A Deep Neural Network-Based Higher Performance Error Prediction Algorithm for Reversible Data Hiding

Bin Ma[1,2], Dongqi Li[1,2], Ruihe Ma[3] and Linna Zhou[4*]
[1]Department of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2]Shandong Provincial Key Laboratory of Computer Network, Jinan, China
[3]Department of Economics, Jilin University, Changchun, China
[4]Department of Cyber Security, Beijing University of Posts and Telecommunications, Peking, China

Keywords: Inception structure; Efficient channel attention; Residual network; Error prediction

Abstract: The traditional error prediction algorithm uses one or more neighboring pixels to linearly predict the target pixel, and it is difficult to use all the information around the original pixel, which affects the prediction performance of the predictor for the target pixel. In this paper, a reversible information hiding error prediction algorithm based on depth neural network is proposed and target images with high prediction accuracy are jointly trained by Inception network, ECA network and residual network. This algorithm extracts the feature images of different receptive fields by using the Inception network, and adds the ECA network after the Inception network to enhance the expression of the important information of the feature images on the high-dimensional channel; at the same time, a residual network structure is added between the Inception network and the ECA network. The feature images of different dimensions of the original image are transmitted to the ECA network, which improves the prediction ability of the target pixel and the convergence speed of the network, and enhances the stability of the network. The algorithm evaluation results using ImageNet database show that the convolutional neural network based on Inception structure has stronger predictive ability compared with the classical error prediction algorithm and other latest research schemes.

1 INTRODUCTION

Since Barton (Barton, 1997) first proposed the reversible data hiding method in the patent, so far, the RDH method has been widely developed. The methods proposed by many researchers can be divided into two categories. The first category is to reduce the embedding distortion by improving the embedding method. Tian (Tian, 2003) proposed a high-capacity DRH method of differential expansion. Ni et al. (Zhicheng Ni et al., 2006) proposed an RDH method based on histogram translation, which exploits the zero or minimum points of the image histogram and slightly modifies the pixel gray value to embed the data into the image. Thodi et al. (Thodi and Rodriguez, 2007) proposed a histogram shifting technique as an alternative to embedding position maps, which improved the distortion performance at low embedding capacities and alleviated the capacity control problem, and also proposed a technique called prediction error Extended Reversible Data Embedding Technique. This new technique exploits the inherent correlation in pixel neighborhoods better than the differential expansion scheme. Ma et al. (Ma and Shi, 2016) proposed a CDM-based reversible data hiding. Weinberger et al. (Weinberger et al., 2000) proposed a low-complexity median edge prediction algorithm (MEDP). The other is to improve the prediction accuracy by designing high-precision predictors. Fallahpour (Fallahpour, 2008) proposed a lossless data hiding method based on the Gradient Adaptive Prediction (GAP) method. Sachnev et al. (Sachnev et al., 2009) proposed an RDH scheme based on diamond interleaved prediction.

Although these prediction methods all utilize the correlation of adjacent pixels in space, they are all local and linear, and the prediction accuracy of some complex images needs to be improved. Therefore, it is valuable to study a global and nonlinear

Ma, B., Li, D., Ma, R. and Zhou, L.
A Deep Neural Network-Based Higher Performance Error Prediction Algorithm for Reversible Data Hiding.
DOI: 10.5220/0012053800003612
In *Proceedings of the 3rd International Symposium on Automation, Information and Computing (ISAIC 2022)*, pages 788-794
ISBN: 978-989-758-622-4; ISSN: 2975-9463

forecasting method to improve forecasting accuracy. With the development of deep learning, convolutional neural networks have been used in image segmentation (Lang et al., 2022; Zhou et al., 2021), super-resolution (Yue et al., 2022), natural language processing (Dong et al., 2015), resolution enhancement (Liang et al., 2022),information hidingand and other fields. extensive development. In the field of information hiding, Luo et al. [30] proposed a CNN-based stereo image RDH method that exploits the correlation of left and right views in stereo images to predict each other. Hu (Hu and Xiang, 2021), et al. proposed a prediction method based on CNN to predict grayscale images, using two steps of feature extraction (for multi-receptive fields) and image prediction (for global optimization) to achieve target pixel prediction, which is similar to the classic one. Compared with the predictor, it has better prediction performance.

The main contributions of this paper are as follows:

(1) A multi-convolution combination model based on the Inception structure is proposed, which extracts image features in a multi-scale parallel manner through convolution kernels of different sizes, making full use of the correlation between adjacent pixels in the image space.

(2) A channel attention model for cross-local channel interaction is proposed. An adaptive one-dimensional convolution kernel is used to determine the coverage of cross-channel interaction. By learning the weights between different channels, the weights of important channels are enhanced. Channel attention cross-channel interaction capability.

(3) An optimization model of adding residual network between Inception modules is proposed. By using skip connections inside the network, the problem of learning degradation of deep networks is solved, the convergence speed of the network is accelerated, and the learning ability of the network is enhanced.

2 RELATED WORK

2.1 Basic structure of GoogLeNet

GoogLeNet was first proposed by Szegedy et al., improving the performance of deep neural networks by increasing the depth and width of the network often results in two problems (Szegedy et al., 2015). First, under the condition that the training set is limited, using a large size usually results in an increased amount of parameters, making the enlarged network prone to overfitting; second, expanding the size of the network results in a substantial increase in the use of computing resources. In order to solve the above two problems, the full connection is transformed into a sparse connection structure.

2.2 Basic structure of ECA-Net

Convolutional neural networks extract features by fusing spatial and channel information together within local receptive fields. Wang et al. proposed an Efficient Channel Attention (ECA) network (Wang et al., 2020).

2.3 Basic structure of Res−Net

Most of the previous models improved the performance of the network by increasing the depth. However, deeper networks often create problems with vanishing gradients and difficulty in training. He et al. proposed a residual learning framework to address network degradation, which is able to simplify previous complex deep networks (He et al., 2016).

3 ERROR PREDICTION ALGORITHM BASED ON DEPTH NEURAL NETWORK

The convolutional neural network of the Inception structure is a sparse structure that can efficiently express features. It can extract image feature information by using convolution kernels of different sizes. Small convolution kernels are used in areas with dense information distribution and strong correlation, and small convolution kernels are used in areas with sparse information distribution. Regions with weak correlation use large convolution kernels. Therefore, this study proposes a convolutional neural network algorithm based on the Inception structure. By optimizing the Inception network structure, the Inception structure can obtain different receptive field ranges and can extract the important features of the original image in space; accurate predictions. The basic model of the convolutional neural network based on the Inception structure consists of three sub-networks: Inception network, ECA-Net, and residual network (as shown in Figure 1).

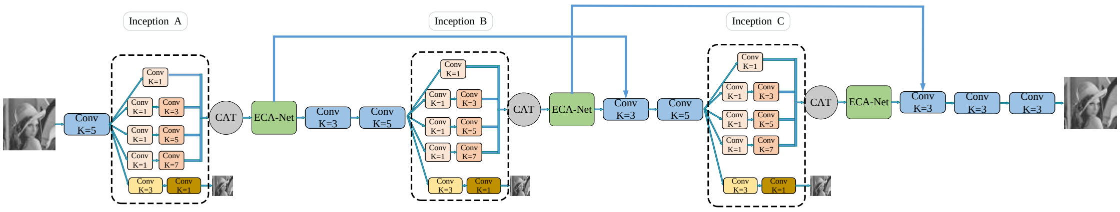

Figure 1: Algorithm architecture based on deep neural network structure.

3.1 Design of Convolutional Neural Network Based on Inception Structure

3.1.1 Inception network structure

In the convolutional neural network based on the Inception structure, the role of Inception is to extract spatial feature information from the original image. Three streamlined Inception module layers are designed in the Inception network in this study. As shown in Figure 2, convolution layers of different sizes are designed inside the Inception module. Each layer of convolution includes image convolution (Conv2d), activation (LeakRelu), and image convolution (Conv2d) three data operations to expand the range of image feature extraction. The convolutional structure of each layer is shown in Figure 3. By adjusting the size of the convolution kernel, different receptive field ranges are obtained, and finally, images with different scale features are spliced in the channel dimension. After the original image is extracted with the initial image features with a convolution kernel size of 5 × 5, the convolution kernel sizes of the Inception structure are respectively 1 × 1, 3 × 3, 5 × 5, and 7 × 7. The first, second, and third, four convolutional layers, and the second, third, and fourth convolutional layers all have 1×1 convolution operations.

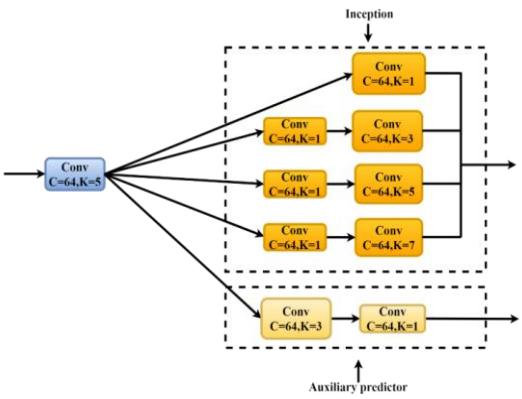

Figure 2: Inception module.

Figure 3: The size of the convolution kernel K in the Inception module is 1, 3, 5, and 7 convolution layer structures

3.1.2 ECA network structure

ECA-Net first uses the global average pooling operation without dimensionality reduction to aggregate features, and uses an adaptive strategy to determine the convolution kernel size K according to the number of channels to achieve local cross-channel interaction, then perform one-dimensional convolution, and finally use the sigmoid activation function to learn channel attention. Table 1 shows the detailed parameter settings of ECA-Net.

Table 1: Detailed parameters of ECA-Net network.

Model	Network layer function	Set
ECA 1	Avg_pool,Conv1d,sigmod	256;[5,1,2]
ECA 2	Avg_pool,Conv1d,sigmod	256;[5,1,2]
ECA 3	Avg_pool,Conv1d,sigmod	256;[5,1,2]

3.1.3 Residual network

In this study, by adding a residual network with a skip layer connection between the Inception network and the ECA network, the image features extracted by the Inception network output layer through the ECA-Net layer are linked to the next Inception network output layer after ECA-Net (the first, two, and three Inception module output features are connected to the second and third ECA-Net respectively) to fuse the shallow feature information with the deep features, as shown in Figure 4, improve the performance of the convolutional neural network based on the Inception structure, and speed up the convergence.

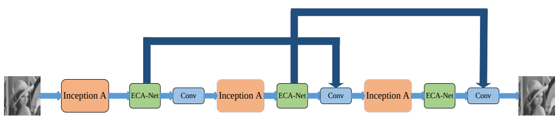

Figure 4: Schematic diagram of the jump-layer connection enhanced network.

3.2 Loss function

The core of the convolutional neural network algorithm based on the Inception structure is to predict the target image based on the Inception network, ECA-Net, and residual network. Different from the standard Inception structure, in the convolutional neural network based on the Inception structure: on the one hand, the range of the receptive field is expanded by deepening the network depth inside the Inception structure, adding auxiliary predictors in parallel with the Inception structure and adding jumps between Inception modules Connections help to better learn image information and speed up the convergence of the network. On the other hand, ECA-Net improves the representation of feature images on important channels by weighting the important channels. In this study, the loss function of the convolutional neural network based on the Inception structure is designed as:

$$loss_{main} = \frac{1}{D}\sum_{i=1}^{D}(\tilde{I}_t - \mathrm{I}_o)^2 \quad (1)$$

$$Loss \ = loss + \lambda \ \ (loss1 + loss2 + loss3) \quad (2)$$

In formula (1), D represents the number of images, $\tilde{I}_o$ is the output image, λ is the target image. In formula (2), loss indicates that the loss function of the backbone network is composed of MSE loss, loss1, 2, and 3 indicate the losses of the first, second, and third auxiliary predictors, respectively. Assigning smaller weights and achieving feature fusion in the final loss can help the network achieve better predictions for the target image. The final loss function is the sum of the loss functions of the backbone and branch networks.

4 EXPERIMENTAL STUDY

4.1 Experimental configuration

In order to verify the performance of the Inception-based convolutional neural network, 3100 images are randomly selected from the ImageNet dataset as experimental data in this experiment. First, the original image is grayscaled, and the grayscale image is resized to 512×512 by bilinear interpolation. The experiment uses the PyTorch framework to implement the convolutional neural network prediction algorithm based on the Inception structure. All experiments are performed on a Dell R740 graphics workstation configured with Intel Gold 5218 CPU, 64GB RAM, and NVIDIA A4000 16G VRAM.

4.2 Evaluation indicators

The prediction performance of the network can be reflected by comparing the similarity between the predicted image and the target image. In the experiment, the mean square error, variance, mean, and prediction error were used as the performance of the prediction network.

Mean Squared Error (MSE) represents the mean squared error between the predicted image and the target image. MSE is calculated as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\ \sum_{j=0}^{n-1}[\tilde{I}(i,j) - I(i,j)]^2 \quad (3)$$

Variance is used to represent the degree of deviation of the forecast error from the mean of the forecast error.

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N} \quad (4)$$

The mean represents the mean of prediction errors.

4.3 Research on the performance of convolutional neural network algorithm based on Inception structure

4.3.1 Effect of different numbers of Inception structures on prediction performance

The bottom layer of the convolutional neural network extracts the local features of the image through convolution. In the local area with weak correlation, a large convolution kernel is used for learning, and in the local area with strong correlation, a small convolution kernel is used for learning. information fusion to obtain a better representation of the image.

In the experiment, different numbers (0, 1, 2, 3) of deep neural network modules were used to verify the performance of the convolutional neural network algorithm based on the Inception structure. The experimental results are shown in Table 3. After adding the Inception module, the similarity between the predicted image and the target image is further improved. The results show that the predicted pixels generated by the convolutional neural network based on Inception structure have strong similarity with the target pixels, and the difference between the predicted image and the target image generated after adding three Inception modules is the smallest.

Table 2: Prediction performance results for different numbers of Inception structures.

	0	1	2	3
Mse	47.6952	47.0069	52.7191	46.3105
Mean	0.1227	0.0759	0.0760	0.07942
Variance	47.6792	46.9971	52.7079	44.5419

4.3.2 Influence of Inception structure with different λ weights on prediction performance

In the experiment, MSE losses with different intensity coefficients (0.1, 0.2, 0.3, 0.5, 0.7, 0.9) were selected to verify the performance of deep neural network structure. The experimental results are shown in Table 3. The experimental results are shown in Table 4. When λ is 0.2, MSE and Variance are the smallest, followed by Mean, and the generated predicted image has the highest similarity with the target image, that is, the convolutional neural network based on Inception structure has good prediction performance.

Table 3: Effect of different λ weights on prediction performance.

	0.1	0.2	0.3	0.5	0.7	0.9
Mse	43.6333	42.8444	46.1708	46.6084	47.8908	46.7753
Mean	0.03268	0.03276	0.0380	0.3331	0.2186	0.1352
Variance	43.6284	42.8339	45.9950	46.4883	47.8151	46.7498

4.3.3 Influence of residual network on prediction performance

MSE is used in experiments to compare and evaluate the image prediction performance of depth convolution neural network based on Incept structure. It can be seen from Figure 5 that for the target image, the predicted image generated by the depth convolution neural network based on the residual network has faster convergence speed and stronger stability. improved.

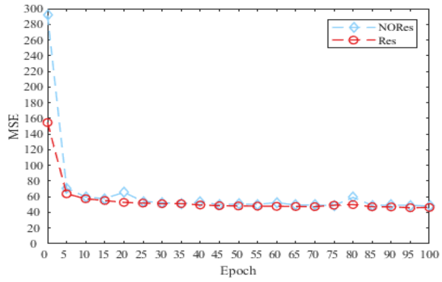

Figure 5: Influence of residual network on prediction performance.

4.3.4 Influence of Eca-Net Network on Prediction Performance

To explore the impact of different numbers of Eca-Net on prediction performance, the experimental results are shown in Table 4. After adding three Eca Net, the prediction accuracy reaches the highest. Mse, Mean and Variance are 42.8444, 0.0328 and 42.8339 respectively.

Table 4: Influence of Different Eca Net Quantities on Prediction Accuracy.

	0	1	2	3
Mse	47.4255	45.0885	47.1813	42.8444
Mean	0.6650	0.2088	0.1893	0.0328
Variance	46.935	45.0369	47.1342	42.8339

4.4 Comparison of prediction performance among different methods

Most of the traditional prediction methods use linear methods to predict image pixels in the local area of target pixels. These prediction methods make full use of the correlation between adjacent pixels in space to predict target pixels. In the experiment, the mean square error, variance, mean value, and prediction error were used to compare and analyze the prediction performance of the algorithm proposed in this paper and other algorithms on the target image. The experimental results are shown in Table 5.

Table 5: Comparison of prediction performance without predictor in Mse, Absolute Mean, and Variance.

	DP	MEDP	GAP	BIP	CNNP	Proposed
Mse	280.5058	214.7554	394.5430	127.2638	47.6952	42.8444
Mean	0.0011	0.4147	1.0949	0.6154	0.1227	0.03276
Variance	280.4984	214.5310	393.1429	126.6566	47.6792	42.8339

5 CONCLUSION

In this paper, reversible data hiding error prediction algorithm based on depth neural network proposed to achieve target pixel prediction. This algorithm makes full use of the correlation between adjacent pixels to generate a prediction image with high precision with the target image through the Inception structure network and ECA network. At the same time, by adding a residual network between the inception networks, the algorithm integrates feature information of different dimensions in the process of network optimization, which enhances the expressive ability of the deep neural network and improves the convergence speed of the network.

ACKNOWLEDGEMENTS

This research presented in this work was supported by the National Natural Science Foundation of China (No: 62272255, 61872203), the National Key Research and Development Program of China (2021YFC3340600), the Shandong Province Natural Science Foundation (ZR2019BF017, ZR2020MF054), Major Scientific and Technological Innovation Projects of Shandong Province (2019JZZY020127, 2019JZZY010132, 2019JZZY010201), Plan of Youth Innovation Team Development of Colleges and Universities in Shandong Province (SD2019-161), Jinan City ''20 universities'' Funding Projects(2020GXRC056 and 2019GXRC031), Jinan City-School Integration Development Strategy Project (JNSX2021030).

REFERENCES

Barton, J.M., 1997. Method and apparatus for embedding authentication information within digital data. United States Patent, 5 646 997.

Dong, L., Wei, F., Zhou, M., Xu, K., 2015. Question Answering over Freebase with Multi-Column Convolutional Neural Networks. Presented at the Natural Language Processing, Beijing, China, pp. 260–269.

Fallahpour, M., 2008. Reversible image data hiding based on gradient adjusted prediction. IEICE Electron. Express 5, 870–876.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 770–778.

Hu, R., Xiang, S., 2021. CNN Prediction Based Reversible Data Hiding. IEEE Signal Process. Lett. 28, 464–468.

Lang, C., Cheng, G., Tu, B., Han, J., 2022. Learning What Not to Segment: A New Perspective on Few-Shot Segmentation.

Liang, J., Zeng, H., Zhang, L., 2022. Details or Artifacts: A Locally Discriminative Learning Approach to Realistic Image Super-Resolution.

Ma, B., Shi, Y.Q., 2016. A Reversible Data Hiding Scheme Based on Code Division Multiplexing. IEEE Trans.Inform.Forensic Secur. 11, 1914–1927.

Sachnev, V., Hyoung Joong Kim, Jeho Nam, Suresh, S., Yun Qing Shi, 2009. Reversible Watermarking Algorithm Using Sorting and Prediction. IEEE Trans. Circuits Syst. Video Technol. 19, 989–999. https://doi.org/10.1109/TCSVT.2009.2020257

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Tian, J., 2003. Reversible data embedding using a difference expansion. IEEE transactions on circuits and systems for video technology 13, 890–896.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks.

Weinberger, M.J., Seroussi, G., Sapiro, G., 2000. The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS. IEEE Trans. on Image Process. 9, 1309–1324.

Yue, Z., Zhao, Q., Xie, J., Zhang, L., Meng, D., Wong, K.-Y.K., 2022. Blind Image Super-resolution with Elaborate Degradation Modeling on Noise and Kernel. Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New Orleans, LA, USA, pp. 2118–2128.

Zhicheng Ni, Yun-Qing Shi, Ansari, N., Wei Su, 2006. Reversible data hiding. IEEE Trans. Circuits Syst. Video Technol. 16, 354–362. https://doi.org/10.1109/TCSVT.2006.869964

Zhou, T., Wang, W., Liu, S., Yang, Y., Van Gool, L., 2021. Differentiable Multi-Granularity Human Representation Learning for Instance-Aware

Human Semantic Parsing. Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, pp. 1622–1631.