



**HAL**  
open science

# A Neural Network Approach For Joint Optimization Of Predictors In Lifting-Based Image Coders

Tassnim Dardouri, M. Kaaniche, A. Benazza-Benyahia, J.-C. Pesquet, G. Dauphin

► **To cite this version:**

Tassnim Dardouri, M. Kaaniche, A. Benazza-Benyahia, J.-C. Pesquet, G. Dauphin. A Neural Network Approach For Joint Optimization Of Predictors In Lifting-Based Image Coders. 2021 IEEE International Conference on Image Processing (ICIP), Sep 2021, Anchorage / Virtual, United States. pp.3747-3751, 10.1109/ICIP42928.2021.9506737 . hal-03526467

**HAL Id: hal-03526467**

**<https://hal.science/hal-03526467>**

Submitted on 14 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A NEURAL NETWORK APPROACH FOR JOINT OPTIMIZATION OF PREDICTORS IN LIFTING-BASED IMAGE CODERS

T. Dardouri<sup>1,3</sup>, M. Kaaniche<sup>1</sup>, A. Benazza-Benyahia<sup>2</sup>, J.-C. Pesquet<sup>3</sup>, and G. Dauphin<sup>1</sup>

<sup>1</sup> Université Sorbonne Paris Nord,  
L2TI, UR 3043, F-93430,  
Villetaneuse, France  
tasnim.dardouri@edu.univ-paris13.fr  
mounir.kaaniche@univ-paris13.fr  
gabriel.dauphin@univ-paris13.fr

<sup>2</sup> University of Carthage, SUP'COM,  
LR11TIC01, COSIM Lab,  
Cité Technologique des Communications,  
2083 Ariana, Tunisia  
benazza.amel@supcom.rnu.tn

<sup>3</sup> Centre de Vision Numérique,  
Université Paris-Saclay,  
CentraleSupélec, INRIA,  
91190 Gif-sur-Yvette, France  
jean-christophe.pesquet@centralesupelec.fr

## ABSTRACT

The objective of this paper is to investigate techniques for learning Fully Connected Network (FCN) models in a lifting based image coding scheme. More precisely, based on a 2D non separable lifting structure composed of three FCN-based prediction stages followed by an FCN-based update one, we first propose to resort to an  $\ell_p$  loss function, with  $p \in \{1, 2\}$ , to learn the three FCN prediction models. While the latter are separately learned in the first approach, a novel joint learning approach is then developed by minimizing a weighted  $\ell_p$  loss function related to the global prediction error. Experimental results, carried out on the standard Challenge Learned Image Compression (CLIC) dataset, show the benefits of the proposed techniques in terms of rate-distortion performance.

**Index Terms**— Lifting schemes, adaptive wavelets, image coding, neural networks, optimization, joint learning.

## 1. INTRODUCTION

Lifting Schemes (LS) [1], known also as the second generation wavelets, were found to be efficient for coding different kinds of visual data such as still images [2], stereo images [3], video [4, 5], etc. For instance, they have been adopted by the JPEG2000 image coding standard [6]. A 1D LS based decomposition consists in using a prediction filter and an update one to generate the detail and approximation coefficients, respectively. By applying such 1D decomposition to the rows and then to the columns of an image (or inversely), one approximation subband and three detail subbands oriented horizontally, vertically and diagonally are obtained. Note that a 2D decomposition with three prediction filters and an update one can also be used to generate similarly the wavelet subbands [7, 8].

The performance of these lifting based coding schemes depends on the involved prediction and update filters. For this reason, many research works have been developed to design optimal filters and make the lifting schemes well adapted to the image to be encoded. Concerning the prediction filters, they are often optimized by minimizing the variance of the detail coefficients [2]. Moreover, instead of minimizing an  $\ell_2$  criterion, an  $\ell_1$ -based minimization technique has been introduced in [9] to produce sparse wavelet representation. Furthermore, the minimization of the entropy of the detail coefficients has also been investigated in [10, 11]. Regarding the update filter, two main optimization techniques have been developed. The first one aims at minimizing the reconstruction error while setting the detail coefficients to zero at the synthesis stage [2, 12]. While

this approach results in a complex linear system of equations, the second one consists in minimizing the difference between the approximation coefficients and the output of an ideal low-pass filter applied to the image [13].

In order to take advantage of the neural networks in providing accurate nonlinear approximation, a Convolutional Neural Network (CNN) based LS for image coding has been introduced in [14]. More precisely, the prediction step is performed using a CNN while the update one is simply replaced by an average operation. More recently, a fully nonlinear transform, where both prediction and update stages are performed using Fully Connected Networks (FCN), has been introduced in [15]. In addition to these neural network based lifting schemes, deep learning based image compression techniques have also attracted a lot of attention in the last years. To this end, different neural network architectures have been developed to perform the analysis and synthesis stages while using standard quantization and entropy coding schemes [16, 17, 18, 19, 20]. These methods, known as end-to-end compression methods, mainly differ in the chosen neural network model or the employed loss function which is often optimized in an end-to-end manner.

In this paper, we propose to further investigate neural networks for the design of lifting based coding schemes. More precisely, based on a recent work where a 2D non separable lifting scheme is developed using three FCN-based prediction stages followed by an FCN-based update one [15], we propose here a novel algorithm to learn the FCN prediction models. To this end, we will first focus on the minimization of an  $\ell_p$  criterion (with  $p \in ]0, +\infty[$ ) instead of the common  $\ell_2$  one as performed in [15]. Moreover, unlike [15] where the involved FCN models are *separately* learned, the main contribution of this work is to *jointly* learn the FCN prediction models by minimizing a weighted  $\ell_p$  criterion related to the global prediction error.

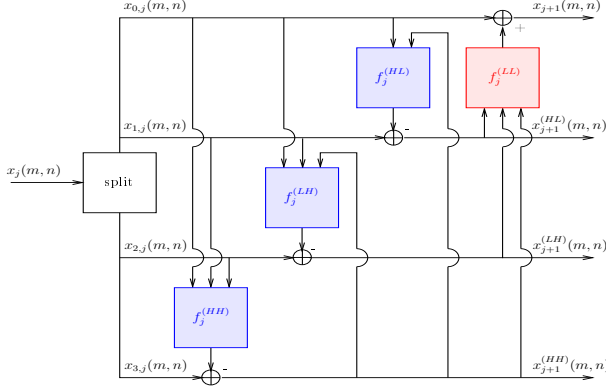
The remainder of this paper is organized as follows. In Section 2, we recall the recent FCN-based lifting scheme. The proposed learning approaches of the different FCN models are described in Section 3. Finally, experimental results are shown in Section 4 and some conclusions are drawn in Section 5.

## 2. FULLY CONNECTED NETWORK BASED LIFTING SCHEME

### 2.1. FCN based lifting architecture

While 1D LS has been used in the JPEG2000 image compression standard, 2D Non Separable LS (NSLS) are found to be more efficient to cope with the different characteristics of edges of the input

images to be encoded. A typical NSLS is composed of three prediction stages followed by an update one. Instead of using linear operators as it is often the case in conventional LS, it has been recently proposed to perform the prediction and update stages using FCN models. The resulting FCN based lifting architecture is shown in Fig. 1.



**Fig. 1.** FCN-based lifting architecture.

For each input image  $x$ , let  $x_j$  denote the approximation subband at resolution level  $j$  (where  $x_0 = x$ ). Thus, each input signal  $x_j(m, n)$  is first split into four polyphase components designated by  $x_{0,j}(m, n) = x_j(2m, 2n)$ ,  $x_{1,j}(m, n) = x_j(2m, 2n + 1)$ ,  $x_{2,j}(m, n) = x_j(2m + 1, 2n)$ , and  $x_{3,j}(m, n) = x_j(2m + 1, 2n + 1)$ . Then, the conventional prediction filters used in classical LS are replaced by the FCN models denoted here by  $f_j^{(HH)}$ ,  $f_j^{(LH)}$ ,  $f_j^{(HL)}$ . The latter allow to produce the diagonal detail coefficients  $x_{j+1}^{(HH)}$ , the vertical ones  $x_{j+1}^{(LH)}$  and the horizontal ones  $x_{j+1}^{(HL)}$  given by

$$\begin{aligned} \forall o \in \{HH, LH, HL\}, \\ x_{j+1}^{(o)}(m, n) &= x_{i,j}(m, n) - \hat{x}_{i,j}(m, n) \\ &= x_{i,j}(m, n) - f_j^{(o)}(\tilde{\mathbf{x}}_j^{(o)}(m, n)) \end{aligned} \quad (1)$$

where  $x_{i,j}(m, n)$  (with  $i \in \{1, 2, 3\}$ ) is the polyphase component to be predicted,  $\tilde{\mathbf{x}}_j^{(o)}(m, n)$  is the input reference vector containing the samples used to generate the detail coefficients  $x_{j+1}^{(o)}$  and  $\hat{x}_{i,j}(m, n) = f_j^{(o)}(\tilde{\mathbf{x}}_j^{(o)}(m, n))$  is the predicted value. Therefore,  $\hat{x}_{i,j}(m, n)$  can be seen as the output of an FCN model applied to the reference vector  $\tilde{\mathbf{x}}_j^{(o)}(m, n)$ . More precisely, the latter will be associated to the input layer of the FCN model. Then,  $H$  hidden layers are used. Note that the neuron responses are computed based on a linear combination with bias followed by a nonlinear activation function. Finally, an output layer with a single neuron is employed to produce the predicted value  $\hat{x}_{i,j}(m, n)$  using only a linear combination of the neuron values of the previous layer (i.e last hidden layer).

After these three FCN-based prediction stages, an FCN model  $f_j^{(LL)}$  is employed in the update stage to generate the approximation coefficients  $x_{j+1}$ :

$$\begin{aligned} x_{j+1}(m, n) &= x_{0,j}(m, n) + \hat{t}_j(m, n) \\ &= x_{0,j}(m, n) + f_j^{(LL)}(\tilde{\mathbf{x}}_{j+1}(m, n)) \end{aligned} \quad (2)$$

where  $\tilde{\mathbf{x}}_{j+1}(m, n)$  is the input reference vector containing the detail coefficients used to generate the approximation coefficients. Simi-

larly to the FCN-based prediction stage,  $\tilde{\mathbf{x}}_{j+1}(m, n)$  will be associated with the input layer of the FCN-based update stage. Then,  $H$  hidden layers followed by an output layer with a single neuron are applied to produce  $\hat{t}_j(m, n)$ , which allows us to deduce the approximation coefficients.

## 2.2. Learning of the FCN-based prediction models

Let  $\Theta_j^{(o)}$ , with  $o \in \{HH, LH, HL\}$ , denote the weight parameters associated to the FCN model  $f_j^{(o)}$ . For a given detail subband with orientation  $o$ , the weight vector  $\Theta_j^{(o)}$  is learned through several forward and backward propagation passes while minimizing a given loss function. In [15], each FCN-based prediction model is trained by minimizing the Mean Square Error (MSE) between the target pixels  $x_{i,j}(m, n)$  and the predicted ones  $\hat{x}_{i,j}(m, n)$ . The corresponding loss function is given by

$$\mathcal{L}_2(\Theta_j^{(o)}) = \frac{1}{M_j N_j} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} (x_{i,j}(m, n) - \hat{x}_{i,j}(m, n))^2 \quad (3)$$

where  $M_j$  and  $N_j$  correspond to the dimensions of the image  $x_j$  divided by 2. Note that  $M_j \times N_j$  represents also the number of samples (i.e., mini-batch size) used during each iteration of the training phase. Thus, a Mini-Batch Gradient Descent (MBGD) algorithm is used to optimize the above loss function [21]. After the convergence of the algorithm, the optimal weights  $\Theta_j^{(o)}$  are obtained and finally applied to the test images to generate the predicted pixels  $\hat{x}_{i,j}(m, n)$  and deduce the detail coefficients  $x_{j+1}^{(o)}$ . It is important to emphasize here that the above learning process is *separately* applied to each prediction step of the wavelet decomposition.

## 2.3. Learning of the FCN-based update model

Once the three FCN prediction models are learned and the three detail subbands of the training images are produced, the FCN update model will be trained. The main difference with the previous FCN-based prediction learning stage is the choice of the loss function. In this respect, we have retained a standard criterion which aims at minimizing the quadratic error between the approximation subband and the decimated version of the output of an ideal low-pass filter applied to the input signal  $x_j$ . Let us recall that this error is defined as follows:

$$\begin{aligned} e_j(m, n) &= y_{j+1}(m, n) - x_{j+1}(m, n) \\ &= y_{j+1}(m, n) - x_{0,j}(m, n) - f_j^{(LL)}(\tilde{\mathbf{x}}_{j+1}(m, n)) \end{aligned} \quad (4)$$

where  $y_{j+1}(m, n) = (\tilde{h} * x_j)(2m, 2n)$  and  $\tilde{h}$  is the impulse response of the ideal rectangular low-pass filter.

According to (4), the optimization of the FCN-based update model becomes similar to that of an FCN-based prediction task and aims at predicting the target signal  $t_j(m, n) = y_{j+1}(m, n) - x_{0,j}(m, n)$  from the input reference vector  $\tilde{\mathbf{x}}_{j+1}(m, n)$ . Therefore, similarly to (3), the FCN-based update model will be trained by minimizing the MSE between  $t_j(m, n)$  and  $\hat{t}_j(m, n)$ .

## 3. FROM INDEPENDENT TO JOINT LEARNING OF THE FCN BASED PREDICTION MODELS

While the learning of the FCN-based update model is kept unchanged, we investigate now two techniques for learning the three FCN-based prediction models.

### 3.1. $\ell_p$ loss function based independent learning approach

In the considered FCN-based NSLS, as shown by (3), each FCN-based prediction model is learned by minimizing the energy (i.e.  $\ell_2$ -norm) of its corresponding wavelet detail coefficients. Generally, wavelet coefficient statistics can be efficiently exploited to further improve the image compression performance [9, 22]. To this end, the Generalized Gaussian Distribution (GGD) can be used to model the resulting wavelet coefficients  $x_{j+1}^{(o)}$  [23]. The latter are viewed as realizations of a random variable  $X_{j+1}^{(o)}$  whose probability density function  $g_{j+1}^{(o)}$  is given by

$$\forall \xi \in \mathbb{R}, g_{j+1}^{(o)}(\xi; \alpha_{j+1}^{(o)}, \beta_{j+1}^{(o)}) = \frac{\beta_{j+1}^{(o)}}{2\alpha_{j+1}^{(o)}\Gamma(\frac{1}{\beta_{j+1}^{(o)}})} e^{-\left(\frac{|\xi|}{\alpha_{j+1}^{(o)}}\right)^{\beta_{j+1}^{(o)}}}$$

where  $\Gamma$  is the Gamma function,  $\alpha_{j+1}^{(o)} \in ]0, +\infty[$  is the scale parameter, and  $\beta_{j+1}^{(o)} \in ]0, +\infty[$  is the shape parameter of the GGD. In the particular case when  $\beta_{j+1}^{(o)} = 2$  (resp.  $\beta_{j+1}^{(o)} = 1$ ), the GGD becomes the Gaussian distribution (resp. the Laplacian one).

Using a uniform scalar quantization scheme, the discrete entropy of the resulting quantized source can be approximated at high bitrate by the differential entropy  $h$  of the variable  $X_{j+1}^{(o)}$  [24]. The latter is approximatively equal to

$$h(X_{j+1}^{(o)}) \approx \frac{1}{M_j N_j \ln(2)(\alpha_{j+1}^{(o)})^{\beta_{j+1}^{(o)}}} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} \left| x_{j+1}^{(o)}(m, n) \right|^{\beta_{j+1}^{(o)}} + \log_2 \left( \frac{2\alpha_{j+1}^{(o)}\Gamma(\frac{1}{\beta_{j+1}^{(o)}})}{\beta_{j+1}^{(o)}} \right). \quad (5)$$

The above equation shows that the  $\ell_{\beta_{j+1}^{(o)}}$ -norm of the detail coefficients can be seen as a good approximation of their entropy that should be minimized. Based on this observation, an analysis of the wavelet coefficients distribution has been conducted for a large set of images taken from the standard CLIC dataset. The latter will be described in Section 4. It has been noticed that the shape parameters of the detail subbands  $\beta_{j+1}^{(o)}$  range from 0.15 to 1.75 and their mean values are between 0.6 and 0.7. As a result, instead of only minimizing the  $\ell_2$ -norm of the detail coefficients as performed in [15], we propose here to consider an  $\ell_p$  criterion. For example, using an  $\ell_1$  loss function allows us to increase the sparsity of the wavelet coefficients to be encoded. Therefore, the three FCN-based prediction models can be independently trained by minimizing the following loss function:

$$\mathcal{L}_p(\Theta_j^{(o)}) = \frac{1}{M_j N_j} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} \left| x_{i,j}(m, n) - \hat{x}_{i,j}(m, n) \right|^p \quad (6)$$

where  $p \in ]0, +\infty[$ . This loss function (6) is optimized using the MBGD algorithm and following the strategy described in Sec. 2.2.

### 3.2. Weighted $\ell_p$ loss function based joint learning approach

Up to now, each FCN-based prediction model has been separately trained by minimizing the  $\ell_p$ -norm of its corresponding detail subband  $x_{j+1}^{(o)}$ . While such loss function seems to be appropriate for learning the first-FCN-based prediction model  $\Theta_j^{(HH)}$ , it may be

suboptimal for learning the second and third ones ( $\Theta_j^{(LH)}$  and  $\Theta_j^{(HL)}$ ). Indeed, according to the lifting architecture shown in Fig. 1, it can be seen that the output of the first FCN-based prediction model, and more specifically the diagonal detail subband  $x_{j+1}^{(HH)}$ , is used as a reference signal during the second and third FCN-based prediction models to generate the vertical and horizontal detail coefficients, respectively. Therefore, it becomes more interesting to learn the first FCN based prediction model by minimizing a global loss function defined on the three resulting detail signals. More precisely, the vector of parameters  $\Theta_j^{(HH)}$  will be learned by minimizing the sum of a weighted  $\ell_p$  criterion, yielding the following loss function:

$$\mathcal{L}_{w\ell_p}(\Theta_j^{(HH)}) = \sum_{o \in \{HH, LH, HL\}} \frac{1}{(\tilde{\alpha}_{j+1}^{(o)})^p} \sum_{m,n} \left| x_{j+1}^{(o)}(m, n) \right|^p \quad (7)$$

where

$$\tilde{\alpha}_{j+1}^{(o)} = \frac{1}{K} \sum_{k=1}^K \left( \frac{p}{M_j N_j} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} |\tilde{x}_{j+1}^{(o,k)}(m, n)|^p \right)^{1/p}, \quad (8)$$

$K$  is the number of training images, and  $\tilde{x}_{j+1}^{(o,k)}$  is the detail subband of the  $k$ -th training image obtained using the independent  $\ell_p$ -based learned FCN model. To learn the three FCN prediction models (per resolution level), two learning approaches can be employed.

First, since the learning of the first FCN model depends on the learning of the two other FCN prediction models and vice-versa, we can resort to a joint learning approach that alternates between the update of the  $\Theta_j^{(HH)}$  and that of  $\Theta_j^{(LH)}$  as well as  $\Theta_j^{(HL)}$ . In this respect, the training phase will consist in applying an iterative algorithm, composed of two main steps, to each image of the training dataset. Starting from a given initialization of  $\Theta_j^{(HH)}$ ,  $\Theta_j^{(LH)}$  and  $\Theta_j^{(HL)}$ , the first step aims to update  $\Theta_j^{(HH)}$  by minimizing the weighted loss function  $\mathcal{L}_{w\ell_p}(\Theta_j^{(HH)})$  using the MBGD algorithm. In the second step, and after computing the detail coefficients  $x_{j+1}^{(HH)}$  using the obtained  $\Theta_j^{(HH)}$ , we update simultaneously  $\Theta_j^{(LH)}$  and  $\Theta_j^{(HL)}$  by minimizing the loss functions  $\mathcal{L}_p(\Theta_j^{(LH)})$  and  $\mathcal{L}_p(\Theta_j^{(HL)})$  using the MBGD algorithm, respectively. This iterative algorithm is applied to each image of the training dataset for many epochs until its convergence.

While the above iterative approach results in a high computational time, the second one consists of directly learning the three FCN prediction models. More precisely, the three sets of weight parameters can be grouped within a single vector of parameters  $\Theta_j = (\Theta_j^{(HH)}, \Theta_j^{(LH)}, \Theta_j^{(HL)})^\top$  which is directly learned by minimizing the  $\mathcal{L}_{w\ell_p}(\Theta_j)$  using the MBGD algorithm. It should be noted here that the two aforementioned learning approaches have been tested and resulted in similar coding performance.

## 4. EXPERIMENTAL RESULTS

Our simulations are carried out using the standard CLIC dataset<sup>1</sup>. More precisely, 585 images with different sizes are used for training the neural network models. In the test phase, 40 crop images of size  $512 \times 512$  are selected from the test CLIC dataset. Since

<sup>1</sup><http://www.compression.cc/2018/challenge/>

the FCN-based NSLS architecture is developed for 2D images, the CLIC images are converted to grayscale.

As an example of lifting structure, we have considered the NSLS(4,2) transform [13] over three resolution levels. In this NSLS architecture, each FCN model is implemented using 4 hidden layers with  $128 \times 64 \times 32 \times 16$  neurons. The Parametric Rectified Linear Unit (PReLU) has been employed as an activation function. For the MGD algorithm, the batch size corresponds to the number of samples to be predicted in each input image  $x_j$ . The learning rate is set to  $10^{-2}$  while applying a decay of  $10^{-4}$  and 200 epochs have been used for training our FCN prediction models. These implementations are performed using Keras and TensorFlow 2.4 on a NVIDIA Tesla V100 32 GB GPU.

The proposed FCN-based lifting coding scheme will be firstly evaluated in the context of lossy compression. To this end, our methods have been compared to the JPEG2000 compression standard as well as to two recent neural network-based coding schemes. The first one, denoted by “Ballé [18]”, is an end-to-end image compression method composed of nonlinear analysis and synthesis transforms learned using a rate-distortion loss function. The second one, more related to our work, employs a neural network-based update-predict lifting structure where the update step is a mean filter and the prediction one is performed using a CNN. The latter is trained by minimizing the MSE between the original image and the reconstructed one. This method will be designated by “CNN-LS [14]”. Our proposed approaches have been implemented using  $p \in \{1, 2\}$ . The methods based on the  $\ell_2$ ,  $\ell_1$ , weighted  $\ell_2$  and weighted  $\ell_1$  criteria are respectively denoted by “FCN-LS-L2”, “FCN-LS-L1”, “FCN-LS-WL2” and “FCN-LS-WL1”. In order to take into account the diversity of textures and contents in the test images, we have also considered a hybrid coding method which selects the best model between the four FCN-based prediction models. Note that this method, designated by “H-FCN-LS”, results in a negligible overhead (2 bits per image) that needs to be sent to the decoder.

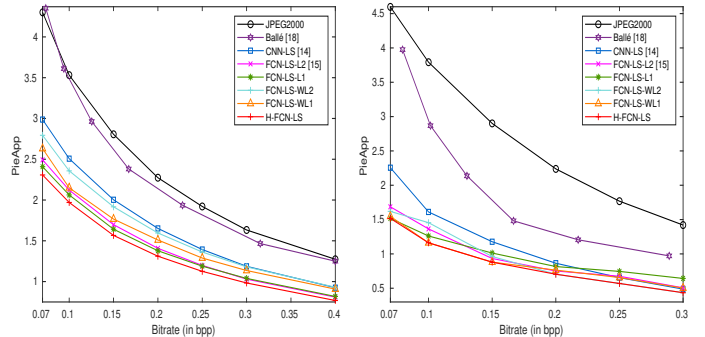
Fig. 2 illustrates the Rate-Distortion (R-D) results, obtained with all the dataset images (Fig. 2(a)) and a given one (Fig. 2(b)), using the Perceptual Image-Error Assessment through Pairwise Preference (PieAPP) metric [25]. Let us recall that this metric was found to be well correlated with human opinion scores. Note that lower PieAPP values indicate better subjective image quality. Thus, it can be seen that the FCN-based LS coding methods outperform the conventional ones. Most importantly, the proposed joint learning approaches, and more specifically the hybrid coding scheme, lead to better coding performance compared to the independent learning approach. Furthermore, Fig. 3 shows the interest of the proposed method in terms of reconstruction quality.

Finally, our methods are evaluated in the context of lossless compression using the entropy of the resulting wavelet representations. Since the aforementioned methods as well as most of neural networks based coding schemes are suitable for lossy compression except in a few recent works [26, 27], we will compare our proposed methods to JPEG2000. Table 1 indicates that  $\ell_1$  and weighted  $\ell_1$  loss functions achieve a gain of about 0.07 bits per pixels (bpp) compared to the  $\ell_2$  and weighted  $\ell_2$  ones, respectively.

## 5. CONCLUSION AND PERSPECTIVES

In this paper, we have proposed a joint optimization approach for learning FCN prediction models in a lifting based image coding scheme. The main idea consists in jointly training three FCN models by minimizing a weighted  $\ell_p$  loss function with  $p \in \{1, 2\}$ . The obtained results show the good performance of the proposed approach

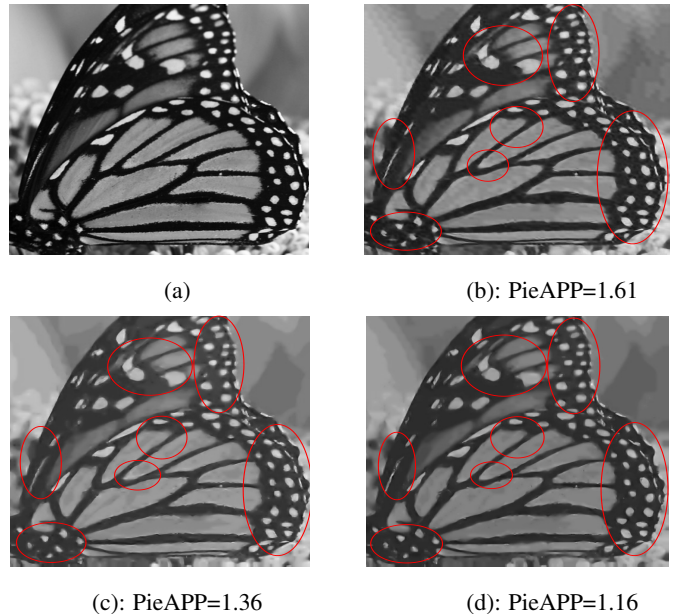
in the context of lossy and lossless compression. In future work, a more general  $\ell_p$  loss function, where various  $p$  values are selected for different wavelet subbands, will be investigated.



(a): All the dataset images

(b): Image 21

Fig. 2. R-D results of the CLIC dataset using the PieAPP metric.



(a)

(b): PieAPP=1.61

(c): PieAPP=1.36

(d): PieAPP=1.16

Fig. 3. Original test image (a) and reconstructed ones at 0.1 bpp using: (b) CNN-LS [14], (c) FCN-LS-L2 [15], (d) H-FCN-LS.

Table 1. Average entropy (in bpp) of wavelet coefficients.

Method	JPEG 2000	FCN- LS-L2	FCN- LS-L1	FCN- LS-WL2	FCN- LS-WL1	H-FCN -LS
Entropy	4.41	4.23	4.17	4.23	4.16	<b>4.15</b>

## 6. REFERENCES

- [1] W. Sweldens, “The lifting scheme: A custom-design construction of biorthogonal wavelets,” *Applied and Computational Harmonic Analysis*, vol. 3, no. 2, pp. 186–200, April 1996.

- [2] A. Gouze, M. Antonini, M. Barlaud, and B. Macq, "Design of signal-adapted multidimensional lifting schemes for lossy coding," *IEEE Transactions on Image Processing*, vol. 13, no. 12, pp. 1589–1603, December 2004.
- [3] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, and J.-C. Pesquet, "Vector lifting schemes for stereo image coding," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2463–2475, 2009.
- [4] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 2001, vol. 3, pp. 1793–1796.
- [5] E. Martinez-Enriquez, J. Cid-Sueiro, F. Diaz de Mari a, and A. Ortega, "Directional transforms for video coding based on lifting on graphs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 4, pp. 933–946, November 2016.
- [6] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Norwell, MA, USA, 2002.
- [7] Y.-K. Sun, "A two-dimensional lifting scheme of integer wavelet transform for lossless image compression," in *International Conference on Image Processing*, Singapore, October 2004, vol. 1, pp. 497–500.
- [8] M. Kaaniche, J.-C. Pesquet, A. Benazza-Benyahia, and B. Pesquet-Popescu, "Two-dimensional non separable adaptive lifting scheme for still and stereo image coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, Texas, USA, March 2010.
- [9] M. Kaaniche, B. Pesquet-Popescu, A. Benazza-Benyahia, and J.-C. Pesquet, "Adaptive lifting scheme with sparse criteria for image coding," *EURASIP Journal on Advances in Signal Processing: Special Issue on New Image and Video Representations Based on Sparsity*, vol. 2012, no. 1, pp. 1–22, January 2012.
- [10] J. Solé and P. Salembier, "Generalized lifting prediction optimization applied to lossless image compression," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 695–698, October 2007.
- [11] A. Benazza-Benyahia, J.-C. Pesquet, J. Hattay, and H. Masmoudi, "Block-based adaptive vector lifting schemes for multichannel image coding," *EURASIP International Journal of Image and Video Processing*, vol. 2007, no. 1, pp. 10 pages, January 2007.
- [12] B. Pesquet-Popescu, *Two-stage adaptive filter bank*, first filling date 1999/07/27, official filling number 99401919.8, European patent number EP1119911, 1999.
- [13] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, and J.-C. Pesquet, "Non separable lifting scheme with adaptive update step for still and stereo image coding," *Elsevier Signal Processing: Special issue on Advances in Multirate Filter Bank Structures and Multiscale Representations*, vol. 91, no. 12, pp. 2767–2782, January 2011.
- [14] H. Ma, D. Liu, R. Xiong, and F. Wu, "iWave: CNN-based wavelet-like transform for image compression," *IEEE Transactions on Multimedia*, vol. 22, no. 7, pp. 1667–1697, July 2020.
- [15] T. Dardouri, M. Kaaniche, A. Benazza-Benyahia, and J.-C. Pesquet, "Optimized lifting scheme based on a dynamical fully connected network for image coding," in *IEEE International Conference on Image Processing*, Abu Dhabi, United Arab Emirates, October 2020, pp. 1–5.
- [16] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell, "Full resolution image compression with recurrent neural networks," in *Computer Vision and Pattern Recognition*, Las Vegas, USA, June 2016, pp. 5306–5314.
- [17] O. Rippel and L. Bourdev, "Real-time adaptive image compression," in *International Conference on Machine Learning*, Sydney, Australia, August 2017, pp. 1–9.
- [18] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," in *International Conference on Learning Representations*, Toulon, France, April 2017, pp. 1–27.
- [19] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang, "Learning convolutional networks for content-weighted image compression," in *Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018, pp. 3214–3223.
- [20] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and V. G. Luc, "Generative adversarial networks for extreme learned image compression," in *International Conference on Learning Representations*, New Orleans, Louisiana, USA, May 2019, pp. 1–31.
- [21] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, San Diego, USA, May 2015, pp. 1–15.
- [22] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation quantization framework," in *Data Compression Conference*, Snowbird, USA, March 1997, pp. 221–230.
- [23] S. Mallat, "A theory for multiresolution signal decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.
- [24] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. 14, no. 5, pp. 676–683, 1969.
- [25] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "PieAPP: Perceptual image-error assessment through pairwise preference," in *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018, pp. 1–10.
- [26] I. Schiopu and A. Munteanu, "Deep-learning based lossless image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 1829–1842, April 2019.
- [27] I. Schiopu and A. Munteanu, "A study of prediction methods based on machine learning techniques for lossless image coding," in *IEEE International Conference on Image Processing*, Abu Dhabi, United Arab Emirates, October 2020, pp. 1–5.