CNN Prediction Based Reversible Data Hiding

Runwen Hu and Shijun Xiang

***Abstract*—How to predict images is an important issue in the reversible data hiding (RDH) community. In this letter, we propose a novel CNN-based prediction approach by luminously dividing a grayscale image into two sets and applying one set to predict the other set for data embedding. The proposed CNN predictor is a lightweight and computation-efficient network with the capabilities of multi receptive fields and global optimization. This CNN predictor can be trained quickly and well by using 1000 images randomly selected from ImageNet. Furthermore, we propose a two stages of embedding scheme for this predictor. Experimental results show that the CNN predictor can make full use of more surrounding pixels to promote the prediction performance. Furthermore, in the experimental way we have shown that the CNN predictor with expansion embedding and histogram shifting techniques can provide better embedding performance in comparison with those classical linear predictors.**

***Index Terms*—Convolutional neural network, reversible data hiding, global optimization capability.**

I. Introduction

REVERSIBLE data hiding (RDH) technologies are well known for the ability to recover the original image and the information without any loss [1], [2]. Based on the characteristic mentioned above, RDH methods are widely used in military, medical, and super-resolution processing fields [3].

So far, many approaches have been developed in RDH. One type of approaches focus on finding new embedding ways on how to deal with the prediction errors for reduction of embedding distortion, such as the difference expansion [4], [5], histogram shifting [6], [7], and prediction-error expansion [8]. The other type of approaches focus on how to design advanced predictors to improve prediction accuracy, including difference predictor (DP) [4], median edge direction predictor (MEDP) [8], gradient adaptive predictor (GAP) [9], [10], bilinear interpolation predictor (BIP) [11], [12], and others by using multi-predictors [13] and adaptive strategies [14]–[17].

In the literature, all of the predictors for grayscale images share a weakness, that is to only apply one or a few adjacent pixels as the context for prediction. That is to say, there still exists room for RDH by making full use of more neighboring pixels as the context of a to-be-predicted pixel. To this direction, we are noting that the multi receptive fields [18], [19] and global optimization capabilities of convolutional neural networks (CNNs), which could be served as a predictor for improvement of the prediction accuracy for RDH. In [20], the authors proposed a CNN-based RDH method for stereo images by using the right image and the left image to predict each other. Also, CNN-based data-driven methods have achieved satisfactory performance in global image analysis [21]–[24] and image coding [25], [26].

In this letter, we propose a new and efficient predictor based on CNN. An 8-bits grayscale image is first divided into two subset images, which can be used to predict each other by using the proposed CNN-based predictor (CNNP). The proposed CNNP is a lightweight and computation-efficient network with the capabilities of multi receptive fields and global optimization, which can be trained well by using 1000 images randomly selected from ImageNet [27]. Experimental results show that the CNNP can promote prediction performance due to the use of more surrounding pixels as the context. Experimentally we have shown that the CNNP with classical expansion embedding and histogram shifting techniques can provide better performance by comparing with several classical linear predictors. To the best of our knowledge, this work is the first report in detail on how to use CNN to predict grayscale images for RDH.

The rest of this letter is organized as follows. The proposed RDH scheme is described in detail in Section II, and the experiments comparing with other predictors are reported in Section III. Finally, we conclude our work in Section IV.

II. Proposed Method

A. Pre-Processing Images

At first, the original image I is divided into two subset images: the "Cross" set image I_C and the "Dot" set image I_D. The image partition pattern is shown in Fig. 1. For the "Cross" set image, the pixel values of the positions belong to the "Dot" set are assigned to 0, and so does the "Dot" set image. Such a partition pattern ensures that the two subset images are independent but relevant. After dividing the original image, the two subset images will be used to train and converge the proposed CNNP.

B. Architecture Overview

Based on the correlation of image pixels and the properties of CNN, the "Cross" set image and the "Dot" set image are designed to predict each other in this letter. Fig. 2(a) illustrates how to use the "Cross" set image to predict the "Dot" set image, including two main steps: the feature extraction step (for the use of the multi receptive fields) and the image prediction step (for the use of global optimization). After the feature extraction

Manuscript received December 9, 2020; revised January 22, 2021; accepted February 6, 2021. Date of publication February 12, 2021; date of current version March 10, 2021. This work was supported in part by the NSFC under Grant 61772234. The associate editor coordinating the review of this manuscript, and approving it for publication was Prof. Mylene Q. Farias. *(Corresponding author: Shijun Xiang.)*

The authors are with the College of Information Science and Technology/College of Cyber Security, Jinan University, Guangzhou 510632, China (e-mail: runwen_hu@qq.com; shijun_xiang@qq.com).

Digital Object Identifier 10.1109/LSP.2021.3059202

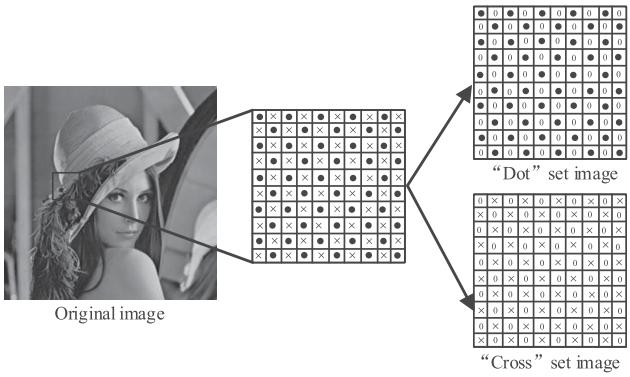
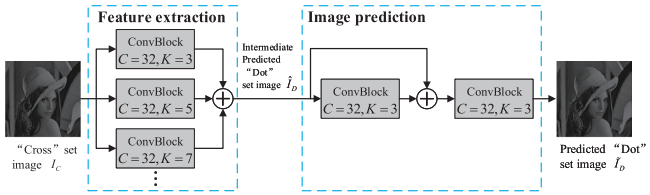


Fig. 1. Illustration to divide an image to "Dot" and "Cross" set images.



(a) Architecture of the proposed CNNP

(b) Structure of the convolution block in (a)

Fig. 2. Overview of the proposed CNNP.

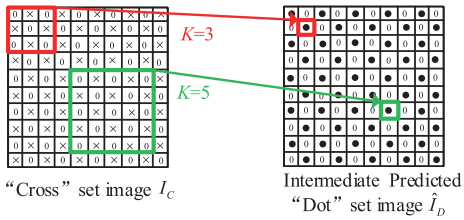
step, we get the intermediate predicted "Dot" set image $\hat{I}_D$, which further is used for the image prediction step. Finally, we get the predicted "Dot" set image $\tilde{I}_D$. The feature extraction step is composed of several convolution blocks arranged in parallel to exploit multi receptive fields. The kernel sizes of these convolution blocks are $K \times K$, where K is an odd number greater than 0 but less than the image size. The structure of the convolution block is shown in Fig. 2(b), where a leaky rectified linear unit (LeakyReLU) activation locates in the middle of two convolution layers.

Examples of the convolution block with $K = 3$ and $K = 5$ for feature extraction are demonstrated in Fig. 3(a). When the "Cross" set image is used to predict the "Dot" set image, the $K \times K$ kernel indicates to use the surrounding $\lfloor K^2/2 \rfloor$ cross set pixels to predict the central dot set pixel, where $\lfloor \cdot \rfloor$ is the floor function. With the increase of K, more surrounding cross set pixels are adopted to predict the central dot set pixel, which can be formulated as,

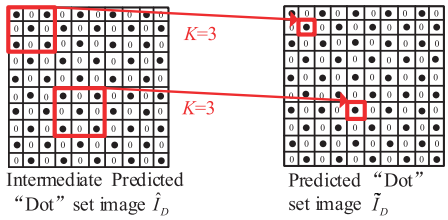
$$\tilde{I}_D(x,y) = \sum_{i,j=1}^{K} I_C(x+i, y+j) \cdot w(i,j) + b, \qquad (1)$$

where $\tilde{I}_D$ is the predicted "Dot" set image, I_C is the "Cross" set image, w is the weight of convolution kernel, and b is the bias.

In the image prediction step in Fig. 2(a), the features extracted from different convolution blocks are added together and fed into two convolution blocks with kernel size of $K = 3$. The pixel values of the intermediate predicted "Dot" set image $\hat{I}_D$ are fine-tuned by five predicted dot set pixels as shown in Fig. 3(b).



(a) When $K = 3$ and $K = 5$ in the feature extraction step



(b) When $K = 3$ in the image prediction step

Fig. 3. Illustration on the use of convolution layers in Fig. 2(a).

By connecting the convolution blocks with different kernel sizes in parallel in the feature extraction step and the fine-tuning convolution blocks in series in the image prediction step, the proposed CNNP can make full use of the global optimization characteristics of CNN to better predict the "Dot" set image.

Considering the lightweight and computational effectiveness of the proposed predictor, the kernel sizes in the feature extraction step are set to 3×3, 5×5, and 7×7, and the number of channels in the proposed CNNP is set to 32. Such a network predictor can be compressed to around 0.7 MB.

C. Training

The proposed CNNP is trained by using ImageNet by randomly selecting 1000 images. In the training, all the images are converted to 8-bits grayscale images with a size of 512×512. Suppose the input is the "Cross" set image I_C, the target is the "Dot" set image I_D and the output of the proposed CNNP is the predicted "Dot" set image $\tilde{I}_D$. Based on back-propagation [28], the Adam optimizer [29] with a batch size of 4 is used to minimize the following loss function:

$$\text{loss} = \frac{1}{P} \sum_{i=1}^{P} \left(\tilde{I}_D - I_D\right)^2 + \lambda \left\|\omega\right\|_2^2, \qquad (2)$$

where P is the number of training data, λ is the weight decay, and ω denotes all the weights in the network. Aiming to effectively suppress over-fitting and accelerate network convergence, λ is set to 10^{-3}. In this letter, the proposed CNNP is trained on an Intel Core i9 CPU (3.6 GHz) with 32 GB RAM and an NVIDIA Titan V GPU.

D. CNNP Based RDH Method

The flowchart of the proposed embedding scheme with the trained CNNP is shown in Fig. 4. The "Cross" set image I_C is firstly used to generate the predicted "Dot" set image $\tilde{I}_D$ by using the proposed CNNP (as illustrated in Fig. 2). Subsequently, I_D and $\tilde{I}_D$ are used to reversibly hide part of the information W_1

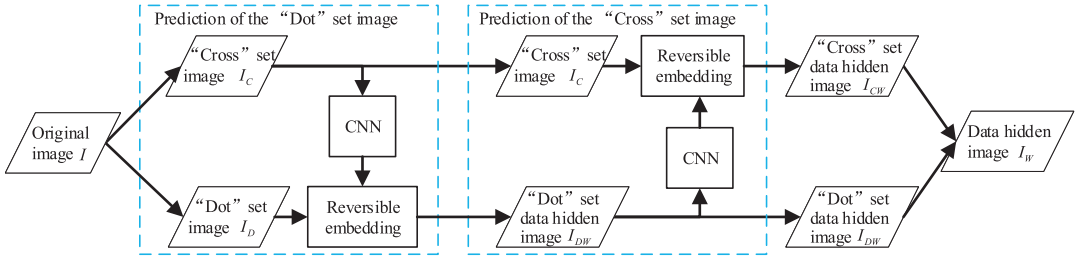


Fig. 4. The proposed reversible data embedding scheme.

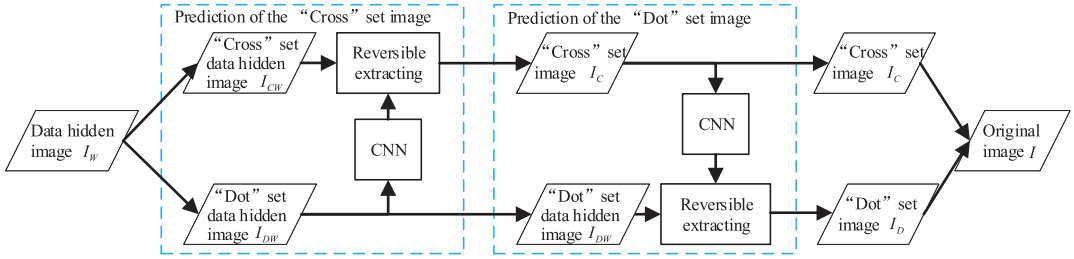


Fig. 5. The proposed reversible data extracting scheme.

and generate the data hidden "Dot" set image I_{DW}. Similarly, I_{DW} is fed into the proposed CNNP to generated the predicted "Cross" set image $\tilde{I}_C$, which is in cooperated with I_C to embed the other part of the information W_2 and generate the data hidden "Cross" set image I_{CW}. After embedding the whole information W ($W = W_1 + W_2$), I_{CW} is added with I_{DW} to generate the data hidden image I_W.

The flowchart of extracting the hidden information and recovering the original image is shown in Fig. 5. The data hidden image I_W is divided into two subset images (i.e., I_{CW} and I_{DW}) with the same partition pattern in Fig. 1. After that, the data hidden "Dot" set image I_{DW} is firstly fed into the proposed CNNP to generate the predicted "Cross" set image $\tilde{I}_C$, which is in cooperated with I_{CW} to extract the information W_2 and recover the original "Cross" set image I_C. The recovered "Cross" set image I_C is fed into the proposed CNNP to generate the predicted "Dot" set image $\tilde{I}_D$, which is used to recover the original "Dot" set image I_D and extract the information W_1. Subsequently, the recovered images I_D and I_C are added together in the spatial domain to recover the original image I. Meanwhile, W_1 and W_2 are combined to recover the information bits W.

III. Experimental Results

The performance of the proposed CNNP is evaluated by comparing it with several classical predictors, including the BIP, MEDP, GAP, and DP. The detailed description of these predictors can be seen in [30].

Aiming to assess the prediction performance of the proposed CNNP, the mean square error (MSE) is adopted since it can reflect the difference between the predicted images and the target images well. For a fair comparison, the predictors mentioned above are combined with the same embedding schemes for RDH, in which the performance of these predictors is measured by computing the peak signal-to-noise ratio (PSNR) values of the watermarked images at the same embedding rates. Independent of the training image set, 100 images are randomly selected from ImageNet for statistical experiments. Without special mention, the results in this section are the average of 100 images.

TABLE I
Average MSE, Absolute Mean and Variance of the Prediction Errors in 100 Images for Five Different Predictors

Predictor	CNNP	BIP	MEDP	GAP	DP
MSE	99.4	154.8	234.2	231.9	230.8
Mean	4.77	6.25	7.37	9.86	5.13
Variance	66.9	100.5	161.3	167.6	196.6

A. Prediction Accuracy

Table I lists the MSE, the absolute mean, and the variance of the prediction errors for the proposed CNNP and four classical predictors. The MSE of the CNNP is 99.4, which is lower than BIP (154.8), MEDP (234.2), GAP (231.9), and DP (230.8). Besides, the absolute mean and the variance in the proposed CNNP are 4.77 and 66.9, respectively, which are more concentrated and lower than BIP (6.25 and 100.5), MEDP (7.37 and 161.3), GAP (9.86 and 167.6), and DP (5.13 and 196.6). From Table I, we can conclude that the prediction accuracy of the proposed CNNP is better than several classical predictors. The basic reason is that the proposed predictor can exploit multi receptive field prediction and global optimization capabilities of CNN by optimizing the network parameters with the loss function. Furthermore, we have plotted the histograms of the prediction error for these five different predictors on the image *Lena*, as shown in Fig. 6, in which the abscissa range is set to $[-30, 30]$ for better display. We can see that the histogram of the prediction errors generated from the proposed predictor is higher and sharper than the histograms of the other four predictors, and the number of the prediction error values in the region $[-3, 3]$ of the proposed CNNP is larger than that of other predictors. From these experimental results, we can

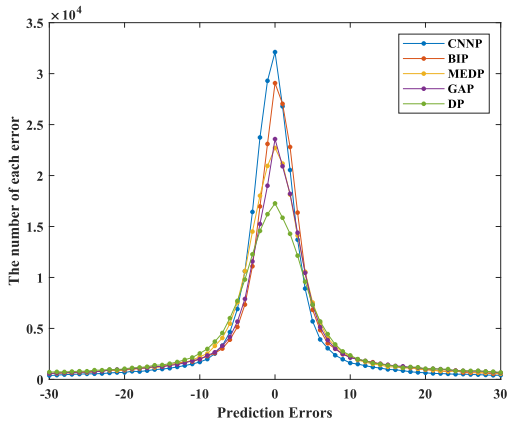


Fig. 6. Histograms of the image *Lena* under five different predictors.

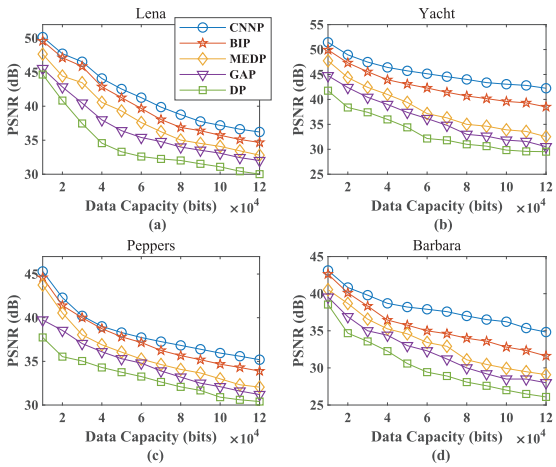


Fig. 7. Comparison of five RDH methods by using four benchmark images.

conclude that the proposed CNNP has achieved better prediction performance in comparison with several classical predictors.

B. Embedding Performance

There are two main embedding techniques for RDH: expansion embedding [8] and histogram shifting [11]. For a better comparison, we first adopted the classical error expansion embedding technique for these five predictors (DP, MEDP, GAP, BIP, and the proposed CNNP). Fig. 7(a)–(d) show the PSNR values of four benchmark images (*Lena*, *Yacht*, *Peppers* and *Barbara*) with embedding capacities from 10 000 to 120 000 bits. From this figure, we can see that the PSNR values of the RDH method with the CNNP are larger. Statistically, we have tested 100 different images and computed their average PSNR values for different embedding capacities in Table II. When the data to be hidden is 10 000 bits, the average PSNR value in the proposed CNNP based RDH method is 47.9 dB, which is higher than BIP (46.4 dB), MEDP (44.2 dB), GAP (41.7 dB), and DP (41.3 dB) based RDH methods. As the data to be hidden increases from 10 000 to 120 000 bits, the average PSNR values of the proposed RDH method are still higher.

To better evaluate the performance of the proposed CNNP, the histogram shifting technique is combined with the proposed

TABLE II
AVERAGE PSNR (IN dB) OF 100 IMAGES FOR FIVE DIFFERENT PREDICTORS BY USING EXPANSION EMBEDDING TECHNIQUE IN [8]

bits	CNNP	BIP	MEDP	GAP	DP
10,000	47.9	46.4	44.2	41.7	41.3
20,000	44.7	43.2	40.3	38.2	37.8
30,000	42.4	41.1	37.9	36.0	35.7
40,000	40.9	39.4	36.0	34.1	33.7
50,000	39.9	38.3	35.4	33.5	32.5
60,000	38.7	37.3	34.7	32.6	31.4
70,000	38.0	36.6	33.5	31.7	30.9
80,000	37.3	35.6	32.7	31.2	30.3
90,000	36.5	35.2	32.1	30.7	29.9
100,000	35.9	34.8	31.6	30.2	29.3
110,000	35.3	34.3	31.0	29.5	28.7
120,000	34.7	33.7	29.5	28.7	28.2

TABLE III
AVERAGE PSNR (IN dB) OF 100 IMAGES OF THE PROPOSED CNNP-BASED METHOD AND THE METHOD [11]

bits	CNNP	BIP
10,000	58.4	56.8
20,000	55.1	53.8
30,000	52.9	51.6
40,000	51.2	50.1
50,000	49.8	48.8
60,000	48.5	47.5
70,000	47.3	46.4
80,000	46.3	45.4
90,000	45.5	44.7
100,000	44.8	44.0
110,000	43.9	43.2
120,000	43.0	42.3

CNNP so as to compare the proposed CNNP-based RDH scheme with the classical BIP-based RDH method [11]. Table III lists the average PSNR values of these two RDH methods. From this table, we can see that the average PSNR values of the proposed CNNP based RDH method is around 1 dB higher than the BIP-based RDH method. The basic reason is that the BIP can be regarded as a convolution operation with $K = 3$, while the proposed CNNP contains three parallel convolution operations with $K = 3$, $K = 5$, and $K = 7$. In other words, the CNNP has exploited more adjacent pixels for prediction due to its multi receptive field and global optimization capacities.

IV. CONCLUSION

In this letter, we design a CNN-based predictor for grayscale images, which can make full use of the capabilities of the multi receptive fields and global optimization of CNN for RDH. The basic idea is to divide an image into two sets and use one set to predict the other set. The proposed CNNP is a lightweight and computation-efficient network and can be trained well by using 1000 different images. Experimental results have shown that the prediction accuracy of the CNNP is higher than those classical linear predictors. Also, we have used the proposed CNNP with two embedding ways to show that the proposed CNN-based RDH scheme can achieve satisfactory performance by testing 100 different images. To the best of our knowledge, there is no detailed report on how to use CNN to predict a grayscale image for RDH. We consider this work makes a room for the RDH community.

REFERENCES

[1] J. M. Barton, "Method and apparatus for embedding authentication information within digital data," US Patent 5 646 997, Jul. 1997.

[2] C. W. Honsinger, P. W. Jones, M. Rabbani, and J. C. Stoffel, "Lossless recovery of an original image containing embedded data," US Patent 6 278 791, Aug. 2001.

[3] Y.-Q. Shi, X. Li, X. Zhang, H.-T. Wu, and B. Ma, "Reversible data hiding: Advances in the past two decades," *IEEE Access*, vol. 4, pp. 3210–3237, 2016.

[4] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 890–896, Aug. 2003.

[5] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 1147–1156, Aug. 2004.

[6] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, Mar. 2006.

[7] X. Li, W. Zhang, X. Gui, and B. Yang, "Efficient reversible data hiding based on multiple histograms modification," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 9, pp. 2016–2027, Sep. 2015.

[8] D. M. Thodi and J. J. Rodríguez, "Expansion embedding techniques for reversible watermarking," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 721–730, Mar. 2007.

[9] D. Coltuc, "Improved embedding for prediction-based reversible watermarking," *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3, pp. 873–882, Sep. 2011.

[10] D. Coltuc, "Low distortion transform for reversible watermarking," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 412–417, Jan. 2011.

[11] V. Sachnev, H. J. Kim, J. Nam, S. Suresh, and Y. Q. Shi, "Reversible watermarking algorithm using sorting and prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 989–999, Jul. 2009.

[12] L. Luo, Z. Chen, M. Chen, X. Zeng, and Z. Xiong, "Reversible image watermarking using interpolation technique," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 1, pp. 187–193, Mar. 2010.

[13] I. F. Jafar, K. A. Darabkh, R. T. Al-Zubi, and R. A. Al Na'mneh, "Efficient reversible data hiding using multiple predictors," *Comput. J.*, vol. 59, no. 3, pp. 423–438, 2016.

[14] X. Li, B. Yang, and T. Zeng, "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3524–3533, Dec. 2011.

[15] X. Li, J. Li, B. Li, and B. Yang, "High-fidelity reversible data hiding scheme based on pixel-value-ordering and prediction-error expansion," *Signal Process.*, vol. 93, no. 1, pp. 198–205, 2013.

[16] I.-C. Dragoi and D. Coltuc, "Local-prediction-based difference expansion reversible watermarking," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1779–1790, Apr. 2014.

[17] H. Chen, J. Ni, W. Hong, and T.-S. Chen, "High-fidelity reversible data hiding using directionally enclosed prediction," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 574–578, May 2017.

[18] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Object detectors emerge in deep scene CNNs," 2014, *arXiv:1412.6856*.

[19] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 4898–4906, 2016.

[20] T. Luo, G. Jiang, M. Yu, C. Zhong, H. Xu, and Z. Pan, "Convolutional neural networks-based stereo image reversible data hiding method," *J. Vis. Commun. Image Representation*, vol. 61, pp. 61–73, 2019.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[23] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856.

[24] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.

[25] I. Schiopu, Y. Liu, and A. Munteanu, "CNN-based prediction for lossless coding of photographic images," presented at the 2018 Picture Coding Symp. (PCS), San Francisco, CA, USA, Jun. 24–27, 2018.

[26] I. Schiopu and A. Munteanu, "Deep-learning based lossless image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 1829–1842, Jul. 2020.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[30] S. Xiang and Y. Wang, "Non-integer expansion embedding techniques for reversible image watermarking," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, 2015, Art. no. 56.