative adversarial capsule network, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018. 0–0
[11] Y. Du, X. Zhao, M. He, W. Guo, A novel capsule based hybrid neural network for sentiment classification, IEEE Access 7 (2019). 39321–3928
[12] Z. Zhu, G. Peng, Y. Chen, H. Gao, A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis, Neurocomputing 323 (2019) 62–75.
[13] M. Singh, S. Nagpal, R. Singh, M. Vatsa, Dual directed capsule network for very low resolution image recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 340–349.
[14] H. Yao, et al., Capsule network assisted IoT traffic classification mechanism for smart cities, IEEE Internet of Things Journal, 2019.
[15] J.C. Bezdek., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[16] L. Zhang, et al., Diverse fuzzy c-means for image clustering, Pattern Recognit. Lett. (2018).
[17] P. Kaur, A.K. Soni, A. Gosain, RETRACTED: A robust kernelized intuitionistic fuzzy c-means clustering algorithm in segmentation of noisy medical images, Pattern Recognit. Lett. (2013) 163–175.
[18] O. Kessemen, O. Tezel, E. Özkul, Fuzzy c-means clustering algorithm for directional data (FCM4DD), Expert Syst. Appl. 58 (2016) 76–82.
[19] C. Qiu, et al., A modified interval type-2 fuzzy c-means algorithm with application in MR image segmentation, Pattern Recognit. Lett. 34 (12) (2013) 1329–1338.
[20] M.S. Yang, S.J. Chang-Chien, Y. Nataliani, Unsupervised fuzzy model-based gaussian clustering, Inf. Sci. 481 (2019) 1–23.
[21] Z. Ju, H. Liu, Fuzzy gaussian mixture models, Pattern Recognit. 45 (3) (2012) 1146–1158.
[22] J. Zeng, L. Xie, Z.Q. Liu, Type-2 fuzzy gaussian mixture models, Pattern Recognit. (2018) 3636–3643.
[23] I.D. Gebru, X. Alameda-Pineda, F. Forbes, R. Horaud, EM Algorithms for weighted-data clustering with application to audio-visual scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 38 (12) (2016) 2402–2415.
[24] H. Larochelle, et al., An empirical evaluation of deep architectures on problems with many factors of variation, in: Proceedings of the 24th International Conference on Machine Learning ACM, 2007, pp. 473–480.
[25] A. Krizhevsky, G. Hinton, Learning Multiple Layers of Features from Tiny Images, Tech. rep., 2009.
[26] K. Diederik, B. Jimmy, Adam: a method for stochastic optimization, 2014. arXiv:1412.6980.