symmetry

MDPI

Article

End-to-End Multimodal 16-Day Hatching Eggs Classification

Lei Geng [1,2], Zhen Peng [1,2], Zhitao Xiao [1,2,*] and Jiangtao Xi [3]

1 School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin 300387, China; genglei@tjpu.edu.cn (L.G.); 1731095307@stu.tjpu.edu.cn (Z.P.)
2 Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin 300387, China
3 School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong NSW2522, Australia; jiangtao@uow.edu.au
* Correspondence: xiaozhitao@tjpu.edu.cn

Received: 13 May 2019; Accepted: 1 June 2019; Published: 4 June 2019

Abstract: Sixteen-day hatching eggs are divided into fertile eggs, waste eggs, and recovered eggs. Because different categories may have the same characteristics, they are difficult to classify. Few existing algorithms can successfully solve this problem. To this end, we propose an end-to-end deep learning network structure that uses multiple forms of signals. First, we collect the photoplethysmography (PPG) signal of the hatching eggs to obtain heartbeat information and photograph hatching eggs with a camera to obtain blood vessel pictures. Second, we use two different network structures to process the two kinds of signals: Temporal convolutional networks are used to process heartbeat information, and convolutional neural networks (CNNs) are used to process blood vessel pictures. Then, we combine the two feature maps and use the long short-term memory (LSTM) network to model the context and recognize the type of hatching eggs. The system is then trained with our dataset. The experimental results demonstrate that the proposed end-to-end multimodal deep learning network structure is significantly more accurate than using a single modal network. Additionally, the method successfully solves the 16-day hatching egg classification problem.

Keywords: end-to-end; multimodal; hatching eggs; CNN; LSTM

1. Introduction

The flu virus has the characteristics of high infectivity and high transmission speed. In addition, the flu is a disease that seriously threatens the health of human beings. Vaccination is universally regarded as the most important method to prevent influenza and eventually eradicate the disease. Vaccines are made using the influenza virus, which is cultured in living hatching eggs before being inactivated. Immunization can be applied to people after vaccination. A key step in the production of vaccines is to inject the virus into special egg embryos. Some egg embryos may die because of their individual differences. The dead egg embryos must be removed in time, otherwise they may contaminate other egg embryos in the same batch and even cause a serious medical safety accident. Therefore, the efficient detection and separation of necrotic hatching eggs is important for the production of vaccines.

Currently, most manufacturers still use the manual method that detects the integrity of blood vessels in hatching eggs under strong light. This method requires large-scale personnel costs, and the result is easily affected by subjective factors. In addition, because workers perform their duties under high-intensity pressure, there are many shortcomings, such as visual fatigue and low detection efficiency, which make it difficult to meet the high standard requirements of the modern hatching eggs

detection and classification industry. Therefore, companies need a new way to replace work to reduce costs and improve the quality of products.

The detection of hatching eggs is usually divided into four periods: 5 days, 9 days, 14 days, and 16 days. Hatching eggs have different blood-vessel features and heartbeat features during the different periods. As such, there are different classification standards in different periods. In particular, because 16 days is the final period of hatching eggs, detection is more rigorous. The 16-day embryos are divided into three categories: Fertile embryos, recovered embryos, and waste embryos. Fertile embryos were used to extract the vaccine. The recovered embryos were recycled for further processing, and the qualified embryos were selected for vaccine extraction. Waste embryos were treated harmlessly. As illustrated in Figure 1, the fertile eggs have regular heartbeats and strong blood vessels. Recovered eggs have three sets of characteristics. The first set is that hatching eggs have slow heartbeats and strong blood vessels. The second kind of hatching eggs have irregular heartbeats and the blood vessels that begin to constrict. In the third set, egg embryos have no heartbeats and the blood vessels begin to constrict and may even disappear completely. The waste eggs have no heartbeat. All blood vessels disappear completely, and the insides of the eggs begin to rot. Approximately 10% of the recovered eggs have the same blood vessel characteristics as fertile eggs. In addition, 50% of the recovered eggs have the same heartbeat signal characteristics as waste eggs. Because different categories may have the same heartbeat or image features, they are difficult to classify with a single heartbeat or image signal. As such, it is of great significance to improve the technical level of classifying 16-day hatching eggs.

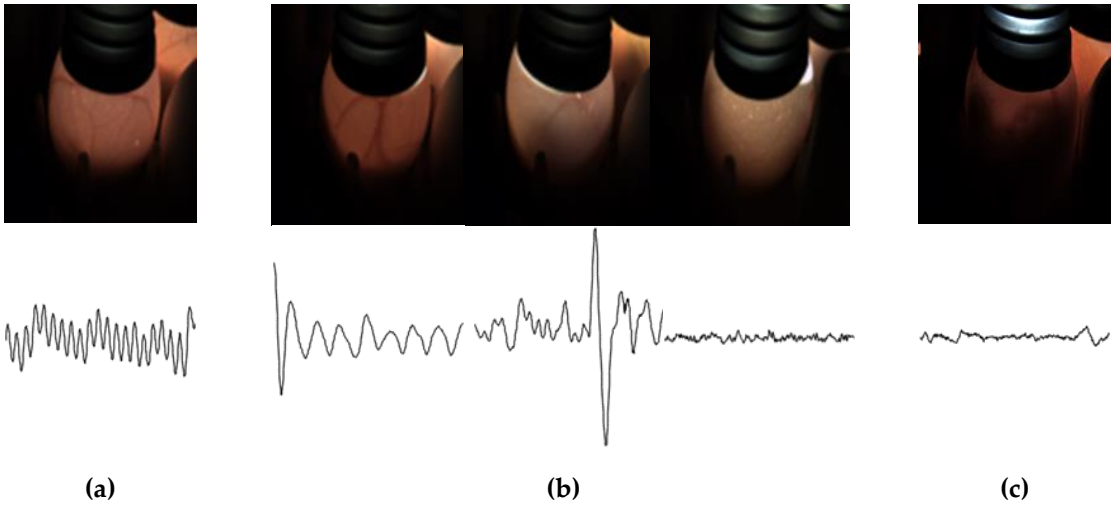


Figure 1. Three categories of 16-day hatching eggs. The first row contains pictures of the hatching eggs. The second row contains photoplethysmography (PPG) signals. (**a**) The fertile egg. (**b**) The recovered egg has three sets of characteristics. (**c**) The waste egg.

In recent years, people have been exploring new methods to classify hatching eggs, such as machine vision technology, hyperspectral imaging technology, and multi-information fusion technology. In 2010, Shan et al. [1] introduced a method to detect the fertility of middle-stage hatching eggs. They used image processing to enhance the picture and obtain the major embryo blood vessels of the hatching egg. Then, they used the weighted fuzzy c-means clustering algorithm to obtain a threshold to detect the fertility. In 2005, Lawrence et al. [2] first used hyperspectral images to detect the development of egg embryos. They designed a hyperspectral imaging system to detect the development of brown- and white-shelled eggs. The detection accuracy was 91% for white-shelled eggs and 83% for brown-shelled eggs. In 2014, Liu et al. [3] proposed a method for detecting infertile eggs using near-infrared hyperspectral imaging. They segmented the region of interest (ROI) of each hyperspectral image and extracted information in the hyperspectral images using the Gabor filter. They used principal component analysis (PCA) to reduce the dimensionality of the spectral transmission characteristics. The final classification accuracy

rates were 78.8% on the first day, 74.1% on the second day, 81.8% on the third day, and 84.1% on the fourth day. In 2014, Xu et al. [4] designed a non-destructive method for detecting the fertility of eggs prior to virus cultivation. Due to the high transmission through the holes in the eggshell, they used a method based on the smallest univalue segment assimilating nucleus to distinguish high-brightness speckle noise pixels in egg images. Additionally, they used the smallest univalve segment assimilating nucleus (SUSAN) principle to detect speckle noise. Then, the blood vessels were restored, and binarized images of the main blood vessels were obtained. By calculating the percentage of the image that the blood vessel area occupies in the ROI image, fertility was evaluated. The final classification accuracy rate was 97.78%.

With the development of deep learning, convolutional neural networks (CNNs) show good performance in solving classification problems. CNNs such as Alexnet [5], GoogLeNet [6], and ResNet [7] are widely used in image classification. In 2018, Geng et al. [8] designed a method for detecting 5-day infertile eggs using a CNN and images of hatching eggs. In 2019, Geng et al. [9] designed a method for detecting 9-day infertile eggs using a CNN and heartbeat signal. Huang [10] designed a CNN architecture in a small images dataset to classify 5- to 7-day embryos, but 5-day to 9-day embryos have no recovered eggs, and there is no overlap between the characteristics of different categories, so using a single heartbeat signal or images of hatching eggs can achieve good results. Therefore, these three CNN methods have achieved good results.

Now, recurrent neural networks (RNNs) are also widely used in the field of processing sequences, such as speech recognition [11]. More and more researchers are combining CNN with RNN to solve new problems. In reference [12], they use CNN-long short-term memory (LSTM) for non-invasive behavior analysis.

The 16-day hatching eggs are divided into three categories. Since different kinds of eggs may have the same heartbeat signal or blood-vessel features, it is not possible to judge the waste eggs and the recovered eggs by the heartbeat signal only, and the embryo image signals cannot be used in isolation to determine the fertile eggs and recovered eggs. As such, we propose an end-to-end, multimodal hatching eggs classification method. Our main contributions are listed as follows:

1. In order to solve the problem of different categories possibly having the same image or heartbeat characteristics, this paper designed a network structure that can simultaneously use the time series heartbeat signals and the egg embryo images.
2. In order to solve time-series classification problems, this paper designed a six layer-deep temporal convolutional network (TCN) architecture that can model the heartbeat signal.
3. We used a pre-training ResNet model to shorten the training time and create a more accurate image classification model.

2. Methods

We proposed a multimodal network structure that can use information in multiple forms. Compared with using a single modal network structure, the recognition accuracy was improved. Our network model is shown in Figure 2. It is divided into a picture processing network called PicNet and a heartbeat signal processing network called HeartNet. The PicNet uses the CNN and the HeartNet uses the TCN architecture. The fusion layer can combine feature maps from the two networks. The fully connected layer produces a distribution of output classes. The details and other variants are given in Table 1, and the structure of ResNet-50 can be found in reference [7].

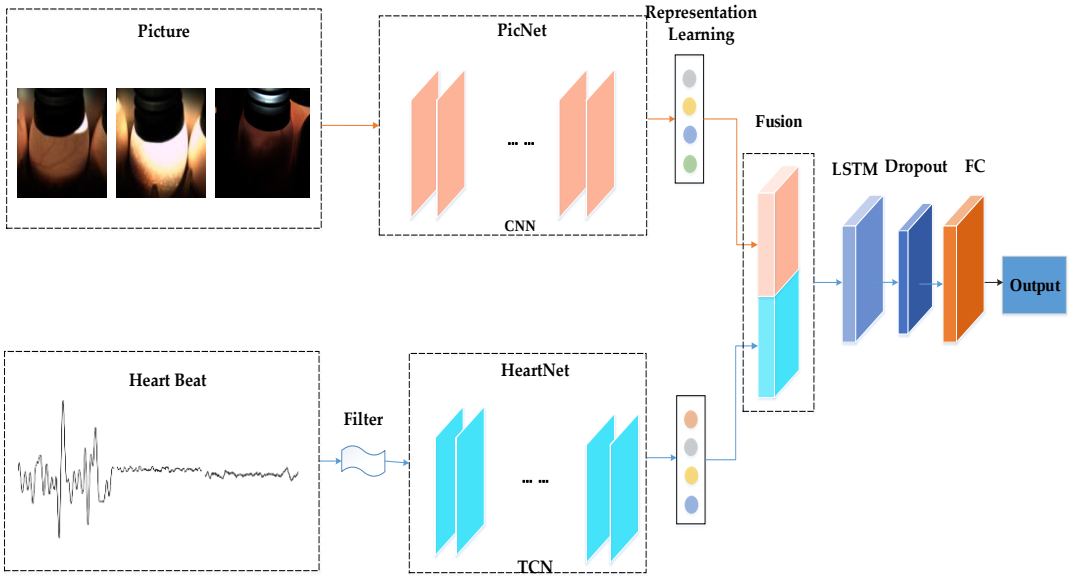


Figure 2. The proposed multimodal network architecture. It is divided into a picture processing network and a heartbeat signal processing network.

Table 1. Related parameters of different layers.

Network	Layer Name	Layer Type	Related Parameters
HeartNet	Conv1	Conv1D	5 kernelsize,1stride,128
	Pool1	Max Pooling	5 kernelsize,1stride
	Conv2	Conv1D	5 kernelsize,1stride,256
	Pool2	Max Pooling	5 kernelsize,1stride
	Conv3	Conv1D	5 kernelsize,1stride,128
	Pool3	Average Pooling	4 kernelsize,1stride
PicNet	ResNet-50 [7]	\	\
Fusion and decision	LSTM	LSTM	350 hidden units
	Dropout	Dropout	dropout-ratio 0.5
	FC	Fully connected	\

2.1. PicNet Design

In this paper, we use a CNN to extract features from hatching egg pictures. We use the existing CNN ResNet-50 as the picture network.

ResNet was proposed in 2015 and won first place in the ImageNet competition classification task. ResNet is simple and practical, so it has been used in areas such as image detection, segmentation, and recognition. The input picture size is 224 × 224 pixels. To reduce the number of parameters, the "bottleneck design" is used in Res-Net-50. Figure 3 shows this architecture. The first 1 × 1 convolution is used to reduce the 256-dimensional channel to 64 dimensions. The second 1 × 1 convolution is used to restore the dimensions. The overall parameters are reduced 16.94 times compared to not using the bottleneck.

Before using ResNet-50, we trained the network structure on the ImageNet 2012 [13] classification dataset. The decay rate is 0.9, and the momentum is 0.1. The batch size is 256. After 100 epochs, we obtained a pre-training model. Using the pre-training model, a more accurate model can be built to shorten the training time.

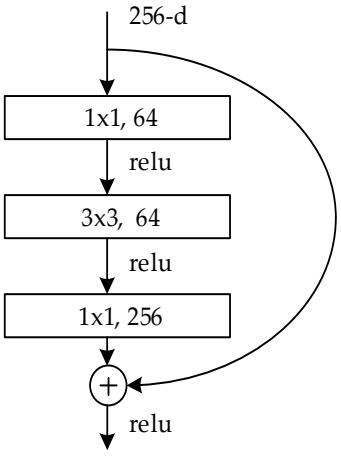


Figure 3. Bottleneck architecture.

2.2. HeartNet Design

The heartbeat signal is a sequence with a duration of 5.6 s after pre-processing, such as filtering and denoising. Its sampling rate is 62.5 Hz. This corresponds to a 350-dimensional vector. Figure 4 depicts the HeartNet architecture.

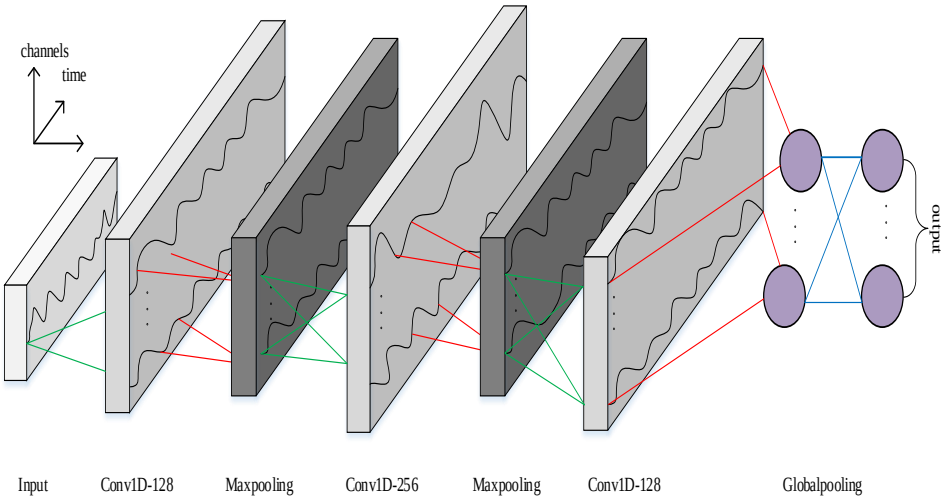


Figure 4. The ARCHITECTURE of HeartNet.

The temporal convolutional networks [14] have proven to be an effective network structure that can solve time-series classification problems [15]. As described in [15], the filters for each layer are parameterized by tensor $W^{(l)} \in \mathbb{R}^{F_l \times d \times F_{l-1}}$ and biases $b^{(l)} \in \mathbb{R}^{F_l}$, where $l \in \{1, \cdots, L\}$ is the layer index, F_l is the length of the input feature of the l layer, and d is the filter duration. For the *l-th* layer of the encoder, the *i-th* component of the (unnormalized) activation $\hat{E}_t^{(l)} \in \mathbb{R}^{F_l}$ is a function of the incoming (normalized) activation matrix $E^{(l-1)} \in \mathbb{R}^{F_{l-1} \times T_{l-1}}$ from the previous layer

$$\hat{E}_{i,t}^{(l)} = f\left(b_i^{(l)} + \sum_{t'}^{d} \left\langle W_{i,t',.}^{(l)}, E_{.,t+d-t'}^{(l-1)} \right\rangle\right) \tag{1}$$

for each time t where $f(\cdot)$ is a rectified linear unit [15].

The heartbeat sequence was fed into 128 filters of size 5 in the first 1D convolutional networks. Afterward, the sequence was downsampled by a max pooling layer size of 5. In the second 1D convolutional network, the sequence was fed into 256 filters of size 5 and then through a pooling layer

of size 5. In the third 1D convolutional network, we used 128 filters of size 5. Finally, the signal was fed into an average pooling [16] layer of size 4.

2.3. Fusion and Decision Layers Design

We connected the last bottleneck architecture of the ResNet-50 network to the average pooling layer and output 2048-dimensional features and then fused them with 448-dimensional features of the heartbeat network's output. The total dimension of the concatenated features was 2496.

The concatenated features were fed into a long short-term memory (LSTM) [16] neural network. LSTM units play a critical role in our network structure. The LSTM unit has three nonlinear gates called the input gate, output gate and forget gate, which can let information pass through and control cell states to be forgotten, updated, or retained. An LSTM maintains a memory vector $\boldsymbol{m}$ and a hidden vector $\boldsymbol{h}$. These vectors control the status update and output at each stage. More concretely, Graves et al. [17] define the computation at time step t as follows,

$$\begin{aligned} g^u &= \sigma(\boldsymbol{W}^u \boldsymbol{h}_{t-1} + \boldsymbol{I}^u x_t) \\ g^f &= \sigma\left(\boldsymbol{W}^f \boldsymbol{h}_{t-1} + \boldsymbol{I}^f x_t\right) \\ g^o &= \sigma(\boldsymbol{W}^o \boldsymbol{h}_{t-1} + \boldsymbol{I}^o x_t) \\ g^c &= \tanh(\boldsymbol{W}^c \boldsymbol{h}_{t-1} + \boldsymbol{I}^c x_t) \\ \boldsymbol{m}_t &= g^f \odot \boldsymbol{m}_{t-1} + g^u \odot g^c \\ \boldsymbol{h}_t &= \tanh(g^o \odot \boldsymbol{m}_t) \end{aligned} \tag{2}$$

where σ is the logistic sigmoid function, $\odot$ represents elementwise multiplication, $\boldsymbol{W}^u$, $\boldsymbol{W}^f$, $\boldsymbol{W}^o$, $\boldsymbol{W}^c$ are recurrent weight matrices, and $\boldsymbol{I}^u$, $\boldsymbol{I}^f$,$\boldsymbol{I}^o$, $\boldsymbol{I}^c$ are projection matrices [18].

We used cross-entropy loss as the loss function,

$$\begin{aligned} loss(x, label) &= -w_{label} \log \frac{e^{X_{label}}}{\sum_{j-1}^{N} e^{X_j}} \\ &= w_{label}\left[-X_{label} + \log \sum_{j-1}^{N} e^{X_j}\right] \end{aligned} \tag{3}$$

where $x \in \mathbb{R}^N$ is the activation value with SoftMax, N is the dimension of X, $label \in [0, C-1]$ is the corresponding label, and $w \in \mathbb{R}^C$ is a vector with dimension C used to represent the weights of labels.

3. Experiments and Results

In this section, we compare our multimodal classification method with a single-mode classification method based on our dataset. Additionally, we evaluate previous methods and the method proposed herein. To evaluate the performance of different methods, we use micro-averaged recall score, micro-averaged precision score and micro-averaged F1 score, which are defined as follows,

$$Accuracy = \frac{\sum_{i=1}^{M} (TP_i)}{N} \tag{4}$$

where TP_i (true positives) is the number of eggs correctly classified into category i;N is the total number of instances; M is the number of categories,

$$Recall_{micro} = \frac{\sum_{i=1}^{M} TP_i}{\sum_{i=1}^{M} (TP_i + FN_i)} \tag{5}$$

$$Precision_{micro} = \frac{\sum_{i=1}^{M} TP_i}{\sum_{i=1}^{M} (TP_i + FP_i)} \tag{6}$$

$$F1_{micro} = \frac{2 \times Recall_{micro} \times Precision_{micro}}{Recall_{micro} + Precision_{micro}} \tag{7}$$

where FP_i(false positives) is the number of eggs that do not belong to class i but are misclassified to class i; TN_i(true negatives) is the number of eggs that do not belong to class i and not classified to class i; FN_i(false negatives) is the number of eggs that belong to class i but were misclassified.

3.1. Dataset

To capture image data, we used a color industrial camera with an 8 mm lens to take pictures of hatching eggs. We used lamps with adjustable brightness to provide a light source and covered the tops of the eggs with a rubber sleeve to prevent light leakage. The size of the original image was 1280 × 960 pixels. We used the photoplethysmography (PPG) technique to acquire the corresponding heartbeat signal. PPG can be used to detect blood volume changes in a microvascular bed of tissue [19]. Because the volume of blood in the blood vessels of egg embryos changes with the heart activity cycle, the light intensity absorbed by the vessels changes synchronously with the beating of the heart. As such, the A/D module can convert light that passes through the tissue into an electrical signal. The signal acquisition equipment is shown in Figure 5. The hatching egg is placed between a laser and a receiving terminal module, which receives light that passes through the egg and converts light into an electrical signal. Finally, the PPG signal is transferred to the microcontroller. The PPG signal is a sequence of 500 data points and the sampling rate is 62.5 Hz.

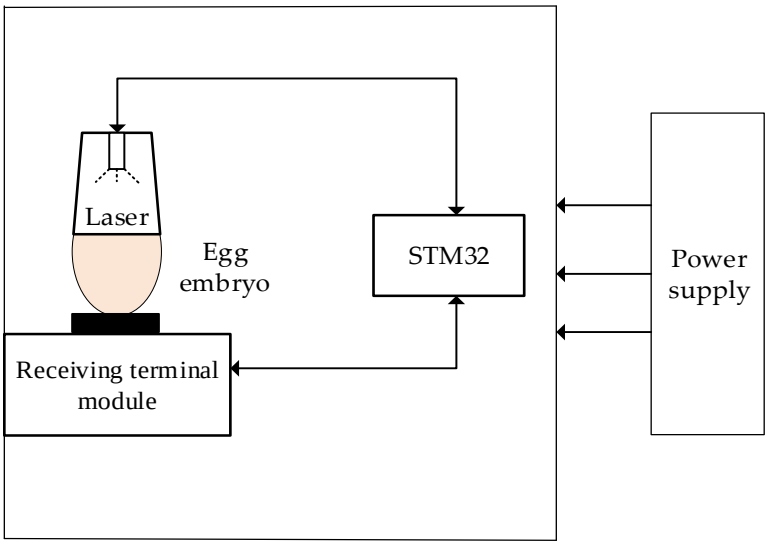


Figure 5. The signal acquisition equipment. The laser source uses a near-infrared source with a wavelength of 808 nm. The receiving terminal module uses the AFE4490 chip, which designed by Texas Instruments for signal denoising and A/D conversion.

Because the background area of the original image was too large, we extracted the region of interest (ROI) to make the embryonic characteristics more obvious. We binarized the image to highlight the outline of the of the egg embryo. For different types of embryos, we used different gray values as thresholds. Then, the maximum contour of the binary image was extracted as the boundary of the ROI region. Finally, all the processed images were scaled to 224 × 224 pixels to fit the required input size of ResNet-50. We designed a second order Butterworth high-pass filter to denoise the heartbeat data

and take the last 350 filtered points as the sampling points. The processed egg embryo pictures and corresponding heartbeat signals are shown in Figure 6.

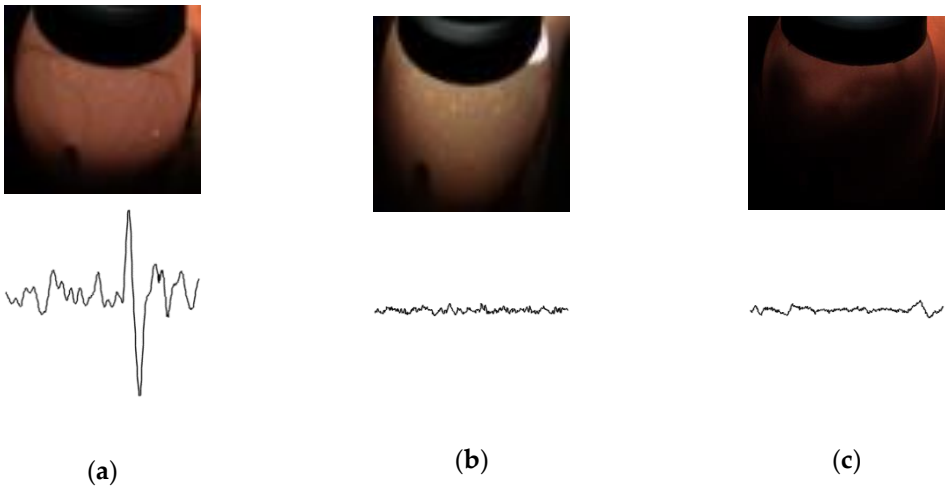


Figure 6. Processed 16-days hatching egg signal. The first row shows pictures of hatching eggs. The blood vessels of the hatching egg are apparent. The second row shows the PPG signal, which reflects heartbeat information. (**a**) The fertile egg. (**b**) The recovered egg. (**c**) The waste egg.

The dataset in this study has a total of 7128 egg embryo images, named the egg picture dataset. Each picture corresponds to a heartbeat signal, and these heartbeat signals are called the heartbeat dataset. In this dataset, there are 2088 samples of fertile eggs, 2160 samples of waste eggs, and 2880 recovered egg samples. The number of embryos in each category is roughly the same, ensuring the balance of the data. All datasets are divided into training sets, validation sets, and testing sets. Table 2 contains more details for each portion of our dataset.

Table 2. The partitioning of the dataset.

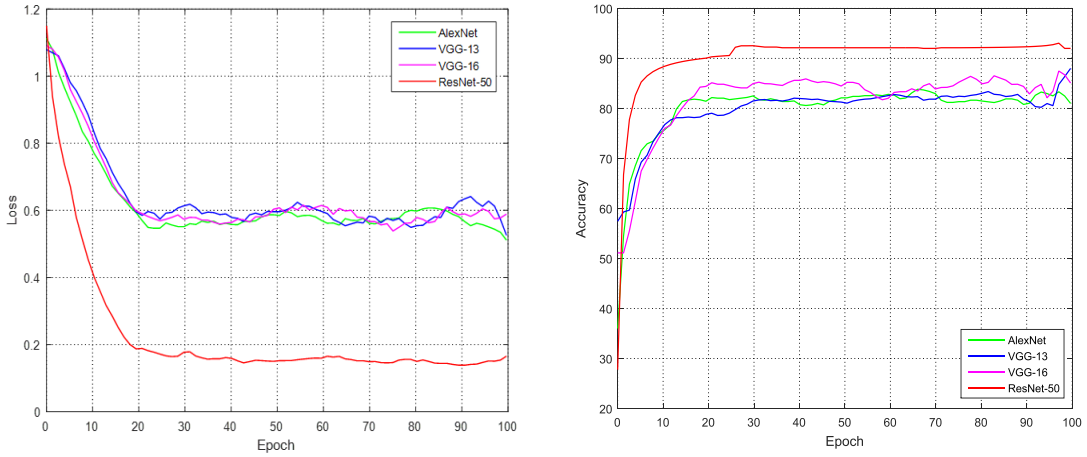
Type	Train	Valid	Test	Total
fertile eggs	1253	418	417	2088
waste eggs	1296	432	432	2160
recovered eggs	1728	576	576	2880
total	4277	1426	1425	7128

3.2. Unimodal Training

We trained PicNet and HeartNet separately on our dataset and compared them to other network structures. The results are as follows.

3.2.1. PicNet Training

We compared existing CNNs on the hatching egg picture dataset. The model was trained for, at most, 100 epochs. The batch size was 32. With eight NVIDIA GTX 1080 Ti GPUs, it took approximately 2 minutes for one epoch. We used the cross-entropy loss function to compute the loss of the PicNet. The varying curves of loss and accuracy are shown in Figure 7. Table 3 contains the accuracies of different CNNs.



(**a**) (**b**)

Figure 7. Loss and accuracy curves of different models on picture dataset. (**a**) Loss curve of different models. (**b**) Accuracy curve of different models.

Table 3. Comparison of performance between different models on picture dataset.

Model	Accuracy
AlexNet	82.56%
VGG-13	85.34%
VGG-16	85.78%
ResNet-50	**90.92%**

Because our egg picture dataset has three types, and approximately 10 percent of the recovered eggs have the same blood vessel characteristics as fertile eggs, the accuracy of using only the picture signal is not high. The best CNN is ResNet-50, which has an accuracy of 90.92%. Based on the results, we used ResNet-50 as the picture network.

3.2.2. HeartNet Training

We studied the effects of different filter sizes k used by each layer of our TCN architecture. We used the cross-entropy loss function to compute the loss of the HeartNet. We performed a series of controlled experiments on the egg heart dataset, the results of which are shown in Table 4. The experimental results show that the TCN model performs best when filter size $k = 5$, so our model's 1D convolution kernel size is 5.

Table 4. The accuracy of different filter size k.

k	Accuracy
3	75.23%
4	77.56%
5	**77.78%**
6	77.68%

We also compared canonical recurrent neural network architectures, such as LSTM and gated recurrent unit (GRU) [20], with the TCN architecture based on our egg heart dataset. To compare all three architectures fairly, the LSTM and GRU architectures have up to six layers so that each model has approximately the same number of parameters, and the optimizers are chosen from adaptive moment estimation (Adam) [21], stochastic gradient descent (SGD) [22], and adaptive gradient algorithm (Adagrad) [23]. The details of the LSTM and GRU architectures are given in Tables 5 and 6.

Table 5. The details of the long short-term memory (LSTM) architecture.

Layer Name	Layer Type	Related Parameters
LSTM1	LSTM	150 hidden units
LSTM2	LSTM	75 hidden units
Dropout	Dropout	dropout-ratio 0.5
FC	Fully connected	

Table 6. The details of the gated recurrent unit (GRU) architecture.

Layer Name	Layer Type	Related Parameters
GRU1	GRU	150 hidden units
GRU2	GRU	75 hidden units
Dropout	Dropout	dropout-ratio 0.5
FC	Fully connected	

All models were trained for, at most, 100 epochs. The batch size was 32. With eight NVIDIA GTX 1080Ti GPUs, it took approximately 1 minute for one epoch. Table 7 contains the accuracies of different networks.

Table 7. Comparison of performance between different models on heart dataset.

Model	Accuracy
LSTM	60.23%
GRU	58.31%
Ours	77.78%

The experimental results show that our TCN architecture performs better than other RNN architectures such as LSTM and GRU. As such, we use our TCN architecture as the HeartNet architecture.

Because our egg heart dataset has three types, and approximately 50% of the recovered eggs have the same heartbeat signal characteristics as fertile eggs, the accuracy of using only the heartbeat signal is low.

3.3. Multimodal Training

We trained the multimodal network and compared it to HeartNet and PicNet. The optimization method we used to train our model is the Adam optimizer. The fixed learning rate is 10^{-4}, the decay rate is 0.9, and the momentum is 0.1. The batch size is 32. With eight NVIDIA GTX 1080 Ti GPUs, it took approximately 3 minutes for one epoch. The loss curve of the training process is shown in Figure 8.

For the training dataset, the loss values of PicNet are slightly lower than those of MultimodalNet. As such, PicNet showed a slightly better performance than MultimodalNet on the training dataset, but for the validation dataset, MultimodalNet had the best performance. Therefore, our proposed method provided the lowest loss among all methods on the validation dataset.

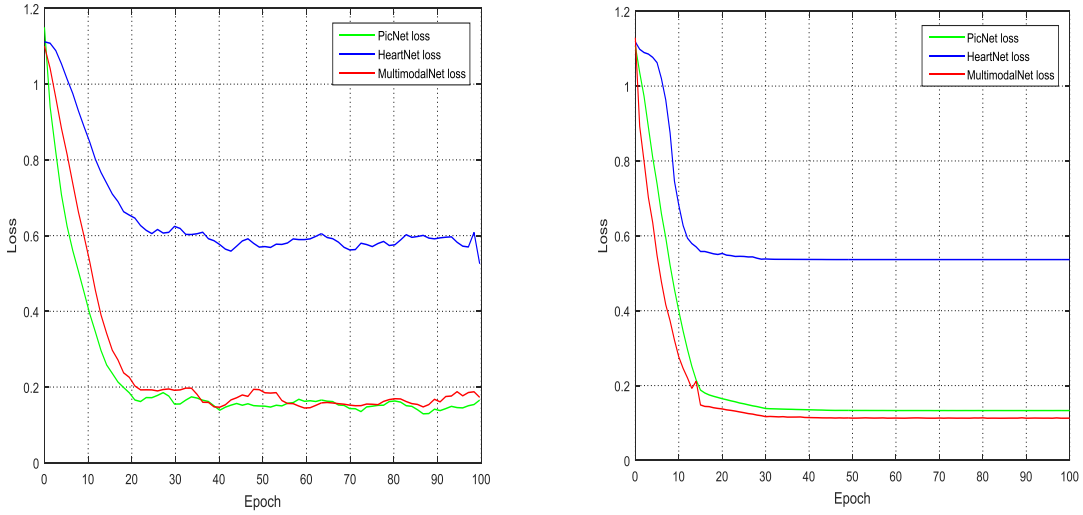


Figure 8. Varying curves of loss. (**a**) Loss curve of training dataset. (**b**) Loss curve of validation dataset.

3.4. Results Evaluation

To verify the feasibility of the network proposed in this paper, we compared the accuracy of single modal networks and the multimodal network. The results are shown in Table 8 and the accuracy curve is shown in Figure 9.

Table 8. Comparison of performance between single modal networks and multimodal network.

Model	Dataset	Signal Type	Accuracy	Recall $_{micro}$	F1 $_{micro}$
PicNet	Egg picture	Picture	90.92%	89.86%	89.99%
HeartNet	Egg heart	Sequence	77.78%	77.82%	77.80%
Multimodal	Mixed	Mixed	98.98%	98.95%	98.90%

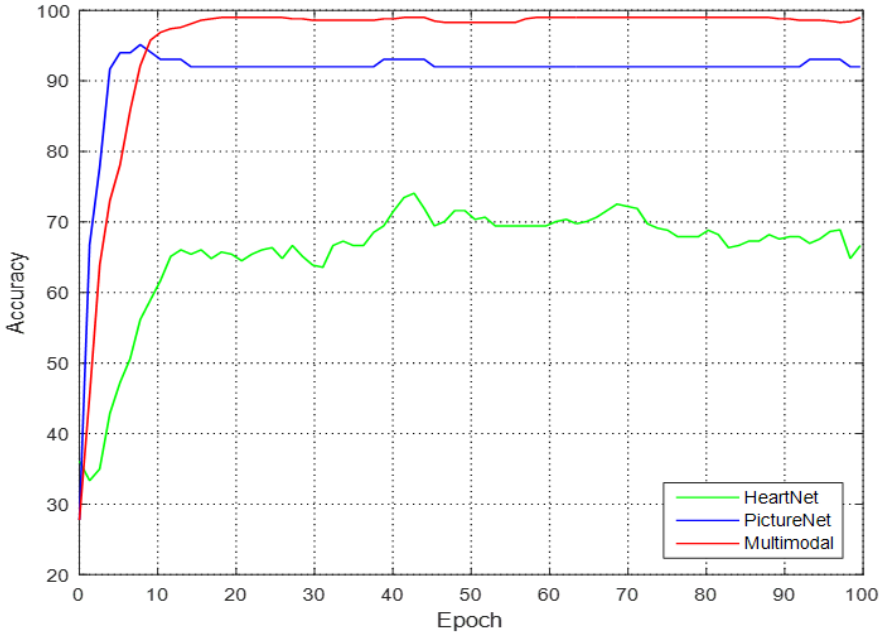


Figure 9. Accuracy curves of single modal networks and the multimodal network.

From Table 8, it is apparent that HeartNet has the lowest accuracy because most of the waste embryos and recovered embryos have no heartbeats, and a small portion of the recovered embryos have heartbeats. Therefore, it is difficult to distinguish the waste embryos and recovered embryos by relying only on the heartbeat signal. Therefore, the accuracy of using HeartNet is very low. There is a small number of embryos with blood vessels but abnormal heartbeats in the recovered embryos, so the use of PicNet led to the inaccurate classification of recovered embryos and fertile embryos. Only by using both signals at the same time can the three types of embryo be correctly classified.

Receiver operating characteristic (ROC) curve can illustrate the diagnostic ability of a classifier system. The larger the area under the ROC curve (AUC), the better the classifier performance, so we also use the ROC chart to illustrate the performance of our model. The ROC chart is shown in Figure 10.

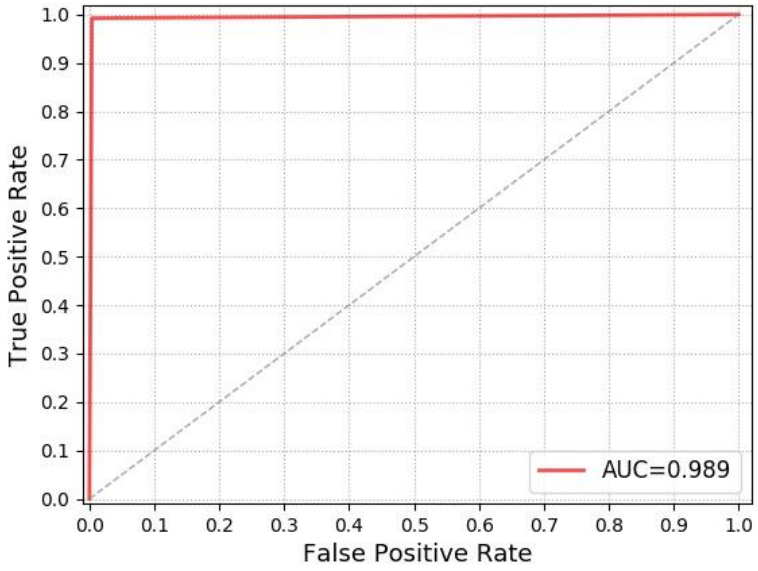


Figure 10. Receiver operating characteristic (ROC) chart of our model.

As can be seen from Figure 10, the AUC indicator of our model is 0.989, which indicates that the performance of our model is outstanding for the 16-day hatching eggs Classification.

We tested our model on the testing sets. The confusion matrix is shown in Figure 11. As can be seen from Figure 11, two of 417 fertile hatching eggs were classified as recovered embryos, five of 432 waste embryos were classified as recovered embryos, and five recovered embryos were also misclassified. A total of 12 embryos were misclassified. The accuracy on the testing sets reached 99.15%.

	Fertile	Waste	Recovered
Fertile	415	0	1
Waste	0	427	4
Recovered	2	5	571

Figure 11. Confusion matrix of our method on the testing sets.

The proposed multimodal network structure inputs two modalities of data at the same time, respectively processes the heartbeat signal and embryo image, and finally fuses them together. The method we proposed can achieve a higher accuracy rate than using a single type of signal.

4. Conclusions

In this paper, we propose an end-to-end multimodal hatching eggs classification method. We designed a deep learning network that includes a picture processing network and a heartbeat

signal processing network. We fed both the heartbeat signals and the egg embryo images into our deep learning network, which overcame the problems that only using heartbeat signals cannot correctly distinguish recovered embryos from waste embryos and that using single-mode embryo images cannot correctly distinguish recovery embryos from fertile embryos. Based on the results of our experiments, the accuracy reached 98.98%. Our method has obvious advantages over other methods that use single modal signals. Additionally, the results show that the proposed method is more suitable for multi-classification of egg embryos.

Our method can replace workers in production and maintain stable operation. This method is not only suitable for hatching eggs classification but also suitable for other aspects. For example, in the fields of face recognition and emotion recognition, video, audio, and other forms of signals can be used for recognition at the same time. In the medical field, we can also combine electrocardiogram and CT images and other signals to improve the accuracy of recognition. Therefore, the method we proposed is very meaningful.

In future work, we will expand our dataset in terms of both the categories of embryos and the amount of experimental data. In addition, we will add more modalities and continue to optimize the network structure to improve its accuracy.

Author Contributions: L.G. and Z.P. wrote the paper; Z.X. and J.X. gave guidance in experiments and data analysis.

Funding: This work was supported by the National Natural Science Foundation of China under grant No.61771340, Tianjin Science and Technology Major Projects and Engineering under grant No.17ZXHLSY00040, No.17ZXSCSY00060 and No.17ZXSCSY00090, the Program for Innovative Research Team in University of Tianjin No.TD13-5034.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shan, B. Fertility Detection of Middle-stage Hatching Egg in Vaccine Production Using Machine Vision. In Proceedings of the 2nd International Workshop on Education Technology and Computer Science, ETCS 2010, Wuhan, China, 6–7 March 2010; pp. 95–98. [CrossRef]
2. Lawrence, K.C.; Smith, D.P.; Windham, W.R.; Heitschmidt, G.W.; Park, B. Egg embryo development detection with hyperspectral imaging. In *Optics for Natural Resources, Agriculture, and Foods*; SPIE: Boston, MA, USA, 2006.
3. Liu, L.; Ngadi, M.O. Detecting Fertility and Early Embryo Development of Chicken Eggs Using Near-Infrared Hyperspectral Imaging. *Food Bioprocess Technol.* **2013**, *6*, 2503–2513. [CrossRef]
4. Xu, Q.; Cui, F. Non-destructive Detection on the Fertility of Injected SPF Eggs in Vaccine Manufacture. In Proceedings of the 26th Chinese Control and Decision Conference, CCDC 2014, Changsha, China, 31 May–2 June 2014; pp. 1574–1579.
5. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
6. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 8–10 June 2015; pp. 1–9.
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
8. Geng, L.; Yan, T.; Xiao, Z.; Xi, J.; Li, Y. Hatching eggs classification based on deep learning. *Multimed. Tools Appl.* **2018**, *77*, 22071–22082. [CrossRef]
9. Geng, L.; Hu, Y.; Xiao, Z.; Xi, J. Fertility Detection of Hatching Eggs Based on a Convolutional Neural Network. *Appl. Sci.* **2019**, *9*, 1408. [CrossRef]
10. Huang, L.; He, A.; Zhai, M.; Wang, Y.; Bai, R.; Nie, X. A Multi-Feature Fusion Based on Transfer Learning for Chicken Embryo Eggs Classification. *Symmetry* **2019**, *11*, 606. [CrossRef]
11. Graves, A.; Mohamed, A.R.; Hinton, G. Speech Recognition with Deep Recurrent Neural Networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.

12. Brattoli, B.; Buchler, U.; Wahl, A.S.; Schwab, M.E.; Ommer, B. LSTM Self-Supervision for Detailed Behavior Analysis. In Proceedings of the 30th Ieee Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 3747–3756.
13. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.H.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
14. Lea, C.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal Convolutional Networks: A Unified Approach to Action Segmentation. In *European Conference on Computer Vision(ECCV)*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 47–54.
15. Wang, Z.; Yan, W.; Oates, T. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks(IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 1578–1585.
16. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
17. Graves, A. Supervised sequence labelling. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 5–13.
18. Karim, F.; Majumdar, S.; Darabi, H.; Chen, S. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* **2018**, *6*, 1662–1669. [CrossRef]
19. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **2007**, *28*, R1–R39. [CrossRef] [PubMed]
20. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.
21. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
22. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
23. Duchi, J.; Hazan, E.; Singer, Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *J. Mach. Learn. Res.* **2011**, *12*, 2121–2159.