IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON TRENDS, PERSPECTIVES AND PROSPECTS OF MACHINE LEARNING APPLIED TO BIOMEDICAL SYSTEMS IN INTERNET OF MEDICAL THINGS

Received December 12, 2018, accepted December 17, 2018, date of publication January 1, 2019,
date of current version January 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2889744

Deep Learning Assessment of Myocardial Infarction From MR Image Sequences

MINGQIANG CHEN[1], LIN FANG[1], QI ZHUANG[2], AND HUAFENG LIU[1]

[1]State Key Laboratory of Modern Optical Instrumentation, Department of Optical Engineering, Zhejiang University, Hangzhou 310027, China
[2]Department of Cardiology, South Campus, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 201112, China

Corresponding author: Huafeng Liu (liuhf@zju.edu.cn)

This work was supported in part by the National Key Technology Research and Development Program of China under Grant 2017YFE0104000, in part by the National Natural Science Foundation of China under Grant 61525106, Grant 61427807, Grant U180920013, Grant 61701436, and Grant 81873908, and in part by the Shenzhen Innovation Fund under Grant JCYJ20170818164343304 and Grant JCYJ20170816172431715.

ABSTRACT The quantitative assessment of the location and size of myocardial infarction has important implications for the diagnosis and treatment of ischemic cardiac diseases. In particular, the tasks of optical flow estimation are of increasing interest in the motion analysis in the field of computer vision. In this paper, we propose a deep learning constrained framework, integrating optical flow features for the classification and localization of myocardial infarction from medical image sequences. The framework is composed of two stages. In the first stage, a stacked denoising autoencoder allows for the extraction of the intensity and motion characteristics from images. Thereafter, a support vector machine model is employed to predict the anomaly scores of each input. Initial experiments are performed with two-dimensional cardiac MRI sequences.

INDEX TERMS Deep learning, support vector machine, myocardial infarction.

I. INTRODUCTION

A major clinical problem is the diagnosis of myocardial infarction heart disease [1]–[3], and the cardiac image analysis has been devoted its tremendous efforts to detecting and isolating the location of ischemic or infarcted myocardia for decades. The segmentation of image sequences typically acquired in 16 to 20 frames consisting of 10 to 16 slices each in the case of magnetic resonance (MR) has long been a starting point in parts of cardiac motion and deformation analysis, but with greater emphasis placed on developing strategies to establish the correspondences of a complete mapping of either the right/left ventricle or entire heart between time frames.

Several multi-frame efforts, involving the stochastic finite element framework [4], state-space approaches [5], and Fourier tracking method [7], have been implemented to make use of the periodic nature of the heart. However, most of these works have focused on a frame-to-frame motion analysis, including the noteworthy studies on mathematically motivated regularization [8], continuum mechanics-based energy minimization [9], deformable superquadrics [10], mesh-free representation and computation framework [27], spatiotemporal B-spline [11], Fisher estimator with smoothness and incompressibility assumptions [12], and continuum biomechanics-based energy minimization [6], [13]. Moreover, to display the status of the heart more clearly, strains have been calculated based on the displacement field estimated by correspondence establishing frameworks.

Despite the success of the above approaches to certain degrees, several issues remain when applied to the clinical environment. Firstly, segmentation-establishing correspondence-calculating strains is a complicated process, which may be a limiting factor. Secondly, these methods are time-consuming and tedious, and require the use of careful modeling techniques. Moreover, the complex spatio-temporal motion of the heart may cause implementation difficulties. Owing to the increase in deep learning in medical images [14], [15], the trend over the past several years is clear in numerous studies, particularly when applied to cardiac segmentation problems [16], [17], which establishes bounds for the similarity between the training and test errors. As the ultimate goal of cardiac motion and deformation analysis is the identification and localization of myocardial infarction through the detection of morphological and kinematic abnormalities, the fundamental question is whether it is possible to predict the status of the heart from a deep learning perspective, instead of a segmentation-tracking-strain point of view.

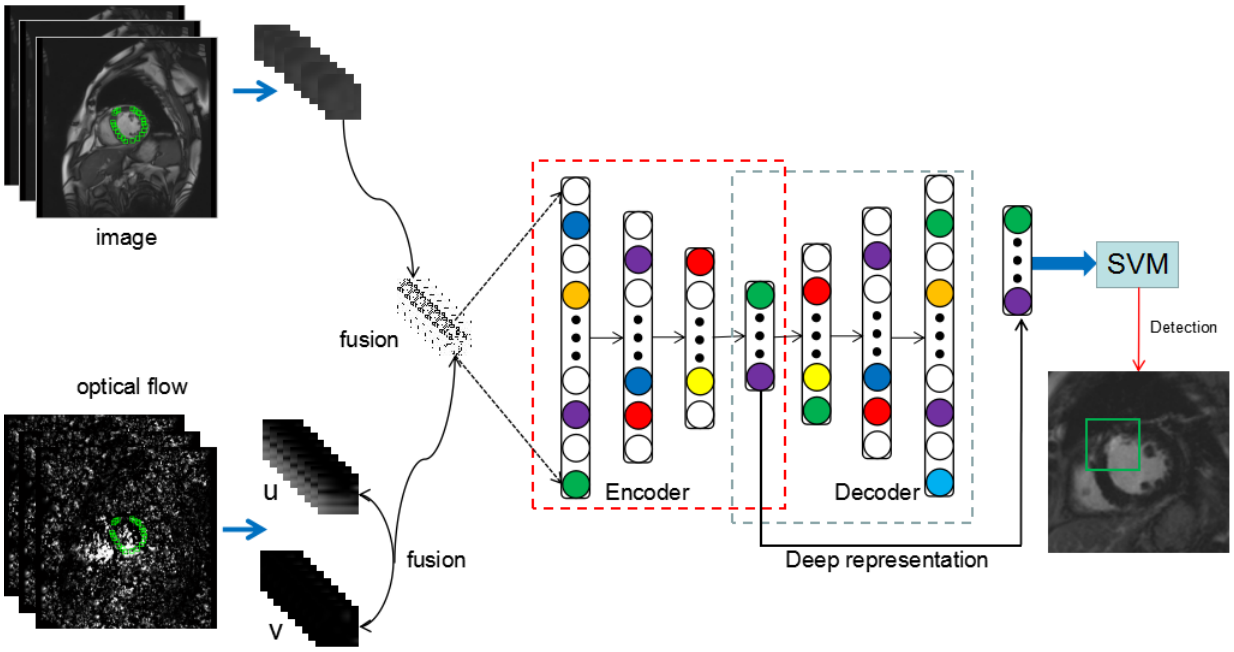


FIGURE 1. Architecture of proposed framework that combining stacked autoencoder and support vector machine for myocardial infarction classification.

In this paper, we propose a framework for the quantitative assessment of the location and size of myocardial infarction simultaneously. The proposed method integrates static and optical flow features [21] into a deep learning framework [22], namely the stack denoising autoencoder (SDAE) [24], [25]. Instead of optimizing the classifier directly, this method minimizes the optimal representation of the feature vector of the classifier error by means of the SDAE prior to training the classifier. Finally, a support vector machine (SVM) is employed for the classification and localization of the region of interest.

We realized that certain work is of relevance to our idea [23], [28]. This study proposes a method based on multiple neural networks from cardiac magnetic resonance imaging (CMRI) to locate myocardial infarction. Firstly, faster region-based convolutional neural network (R-CNN), an object detection network, was utilized to detect the area of the left ventricle. Thereafter, motion feature extraction layers were adopted to extract the local motion information and global motion information of the MRI sequence. Eventually, the final prediction was achieved by the stacked autoencoder. The advantage of this method is that it can detect the myocardial infarction area at the pixel level. However, the overall process was complicated, and the error of each step resulted in subsequent continuous effects. Meanwhile, not only was heavy label calibration required, but the model also contained a huge number of parameters compared with our entire model, which requires more than 138 M parameters. Our proposed method is relatively simple but similar experimental results are obtained. The static and dynamic information of MRI sequence images was explored to learn the deep expression through SAE, and finally complete the prediction by means of the more mature SVM. The parameters required to learn the entire process model were significantly reduced, thus the number of parameters can be neglected.

The remainder of this paper is organized as follows. A brief review of the basic autoencoder and SVM, as well as a detailed description of our proposed framework, are presented in section II. The experimental comparative strategies and results are reported in section III. Conclusions are presented in section IV.

II. METHOD

An overview of the proposed framework is illustrated in Fig. 1. The static image information from the cine CMRI and motion information described by the optical flow were fused and fed into the SDAE. Thereafter, the discriminative feature representation learned by the SDAE was extracted, which served as the input of the SVM to detect the abnormalities of the myocardium.

A. SUBJECTS

The subjects consisted of 51 males and 22 females, aged 60 to 75 years. Cine CMRI sequences and delayed enhancement images were collected from all 73 subjects, as indicated in Fig. 2. The cine CMRI dataset contained a short-axis cardiac sequence of 12 frames, with an acquisition matrix of 256×208 pixels over a cardiac cycle. The delayed enhancement images were obtained approximately 20 min following intravenous injection of 0.2 mmol/kg gadolinium.

We randomly selected 6 subjects from the entire dataset as the validation set, and 13 subjects as the test set. The remaining 54 subjects were used for framework training. Moreover, to ensure a balanced male/female ratio, the ratio

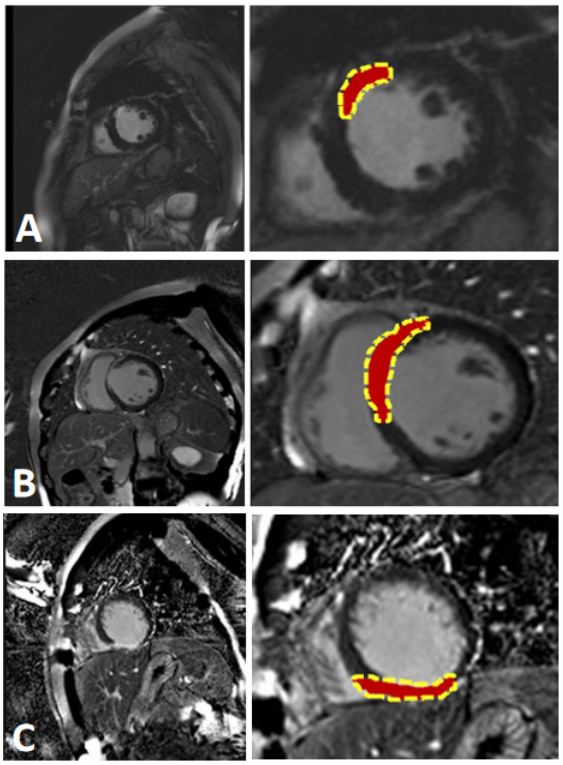


FIGURE 2. Cardiac magnetic resonance imaging (CMRI) of three patients (A, B and C) with heart infarction. All patients are admitted with acute ST-segment elevation myocardial infarction (at least 2 contiguous electrocardiogram leads). The left panels show the delayed enhancement images, while the right panels show the areas of myocardial infraction depicted in red.

of males to females in these three datasets was controlled at around 2:1.

B. STATIC AND MOTION REPRESENTATION

Firstly, fixed-size ($W_w \times H_h$) static patches were obtained by sliding a fixed-sized window through the left ventricular myocardial region. These patches did not need to be strictly separated during the experiment, and there could be some overlap between patches, as indicated in Fig. 3. The intensities of all patches were normalized into the range [0, 1]. Each patch could be expanded as a column vector x_s with $W_w \times H_h \times 1$ elements. To extract the motion information at the same position later, we recorded each extracted position.

The image sequence of each subject was spatially filtered using 11 quadrature pairs of Gabor filters with center frequencies of (f_x, f_y) and a spatial radially symmetric Gaussian σ. At every spatial location $p(p_1, p_2)$, according to the temporal evolution of contours of the constants phase, the point p satisfied $\phi(p, t) = c$, where c was a constant. The temporal phase gradient $\phi_t(p)$ of the CMRI frame for each filter was computed, following which the component velocity V could be derived.

$$V = \frac{-\phi_t(p)}{2\pi(f_x^2 + f_y^2)}(f_x, f_y) \tag{1}$$

Finally, in each space position, the component velocity from the different filter pairs was combined to produce an estimate

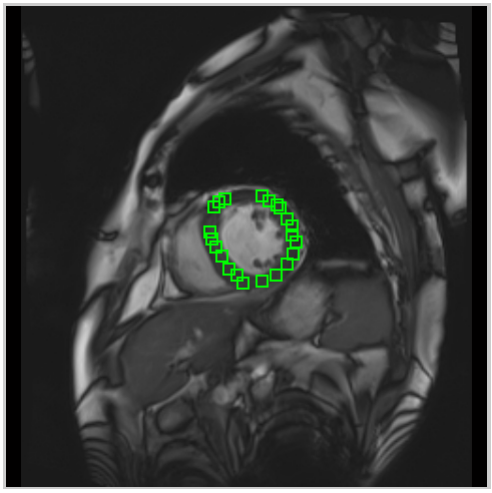


FIGURE 3. One sequence of the short-axis cine cardiac magnetic resonance imaging (CMRI). The patches of left ventricle myocardium are extracted manually enclosed by green boxes. Two types of patches were extracted from the patient cine CMRI: infarcted and non-infarcted.

of the total velocity V at that location. Through the previously recorded extraction location, we extracted the corresponding motion patch for each static patch with windows of the same size, ($W_w \times H_h$). Once again, the velocities of all patches were normalized into the range [0, 1]. The quadratic sum of both the horizontal velocity V_x and vertical velocity V_y served as our motion information $x_m = \sqrt{V_x^2 + V_y^2}$. Therefore, each patch of motion representation could be expanded as a column vector x_m.

C. SDAE FOR HEIGH-LEVEL FEATURE LEARNING

In this study, $x = x_s + x_m$ acted as the input vector for the SDAE. The SDAE is a neural network consisting of multiple layers (see Fig. 4), in which the outputs of one layer are wired to the inputs of its successive layer [18]. Let $X = (x(1), x(2), ..., x(N))^T$ represent all training patches, where $x(k) \in R^d$, N is the number of training patches, d is the number of pixels in each patch, $h^s(k) = (h_1^s(k), h_2^s(k), ..., h_n^s(k))^T$ denotes the learned high-level features at layer s for the $k-th$ patch, and n is the number of hidden units in layer s.

The basic autoencoder was used to reconstruct an original vector $x \in R^d$ from its corrupted version $\tilde{x} \in R^d$. We trained the basic autoencoder by minimizing the objective function as with the squared error. The corrupted inputs were obtained by adding a conditional distribution of Gaussian additive noises into the samples. Thereafter, the basic autoencoder model was optimized by minimizing the average reconstruction error:

$$\begin{aligned}\theta^*, \theta'^* = \arg\min_{\theta,\theta'} \frac{1}{n}\sum_{i=1}^{n} L_r(x^{(i)}, g_{\theta'}(f_\theta(\tilde{x}^i))) \\ + \lambda_1 \sum_{j=1}^{n} KL(\rho||\hat{\rho}) \\ + \lambda_2 ||W||_2^2\end{aligned} \tag{2}$$

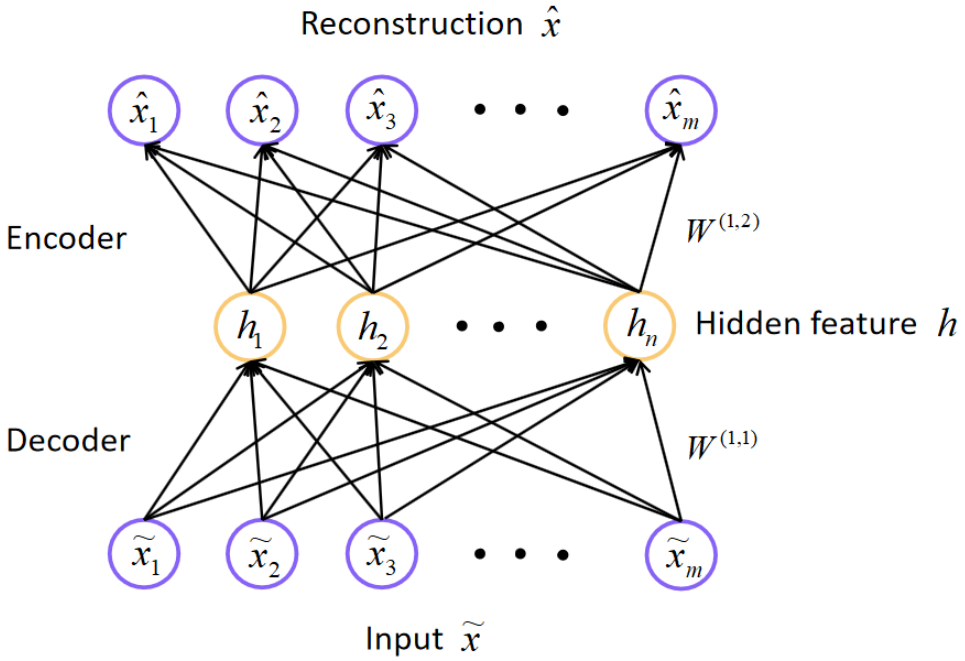


FIGURE 4. Illustration of architecture of basic denoising autoencoder (DAE) with encoder and decoder network. The DAE tries to learn an approximation to the identity function, so as to output $\hat{x}$ that is similar to $\tilde{x}$.

The first term of Eq. (2), namely $\frac{1}{n}\sum_{i=1}^{n} L_r(x^{(i)}, g_{\theta'}(f_\theta(\tilde{x}^i)))$, represents the loss of squared errors, which expresses the difference between the original patch x and input patch $\tilde{x}$. The encoder $f_\theta(\cdot)$ maps the input $\tilde{x}$ to the learned feature h, which can be defined by $h = f_\theta(\tilde{x}) = \sigma(W\tilde{x} + b^{(1,1)})$, where W is a $n \times d$ weight matrix and $b^{(1,1)} \in R^n$ is a bias vector. Here, σ is an activation function defined by the sigmoid logistic function $\sigma(z) = \frac{1}{1+\exp(-z)}$. The decoder $g_{\theta'}(\cdot)$ maps the learned feature back into the original space $\hat{x} = g_{\theta'}(W^T h + b^{(1,2)})$, where W^T is a $d \times n$ weight matrix and $b^{(1,2)} \in R^d$. The weight matrix W^T is the transpose of the weight matrix W, which effectively halves the amount of learned parameters. This is an effective phenomenon for avoiding overfitting in the medical field, in which the amount of data is limited.

The second term of Eq. (2), namely $\lambda_1 \sum_{j=1}^{n} KL(\rho||\hat{\rho})$, represents the Kullback–Leibler (KL) divergence, which is incorporated to ensure the sparseness of the hidden layer. Here, λ_1 represents the loss weight as a percentage of the total loss, n represents the number of units in the hidden layer, and j is the summing over the hidden units in the network. Moreover, $\hat{\rho}_j = \frac{1}{n}\sum_{i=1}^{n} h_j(k)$ represents the average activity of all activation values in a hidden layer, where ρ is the sparsity parameter and is usually a small value close to 0 (for example, $\rho = 0.01$). That is, the aim is to make the average activity of the hidden neurons close to 0.01. To satisfy this condition, the hidden neuron activity must be close to zero. The KL can be defined as $KL(\rho||\hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho)\log \frac{1-\rho}{1-\hat{\rho}_j}$, which can avoid overfitting while increasing the number of hidden layer units, to improve the model expression ability. In this study, two hidden layers were used. It should be noted that, in the first hidden layer, a larger number of units were set and sparsity was employed.

The third term of Eq. (2), namely $\lambda_2||W||_2^2$, is the weight decay term, which can prevent overfitting by reducing the weight magnitude. In practice, this term penalizes large weights and effectively limits the freedom in our model. The regularization parameter λ_2 determines the manner in which to trade off the original cost with the large weight penalization.

In this study, two basic denoising autoencoders were combined into a SDAE, the architecture of which is illustrated in Fig. 5. For the first layer, the network input layer is the corrupted patch $\tilde{x}$, which is represented as a column vector with size 10×10. There were $d = 10 \times 10 = 100$ input units in the input layer. The first hidden layer had $n = 200$ units, while the second hidden layer had $m = 32$ units.

D. CLASSIFICATION AND LOCALIZATION WITH DEEP REPRESENTATION

After training the SDAE containing two hidden layers, the fixed-length ($m \times 1$ vector) features were extracted from the second hidden layer. Hence, the diagnosis procedure can be stated as the problem of classifying N patches in m-dimensional real space (compactly represented by a $N \times m$ matrix). The membership of each vector x_i in the infarct and non-infarct classes was classified by the SVM.

All of the training patches can be expressed as $\{x_i, y_i\}_i^N$, where $i \in \{1, 2, ..., N\}$ is the training patch index. The label y_i was obtained by the delayed enhancement image.

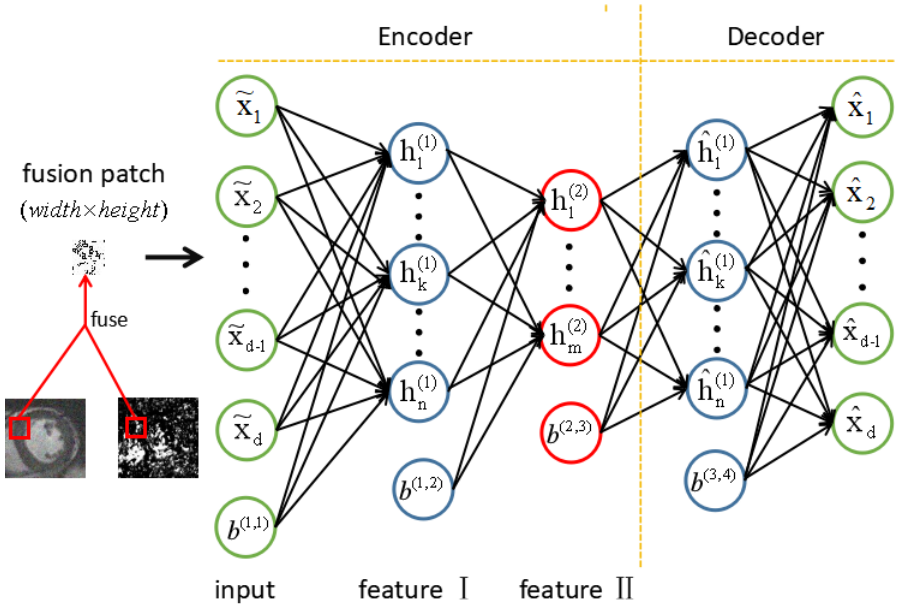


FIGURE 5. Illustration of architecture of stack denoising autoencoder with two hidden layers cascaded by two autoencoders for high level feature learning.

For the two-class classification problem considered in this study, the label of the $i-th$ patch was $y_i \in \{-1, 1\}$, where 1 and -1 refer to infarct and non-infarct patches, respectively. It should be noted that the label information was not used in this model. Compared with the works of other groups, which use supervised learning in medical diagnosis, our approach is an unsupervised learning scheme. Back-propagation was used to compute the loss gradients, which is typically carried out by stochastic gradient descent [19].

This problem can be formulated by a convex quadratic programming problem, as follows (linear SVM [20]):

$$\begin{cases} \min\limits_{w,b} & \frac{1}{2}||w||^2 + C\sum\limits_{i}^{N}\xi_i \\ s.t. & y_i(wx_i + b) \geq 1, \quad i = 1, 2, ...N \\ & \xi_i \geq 0, \quad i = 1, 2, ...N \end{cases} \tag{3}$$

where $C > 0$ is the penalty parameter, generally determined by the application problem. Different selections of C values had varying effects on the experiments, as described in the experimental section.

This problem was transformed into an optimization problem for the m-dimensional vector w and one-dimensional bias b. The infarct and non-infarct vectors correspond to the bounding hyper-planes $wx_i + b = 1$ and $wx_i + b = -1$. If the problem is linearly solvable, the ultimate optimization goal is to obtain as few miscarriages of infarct cases as possible.

However, in reality, most classifications are not linearly solvable problems. Therefore, an urgent need exists for a classifier to complete the situation. The kernel-based SVM model offers such capabilities, and can generate a nonlinear surface by using the $N \times N$ kernel matrix $K(X, X^T)$, where N is the number of training patches, X is the $N \times m$ dataset matrix. The basic concept underlying the kernel function is to map the original training data to a higher-dimensional feature space through appropriate transformation:

$$\varphi(x) : X \rightarrow H \tag{4}$$

where the separation hyper-plane can be determined to maximize the boundary. For the selection of a kernel function to solve practical problems, a commonly used method is the use of prior expert knowledge to preselect the kernel function. However, in the field of medicine, there may not be substantial prior knowledge available for selecting the kernel function. In this study, the Gaussian kernel function was eventually selected. By comparing the performance with other kernel functions, such as the linear or polynomial kernel, it was found that the Gaussian kernel can often optimize the specific problem to a more reasonable solution.

III. EXPERIMENT AND RESULTS

A. TRAINING OF SDAE

The proposed method was mainly implemented in Python on the Pytorch framework. In this study, the number of first hidden layers was set to 200 and that of the second to 32. The architecture of the SDAE is presented in Fig. 5. Prior to training, the neuron weights were initialized in the range of $(\frac{-1}{\sqrt{l}}, \frac{1}{\sqrt{l}})$, where l is the number of inputs to a given neuron. We applied the greedy layer-wise method [26] to train each layer in a sequential manner for the SDAE training. Each new hidden layer was used as a hidden layer supervisory neural network, following which the output layer of the neural network was discarded. Finally, we used the parameters of the hidden layer as pre-training initialization of the new top layer of the deep net to map the output of the previous layer to an

TABLE 1. Confusion matrix, where P represents the infarct class and N represents the non-infarct class.

True class	Predicted class	
	P	N
P	TP	FN
N	FP	TN

improved representation. In this study, the training program consisted of the following three steps: 1) Firstly, we employed the original data, our corrupted fusion patch $\tilde{x}$, to train the first autoencoder to obtain the parameter $W^{(1)}$ and learned feature representation $h^1(\tilde{x})$. It should be pointed out that training the first autoencoder involved sparseness constraints, so the training result was a sparse autoencoder. 2) Based on the first trained autoencoder, taking the corrupted data $\tilde{x}$ as input, the learned feature for each data point was obtained through forward calculation. Thereafter, these learned features were used as the input to the other autoencoder to learn the high-level representation $h^2(\tilde{x})$ by adjusting the weight $W^{(2)}$. The initial learning rate of 10^{-3} was gradually reduced to 10^{-5} during training. 3) Finally, the two layers were combined to form the SDAE, as indicated in Fig. 5, which could map the raw data $\tilde{x}$ to a high-level representation that was fed into the SVM to detect whether or not the patch at a particular location was an infarct.

B. TRAINING OF SVM FOR CALSSIFICATION AND LOCALIZATION

The SVM required that each data instance should be a vector of real numbers. Hence, we first had to convert the deep compact representation of the second hidden layer into numeric data. Simple linear and polynomial kernels were deprecated after multiple trials, because the optimal performance achieved by adjusting the parameters could not provide the desired effects. In this work, we finally selected the Gaussian kernel, which can efficiently learn complex classification functions, and employed powerful regularization principles to avoid overfitting. Our goal was to determine effective C and γ values so that the classifier could accurately predict unknown data (testing data). Examples that had been consistently misclassified in all tests were identified. These examples were provided to a biologist to study again, if it was determined that the original label was incorrect, following which a correction was made, and the process was repeated. In this work, we used a grid search to determine C and γ by cross-validation, which could prevent the overfitting problem.

A confusion matrix, as presented in Table. 1, is often used to describe the performance of a classification model. The sensitivity or true positive rate (TPR) is the conditional probability of correctly identifying infarcted subjects: $TPR = \frac{TP}{TP+FN}$. The false positive rate (FPR) is the conditional probability of a positive test in non-infarcted subjects: $FPR = \frac{FP}{FP+TN}$. Figure 6 illustrates the receiver operating characteristic (ROC) curves and area under the curve (AUC)

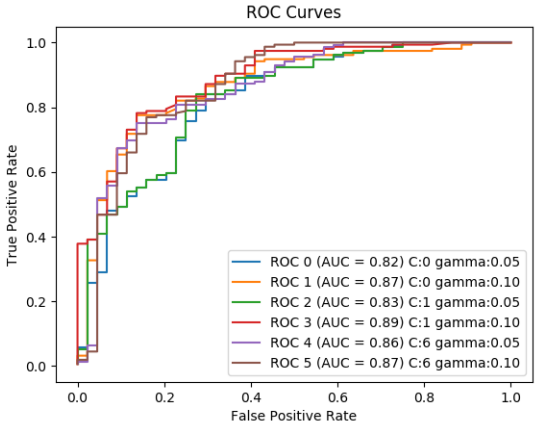


FIGURE 6. SVM performance presented under different C and γ of SVM. The C parameter determines the SVM optimization how much you want to avoid misclassifying each training example. The γ parameter defines how far the influence of a single training example reaches, with low values meaning far and high values meaning close.

obtained from the combination of different parameters. The AUC is equal to the probability of ranking a random positive example over a random negative example. As can be observed from the above experimental results, the optimal results were obtained when $C = 1$ and $\gamma = 0.1$.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

C. WHY USE SVM AFTER THE SDAE

There remains the issue of why SVM training was carried out after the SDAE was trained. It would have been easier to simply apply the final layer of the SDAE, which could provide a two-way softmax regression classifier (SDAE + SM), as illustrated in Fig. 7. We attempted this and found that the performance decreased. The specific implementation process of our approach involved using six-fold cross-validations to train the SDAE model, with a total of 54 subjects in the training set. Firstly, all datasets were divided into six parts. The process was repeated every time to obtain a test set without repetition, the other five were used for the training set of the training model. Each fold consisted of 45 training subjects and 9 validation subjects. For the training dataset in each fold, a SDAE model was trained according to the training method described in a previous section. Thereafter, this trained model could be employed to detect infarcts in the validation set. The accuracy and precision quantitative indicators defined in Eqs. (5) and (6) were employed to evaluate the performance of each model and calculate the average performance as being the final model performance. As a result, the accuracy of the SDAE model was 83.4%, and the precision was 85.4%, as indicated in Table. 2. In line with expectations, the combined SVM effectively improved the overall accuracy and precision of the model. We assume

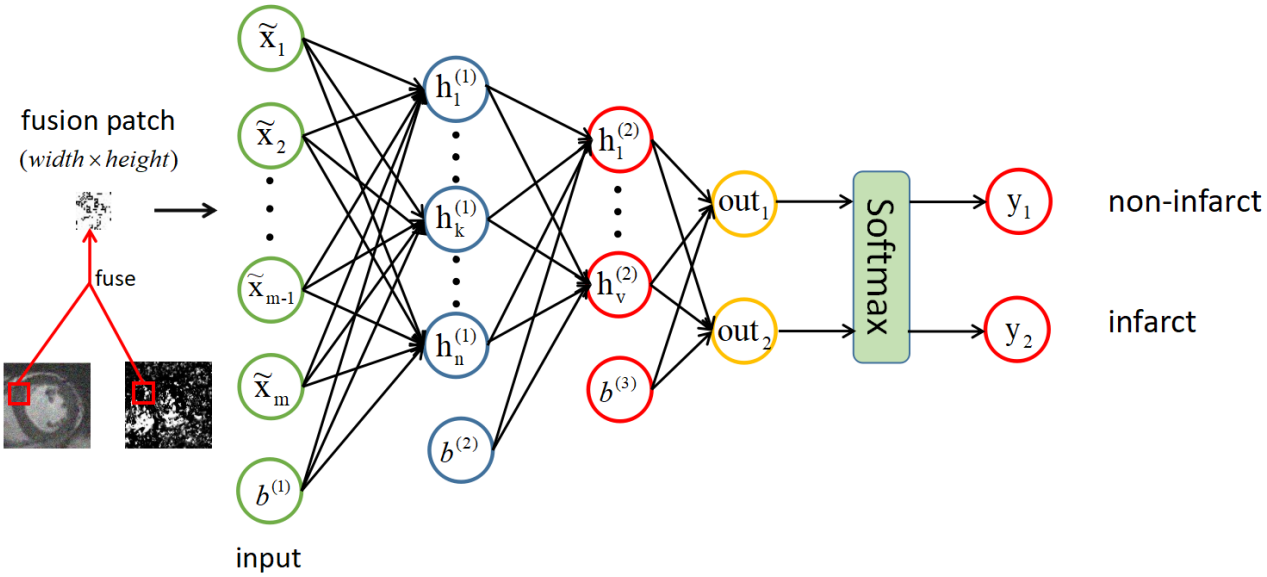


FIGURE 7. Illustration of architecture of stack denoising autoencoder (SDAE) with softmax for myocardial infarction classification.

TABLE 2. The prediction accuracy and precision of different methods in assessment of myocardial infarction. SDAE+SVM operates optimally in comparison with other frameworks.

	SVM	PCA+SVM	SDAE+SVM
Accuracy	0.795	0.848	0.876
Precision	0.822	0.860	0.862

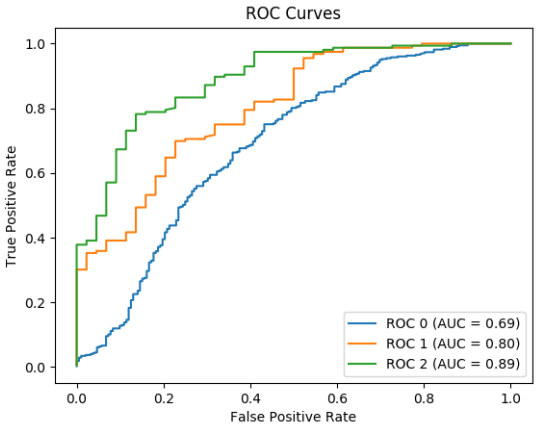


FIGURE 8. The ROC curves represent the results of our proposed framework in three types of data training, namely only static information (blue line), only motion information (yellow line), and fusion of image and motion information (green line).

two possible reasons for this: SVMs offer particularly great power in small sample datasets, and the SDAE may be more adept at data dimension reduction and feature extraction.

D. RESULTS

1) FUSION IS AN ESSENTIAL OPERATION

To verify the fusing effects, an additional experiment was performed to verify the benefits thereof. The experimental results are presented in Fig. 8. The ROC curve indicates the results of our proposed framework for three types of input

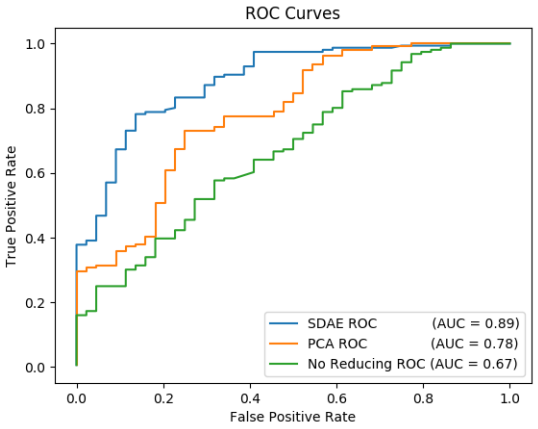


FIGURE 9. The ROC curves demonstrate the effects of different dimensionality reduction methods, in which the green ROC curve is obtained without the use of a dimensionality reduction method, the yellow ROC curve is obtained using principal component analysis to reduce the dimensions, and the blue ROC curve is obtained using the SDAE for dimension reduction.

data, namely only static information, only motion information, and a fusion of these. Evidently, this fusion operation offers considerable benefits and is more conducive to the classification of myocardial infarction. The fusion of cardiac static and dynamic information facilitates a more comprehensive analysis of cardiac infarct conditions.

2) SADE IS AN EFFECTIVE DIMENSION REDUCTION METHOD

To demonstrate the effectiveness of the SDAE used in this study, we conducted a comparative experiment. The experimental results are presented in Fig. 9. Two major observations can be obtained from Fig. 9. Firstly, it is necessary to perform dimensionality reduction of the original data. Secondly, the effect of using the SDAE to for dimension reduction

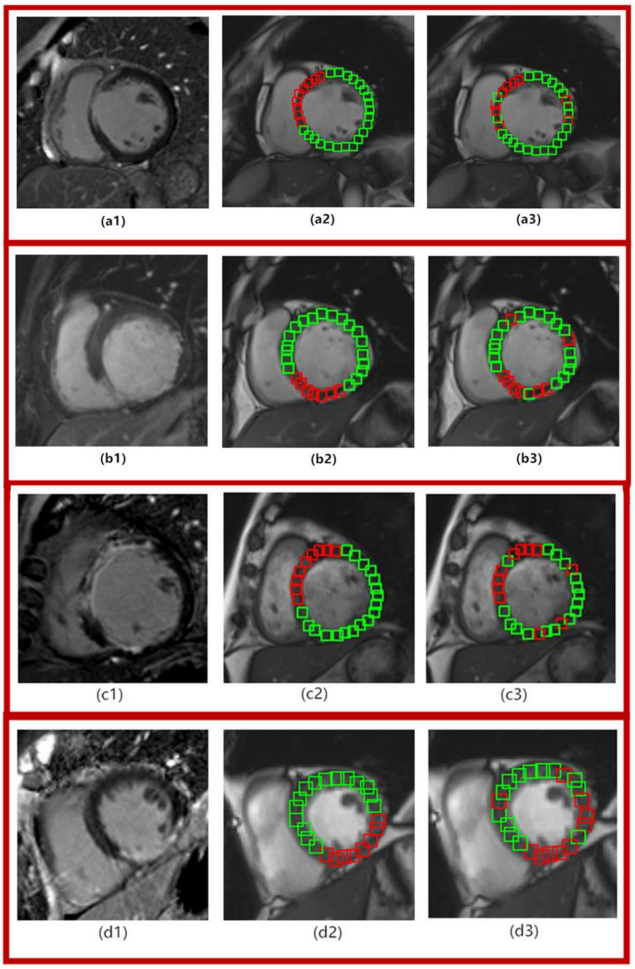


FIGURE 10. Four patients (a, b, c and d) of the infarction detected by our proposed method. The first column (a1, b1, c1 and d1) are the delayed enhancement images from four patients with myocardial infarction. The second column (a2, b2, c2 and d2) represents the ground truth determined by delayed enhancement images. The areas of myocardial infraction enclosed by red boxes. The areas of myocardial non-infraction enclosed by green boxes. The third column (a3, b3, c3 and d3) represents the results of the experimental prediction.

is superior to that of the traditional principal component analysis method. This indicates that SDAE plays a significant role in MI classification. Table. 2 reports the accuracy and precision of the classification performance when using the other frameworks (SVM and PCA + SVM). SDAE + SVM exhibits improved accuracy and precision over all other frameworks, with the highest accuracy of 87.6% and highest precision of 86.2%.

3) TEST OF FRAMEWORK

We set the problem of myocardial infarct detection as a patch-based binary classification and localization problem. We use our framework to detect each fusion patch and reveal its probability of myocardial infarction and specific location. Furthermore, we recorded the specific location of each patch. Thus, we could map to the specific location of the particular patient. Fig. 10 illustrates the results of myocardial infarct detection from four infarcted cases. From the figure, it can be observed that our framework could basically predict the detection of myocardial infarction. This effect could provide cardiologists with significant help in diagnoses.

E. COMPUTATIONAL CONSIDERATION

All of the experiments were carried out on a PC (Intel Core (TM) 3.4 GHz processor with 16 GB of RAM) and a Quadro M5000 NVIDIA graphics processor unit. The proposed method was mainly implemented in the Pytorch framework. The training set included 54 subjects with 3734 patches, and the size of each patch was 10×10. In terms of training time, the SDAE and SVM required 2.35 h. In terms of testing time, testing a patch required 40 s.

IV. CONCLUSION

We have presented a deep learning-based framework for the classification and localization of myocardial infarction from medical images. The framework can capture high-level feature representation in an unsupervised manner. These high-level features enable the classifier to operate very efficiently for predicting infarct locations. Our method eventually performed effectively on a large amount of real patient data.

ACKNOWLEGMENT

(Mingqiang Chen, Lin Fang, and Qi Zhuang contributed equally to this work.)

REFERENCES

[1] J. A. Martin, B. E. Hamilton, S. J. Ventura, and S. Kirmeyer, "Births: final data for 2010," *Nat. Vital Stat. Rep.*, vol. 54, no. 2, pp. 1–116, 2012.

[2] M. Dellborg, "Universal definition of myocardial infarction," *Eur. Heart J.*, vol. 28, no. 20, p. 2525, 2007.

[3] D. Mozaffarian *et al.*, "Executive summary: Heart disease and stroke statistics—2016 update: A report from the american heart association," *Circulation*, vol. 133, no. 4, pp. 447–454, 2016.

[4] P. Shi and H. Liu, "Stochastic finite element framework for simultaneous estimation of cardiac kinematic functions and material parameters," *Med. Image Anal.*, vol. 7, no. 4, pp. 445–464, 2003.

[5] H. Liu and P. Shi, "State-space analysis of cardiac motion with biomechanical constraints," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 901–917, Apr. 2007.

[6] X. Papademetris, A. J. Sinusas, D. P. Dione, and J. S. Duncan, "Estimation of 3D left ventricular deformation from echocardiography," *Med. Image Anal.*, vol. 5, no. 1, pp. 17–28, 2001.

[7] Y. Zhu, M. Drangova, and N. J. Pelc, "Fourier tracking of myocardial motion using cine-PC data," *Magn. Reson. Med.*, vol. 35, no. 4, pp. 471–480, 2015.

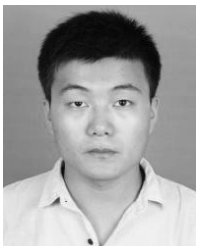
[8] Y. Yu, S. Zhang, K. Li, D. Metaxas, and D. Metaxas, "Deformable models with sparsity," *Med. Image Anal.*, vol. 18, no. 6, pp. 927–937, 2014.

[9] P. Shi, A. J. Sinusas, R. T. Constable, and J. S. Duncan, "Volumetric deformation analysis using mechanics-based data fusion: Applications in cardiac motion recovery," *Int. J. Comput. Vis.*, vol. 35, no. 1, pp. 87–107, 1999.

[10] J. Park, D. Metaxas, and L. Axel, "Analysis of left ventricular wall motion based on volumetric deformable models and MRI-SPAMM," *Med. Image Anal.*, vol. 1, no. 1, pp. 53–71, 1996.

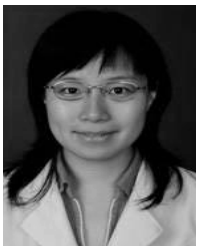
[11] J. Huang, D. Abendschein, V. G. Davila-Roman, and A. A. Amini, "Spatio-temporal tracking of myocardial deformations with a 4-D B-spline model from tagged MRI," *IEEE Trans. Med. Imag.*, vol. 18, no. 10, pp. 957–972, Oct. 1999.

[12] T. S. Denney and J. L. Prince, "Reconstruction of 3-D left ventricular motion from planar tagged cardiac MR images: An estimation theoretic approach," *IEEE Trans. Med. Imag.*, vol. 14, no. 4, pp. 625–635, Dec. 1995.
[13] X. Papademetris, A. J. Sinusas, D. P. Dione, R. T. Constable, and J. S. Duncan, "Estimation of 3-D left ventricular deformation from medical images using biomechanical models," *IEEE Trans. Med. Imag.*, vol. 21, no. 7, pp. 786–799, Jul. 2002.
[14] H. Greenspan, B. V. Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, Mar. 2016.
[15] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[16] S. Queirós *et al.*, "Fast automatic myocardial segmentation in 4D cine CMR datasets," *Med. Image Anal.*, vol. 18, no. 7, pp. 1115–1131, 2014.
[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2014.
[18] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[19] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. Int. Conf. Comput. Statist.*, 2010, pp. 177–186.
[20] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
[21] B. Gao *et al.*, "Estimation of cardiac motion in cine-MRI sequences by correlation transform optical flow of monogenic features distance," *Phys. Med. Biol.*, vol. 61, no. 24, p. 8640, 2016.
[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[23] C. Xu *et al.*, "Direct detection of pixel-level myocardial infarction areas via a deep-learning algorithm," *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2017, pp. 240–249.
[24] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
[25] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.
[26] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 153–160.
[27] H. Zhang *et al.*, "A meshfree representation for cardiac medical image computing," *IEEE J. Transl. Eng. Health Med.*, vol. 6, 2018, Art. no. 1800212.
[28] C. Xu *et al.*, "Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture," *Med. Image Anal.*, vol. 50, pp. 82–94, Dec. 2018.



MINGQIANG CHEN received the B.S. degree from the School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, in 2016. He is currently pursuing the master's degree with the Department of Optical Engineering, Zhejiang University. His research focuses on computer vision and cardiac image processing.

LIN FANG received the B.S. degree from the School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, in 2016. She is currently pursuing the master's degree with the Department of Optical Engineering, Zhejiang University. Her research interest includes cardiac electrophysiological imaging.



QI ZHUANG was born in Shanghai, China, in 1985. She received the B.E. and M.E. degrees from the School of Medicine, Shanghai Jiao Tong University, Shanghai, China, in 2008 and 2010, respectively. In 2010, she joined the Renji Hospital as a Physician. Since 2012, she has been a Cardiologist, and her main areas of research interest are cardiac imaging and interventional cardiology, especially in pulmonary hypertension.

HUAFENG LIU received the B.S. degree in optical engineering, the M.S. degree in measurement techniques and instruments, and the Ph.D. degree in positron emission tomography from the Department of Optical Engineering, Zhejiang University, in 1995, 1998, and 2001, respectively. From 2001 to 2003, he was a Postdoctoral Fellow with The Hong Kong University of Science and Technology, working on statistical filtering and inverse mechanics strategies for cardiac image analysis and PET image reconstruction. He is currently a Full Professor with Zhejiang University. He is also the Director of the ZJU-HAMAMATSU Joint Photonics Laboratory, which was co-founded by Hamamatsu Photonics K. K. and the Department of Optic-Electronic Information Engineering, Zhejiang University, in 1995. Since 1995, his lab has focused on biomedical imaging instrumentation (PET/MRI), image reconstruction (PET/MRI), and medical image analysis.