Convolutional Neural Network for Image Compression with Application to JPEG Standard

Dariusz Puchala, *PhD* and Kamil Stokfiszewski, *PhD*
Institute of Information Technology
Lodz University of Technology
ul. Wolczanska 215, 90-924 Lodz, Poland

{dariusz.puchala,kamil.stokfiszewski}@p.lodz.pl

Data
Compression
Conference

Introduction

- Convolutonal neural network (CNN)
  - Linear with 8x8 pixel input/output
  - Trained on the *FFHQ* database of face images
- Embedded in JPEG image compression standard
  - Replaces standard Discrete Cosine Transform (DCT)
- Compared with other embedded discrete transforms
  - DCT (suboptimal)
  - KLT (optimal)
  - Lapped transforms MLT and LOT (tiling effect reduction)
- Performance
  - better than the remaining approaches in terms of objective image quality measures PSNR/SSIM
  - better than the remaining approaches in terms of tiling effect reduction

The JPEG Image Compression Standard

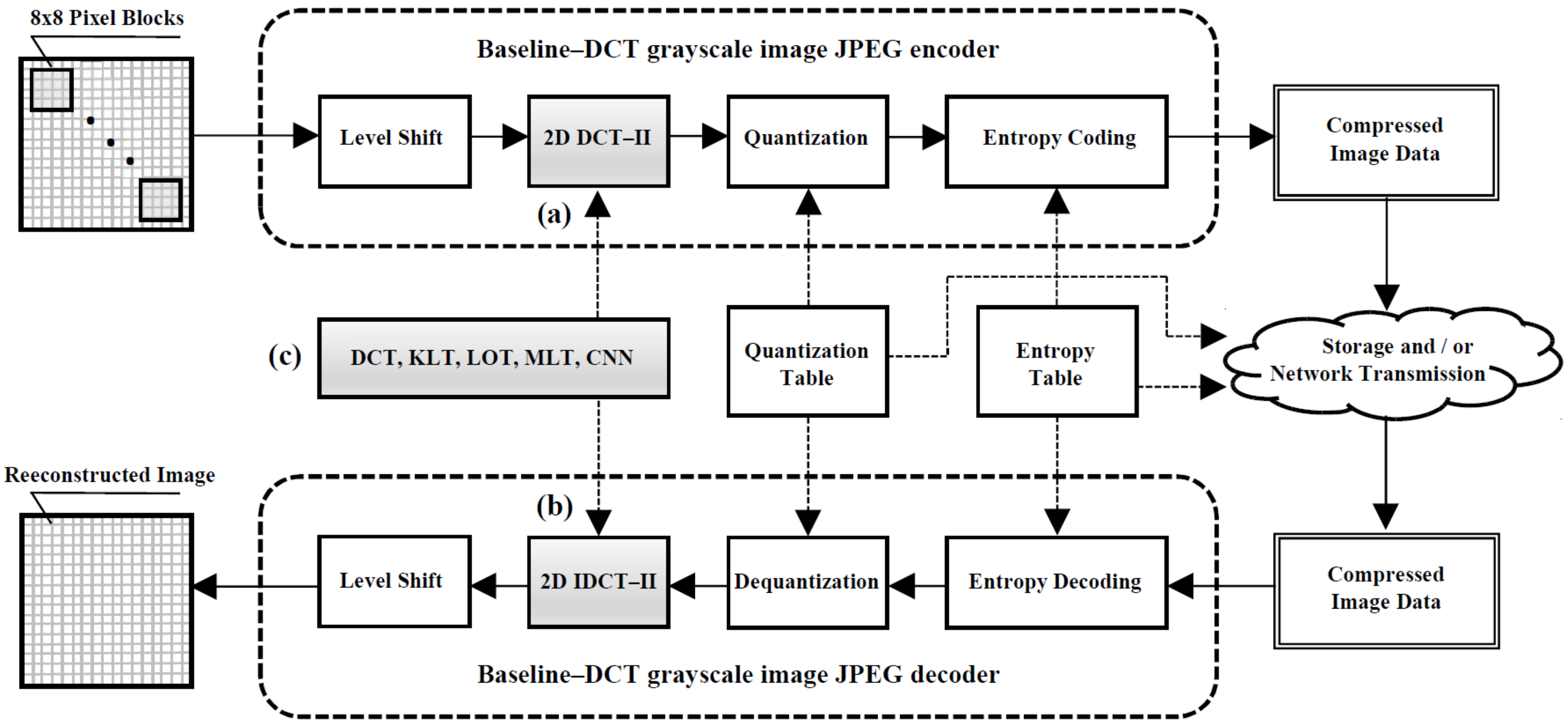


Fig. 1. The JPEG image compression standard operation diagram, **(a)** transform coding stage with standard DCT, **(b)** transform decoding stage with standard IDCT, **(c)** standard transform replacement with KLT, LOT, MLT and CNN.

The JPEG Image Compression Standard

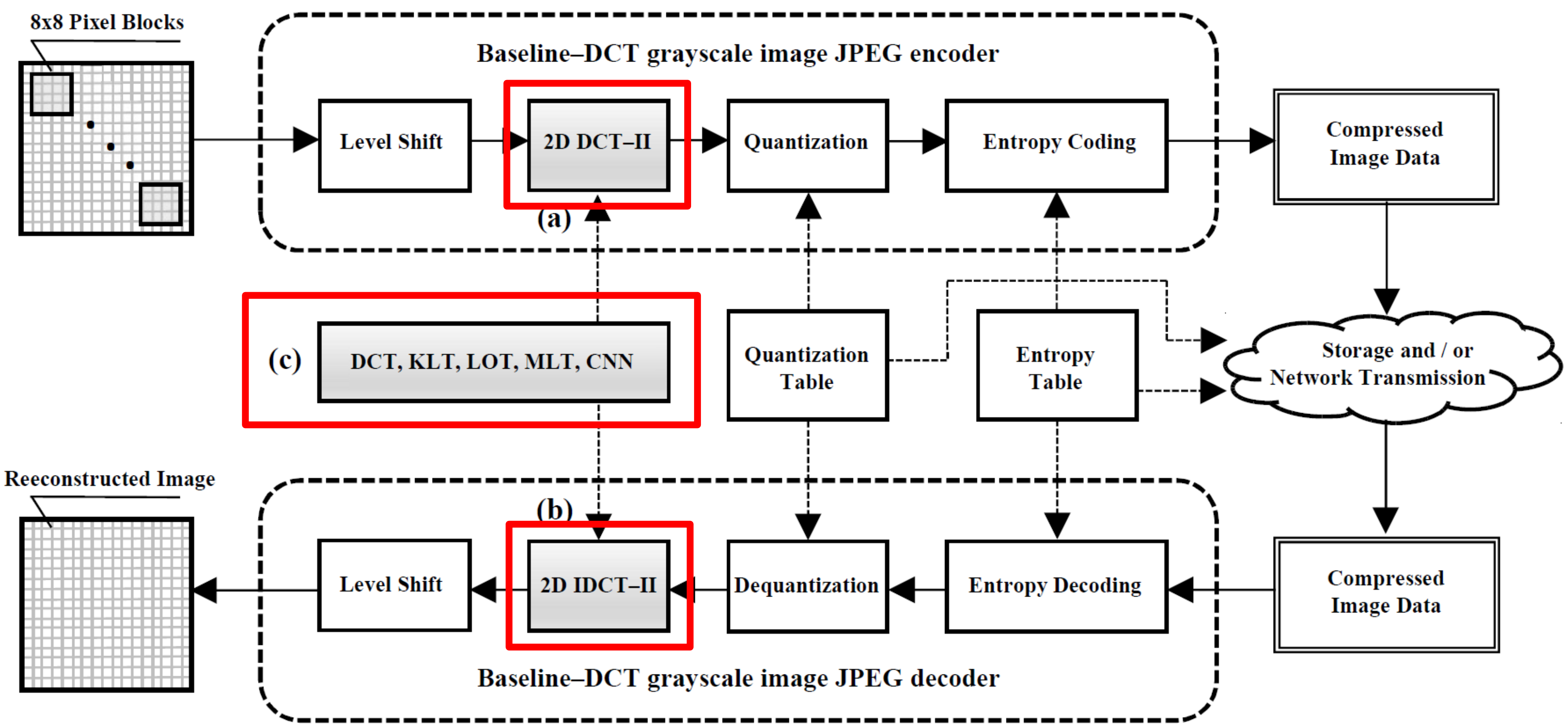


Fig. 1. The JPEG image compression standard operation diagram, **(a)** transform coding stage with standard DCT, **(b)** transform decoding stage with standard IDCT, **(c)** standard transform replacement with KLT, LOT, MLT and CNN.

Orthogonal Transforms in Image Compression

- 2D Discrete Cosine Transform (suboptimal)

$$\mathbf{G} \triangleq \mathbf{C} \otimes \mathbf{C}, \quad (\mathbf{C})_{mn} \triangleq c_m \sqrt{\frac{2}{N}} \, cos\left[\frac{(2\,n+1)\,m\,\pi}{2\,N}\right], \quad c_m \triangleq \begin{cases} \frac{1}{\sqrt{2}} & \text{for } m=0 \\ 1 & \text{for } m=1,\ldots,N-1 \end{cases}$$

- 2D Karhunen-Loève Transform (optimal)

$$\mathbf{U}\,(\,\mathbf{W}\cdot\mathbf{W}^T)\,\mathbf{U}^T = diag\,(\lambda_1,\ldots,\lambda_{N^2}), \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{N^2} \in \mathbb{R}_+$$

- 2D Lapped Orthogonal Transform (tiling effect reduction)

$$\mathbf{U} = \frac{1}{2}\mathbf{Z}\begin{bmatrix} \mathbf{D}_s - \mathbf{D}_a & \mathbf{D}_s + \mathbf{D}_a \\ \mathbf{D}_s - \mathbf{D}_a & -(\mathbf{D}_s + \mathbf{D}_a) \end{bmatrix}$$

- 2D Modulated Lapped Transform (tiling effect reduction)

The MLT is a modulated transform having a property of aliasing cancellation bank of *N* filters with the length of impulse responses equal to 2*N*. It is constructed by an appropriate selection of the low-pass filter prototype and the proper choice of the window function.

Proposed Convolutional Neural Network Architecture

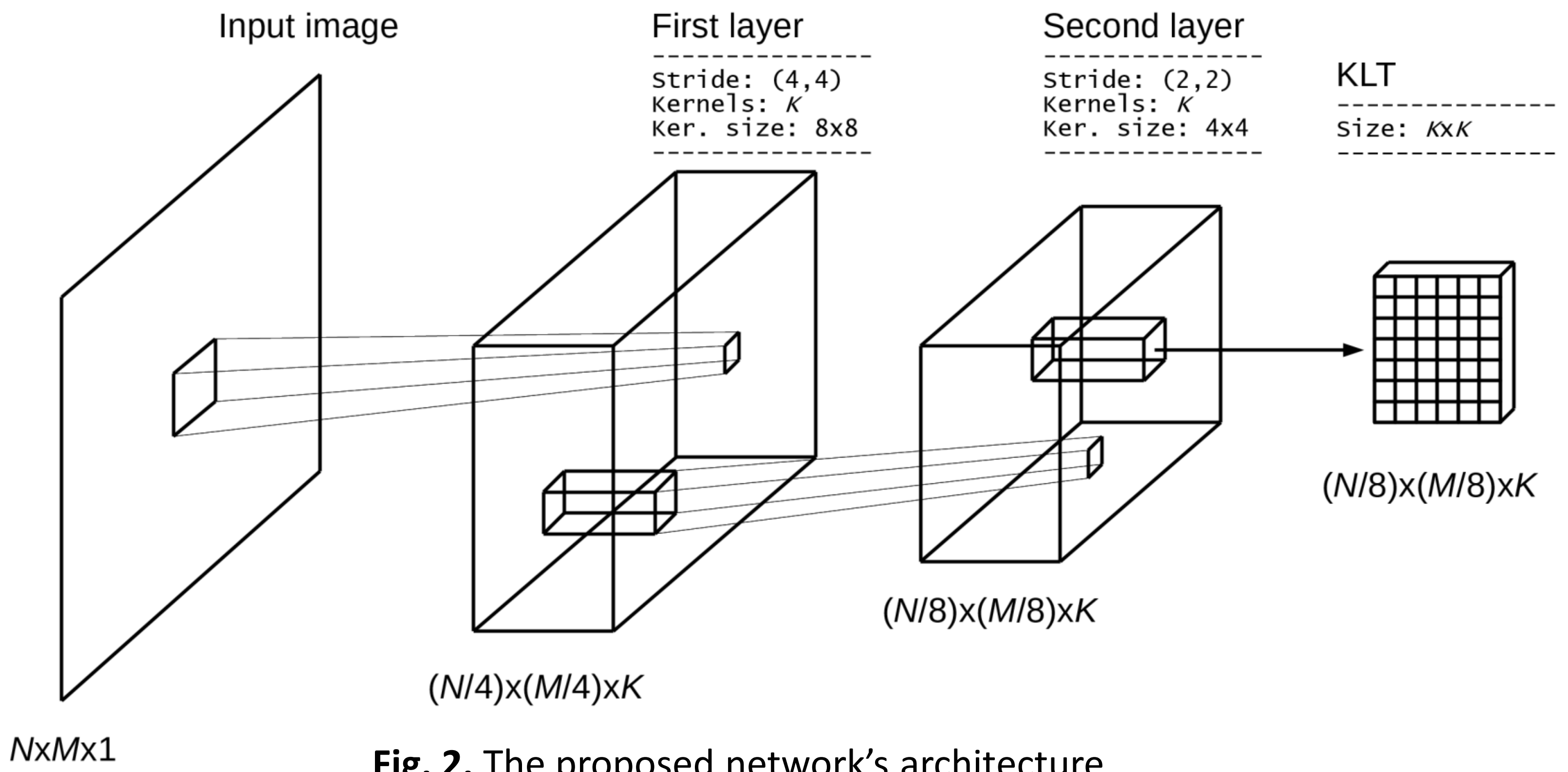


Fig. 2. The proposed network's architecture.

- 1st convolutional layer - *K* kernels of size *8*x*8* elements and the stride steps (*4*, *4*).
- 2nd convolutional layer - *K* kernels of size *4*x*4* elements and the stride steps (*2*, *2*).
- 3rd stage - *K*-point KLT calculated for vectors of pixels coming from subsequent feature maps. Improvement of the energy compaction based on the correlation between feature images.

KLT application in the last processing stage

Fig. 3a. Feature map images for 'lena1024' image at output of the CNN's 2nd layer.

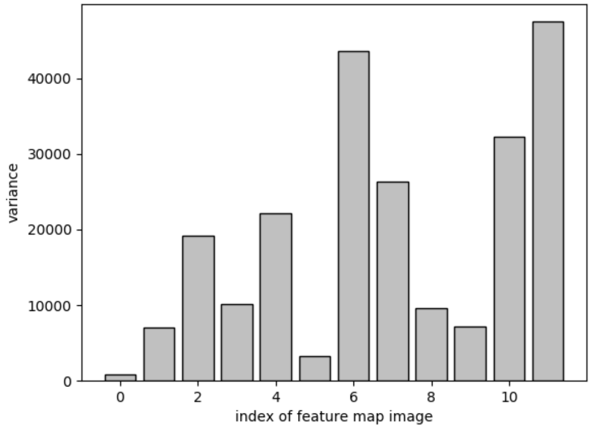


Fig. 3b. Energy distribution for 'lena1024' image at the output of the CNN's 2nd layer.

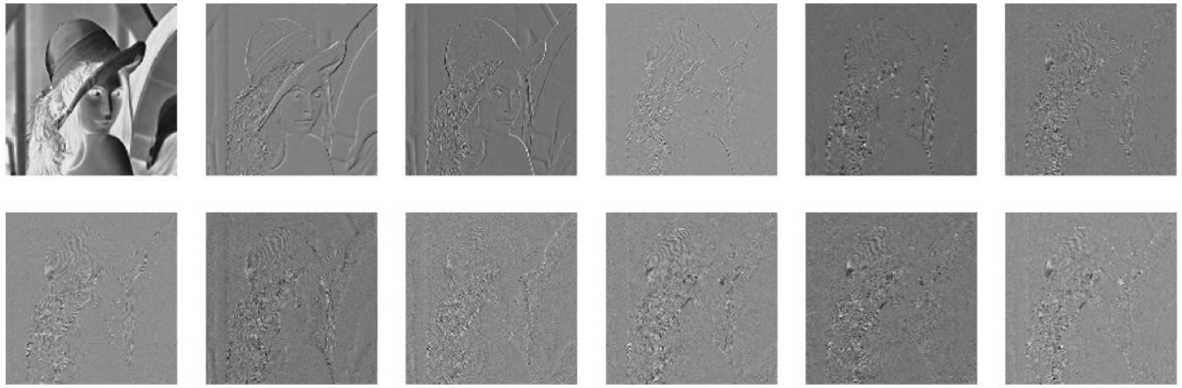


Fig. 4a. Feature map images for 'lena1024' image after application of KLT.

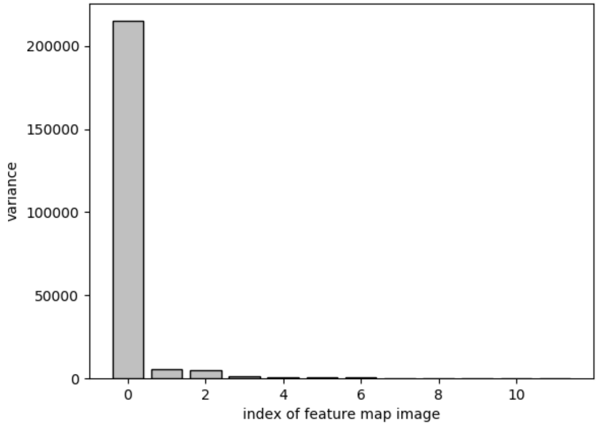


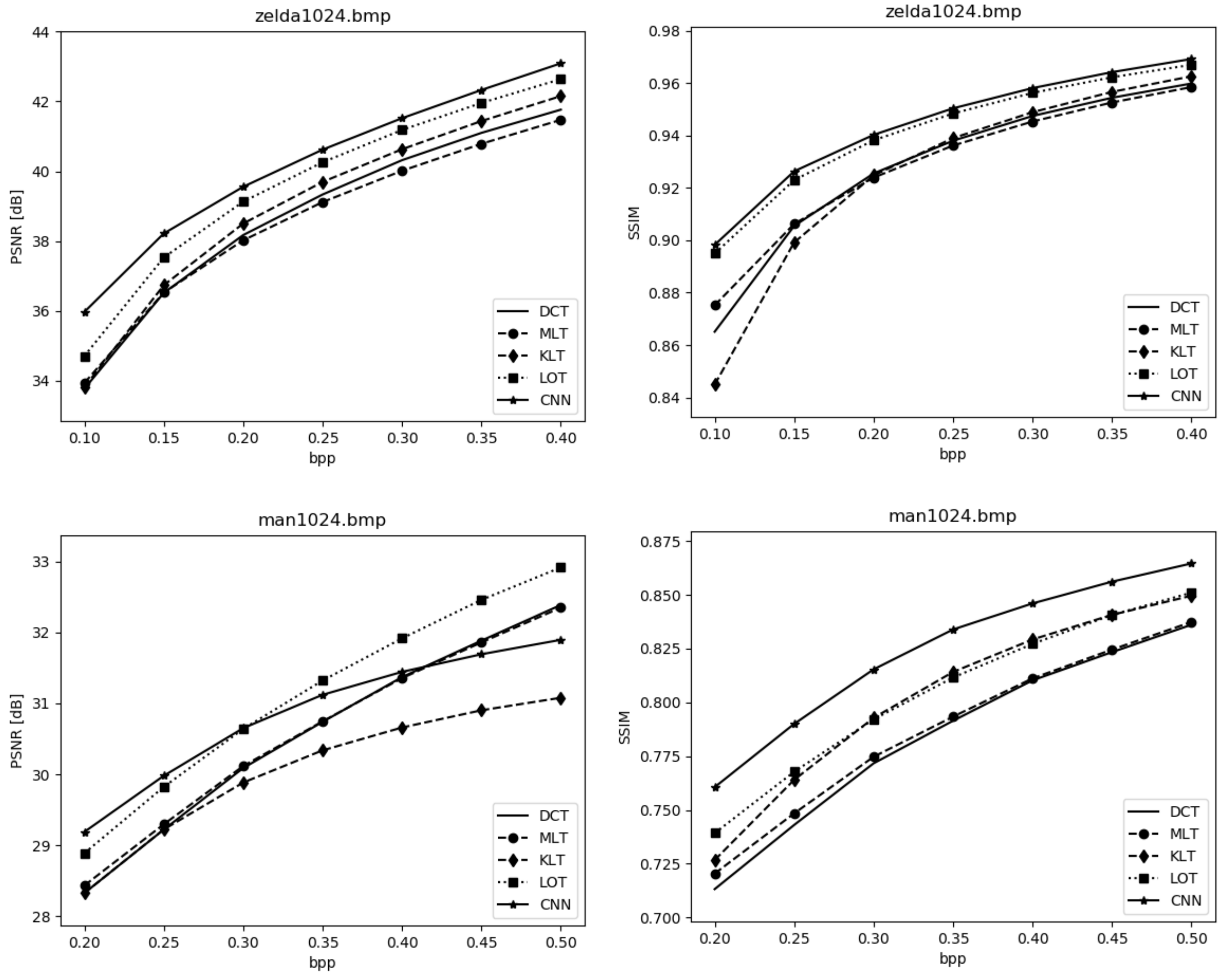
Fig. 4b. Energy distribution for 'lena1024' image after application of KLT.

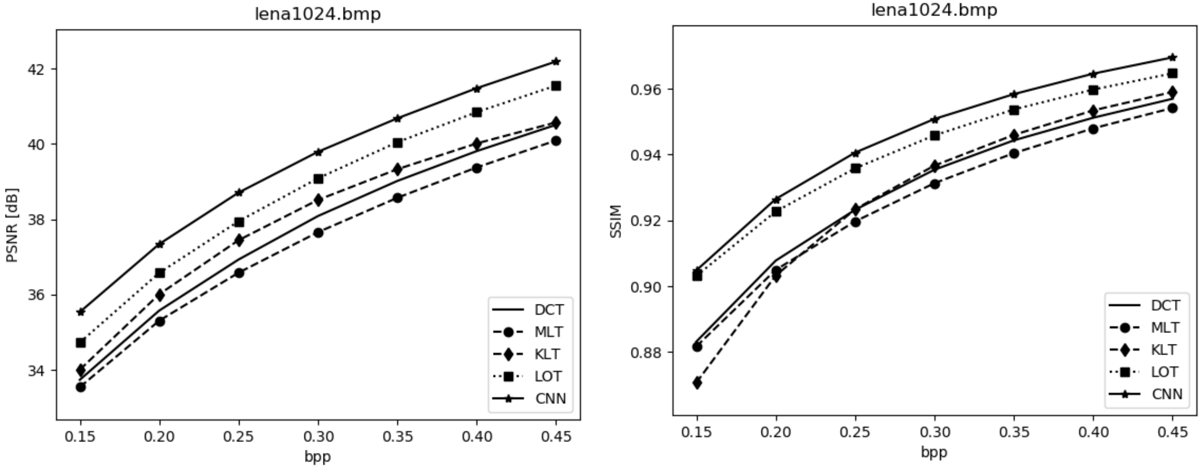
Quality Measures Results for Individual Test Images

Fig. 5. Test images - 'lena1024', 'zelda1024' and 'man1024'.

lena1024.bmp zelda1024.bmp man1024.bmp

Fig. 6. PSNR and SSIM for reconstructed test images.



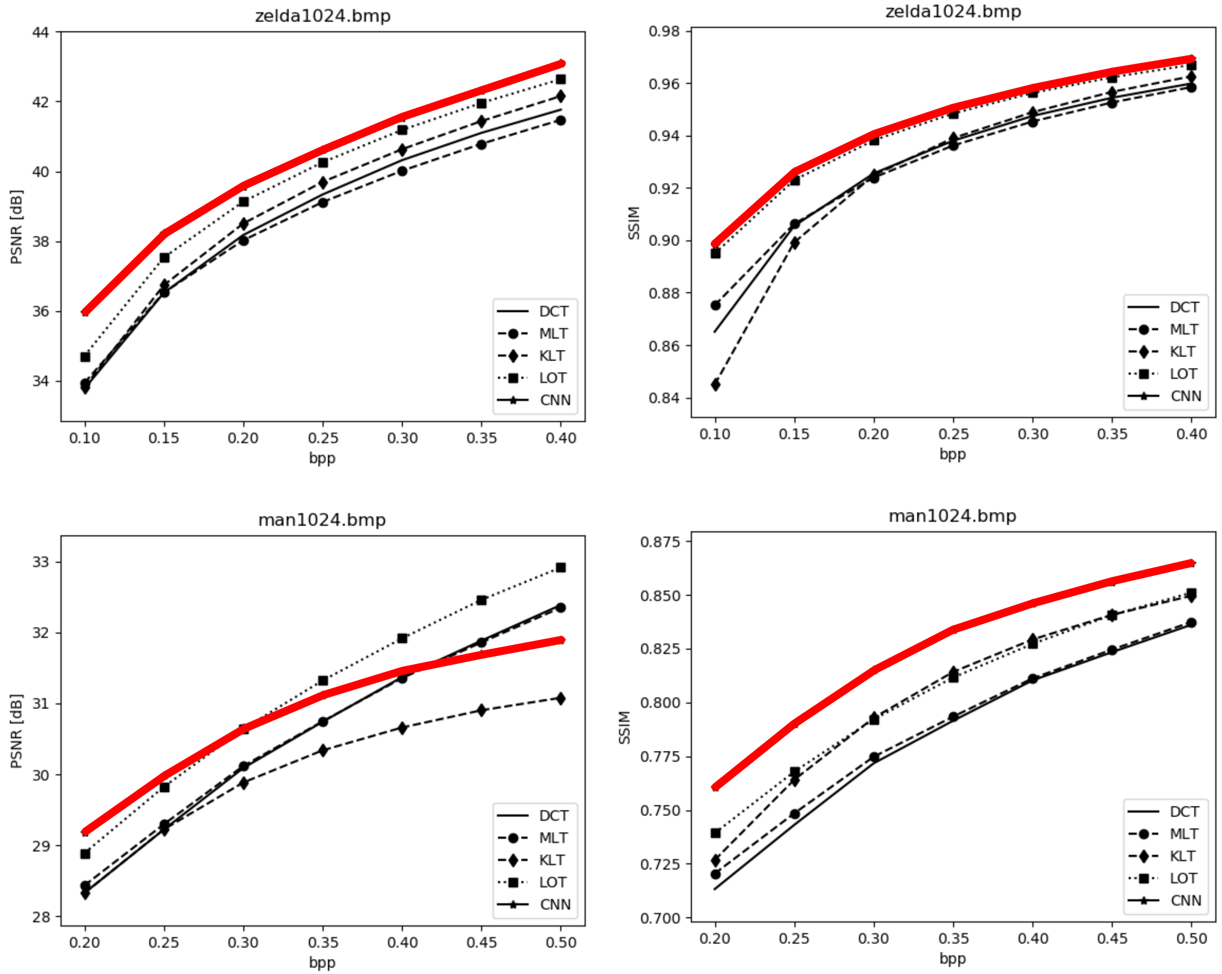


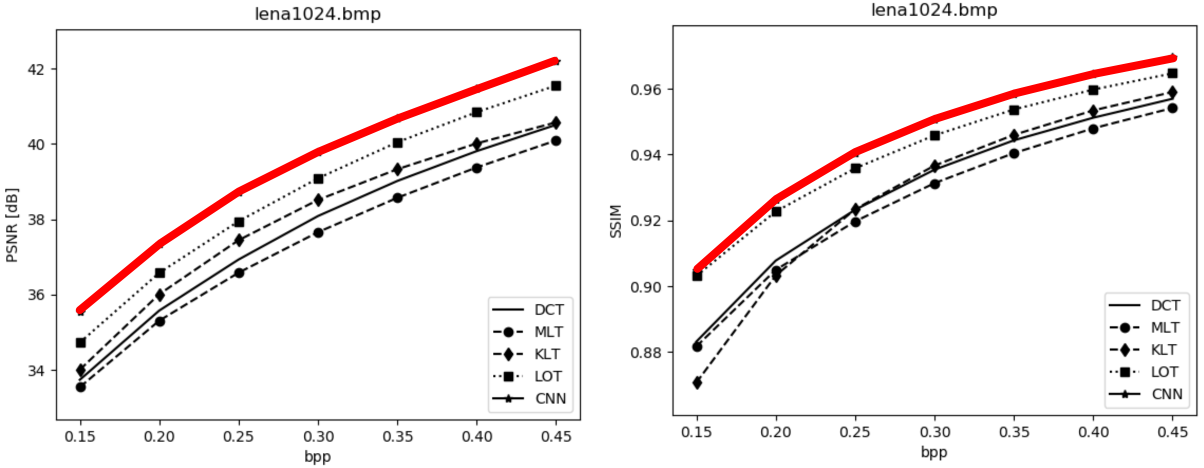
Quality Measures Results for Individual Test Images

Fig. 5. Test images - 'lena1024', 'zelda1024' and 'man1024'.

lena1024.bmp | zelda1024.bmp | man1024.bmp

Fig. 6. PSNR and SSIM for reconstructed test images.





CNN's Tiling Effect Reduction Comparison

Fig. 7. Tiling effect reduction of CNN and other transforms based JPEG image compression scheme for 'lena1024' image at 0.15 *bpp*.

Original fragment | CNN | DCT | KLT | LOT | MLT

Fig. 8. Tiling effect reduction of CNN and other transforms based JPEG image compression scheme for 'zelda1024' image at 0.15 *bpp*.

Original fragment | CNN | DCT | KLT | LOT | MLT

Averaged Results for *FFHQ* Database Images

Tab. 1. The averaged results for 100 test images from *FFHQ* database [1], CNN trained on 20 random images from [1].

Peak Signal to Noise Ratio (PSNR) [dB]						Structural Similarity Index Measure (SSIM)					
Method	bits per pixel [bpp]					Method	bits per pixel [bpp]				
	0.15	0.20	0.25	0.30	0.35		0.15	0.20	0.25	0.30	0.35
DCT	31.79	32.64	33.21	33.67	34.05	DCT	0.756	0.778	0.794	0.808	0.819
LOT	32.20	33.01	33.56	33.99	34.35	LOT	0.772	0.792	0.806	0.818	0.829
MLT	31.78	32.62	33.18	33.63	33.10	MLT	0.758	0.779	0.795	0.808	0.819
KLT	31.90	32.81	33.37	33.74	33.10	KLT	0.756	0.784	0.801	0.813	0.822
CNN	32.56	33.31	33.81	34.14	34.38	CNN	0.784	0.803	0.816	0.825	0.831

[1] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *CoRR*, vol. abs/1812.04948, January 2018.

Averaged Results for *FFHQ* Database Images

Tab. 1. The averaged results for 100 test images from *FFHQ* database [1], CNN trained on 20 random images from [1].

Peak Signal to Noise Ratio (PSNR) [dB]						Structural Similarity Index Measure (SSIM)					
Method	bits per pixel [bpp]					Method	bits per pixel [bpp]				
	0.15	0.20	0.25	0.30	0.35		0.15	0.20	0.25	0.30	0.35
DCT	31.79	32.64	33.21	33.67	34.05	DCT	0.756	0.778	0.794	0.808	0.819
LOT	32.20	33.01	33.56	33.99	34.35	LOT	0.772	0.792	0.806	0.818	0.829
MLT	31.78	32.62	33.18	33.63	33.10	MLT	0.758	0.779	0.795	0.808	0.819
KLT	31.90	32.81	33.37	33.74	33.10	KLT	0.756	0.784	0.801	0.813	0.822
CNN	32.56	33.31	33.81	34.14	34.38	CNN	0.784	0.803	0.816	0.825	0.831

[1] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *CoRR*, vol. abs/1812.04948, January 2018.

Averaged Results for *FFHQ* Database Images

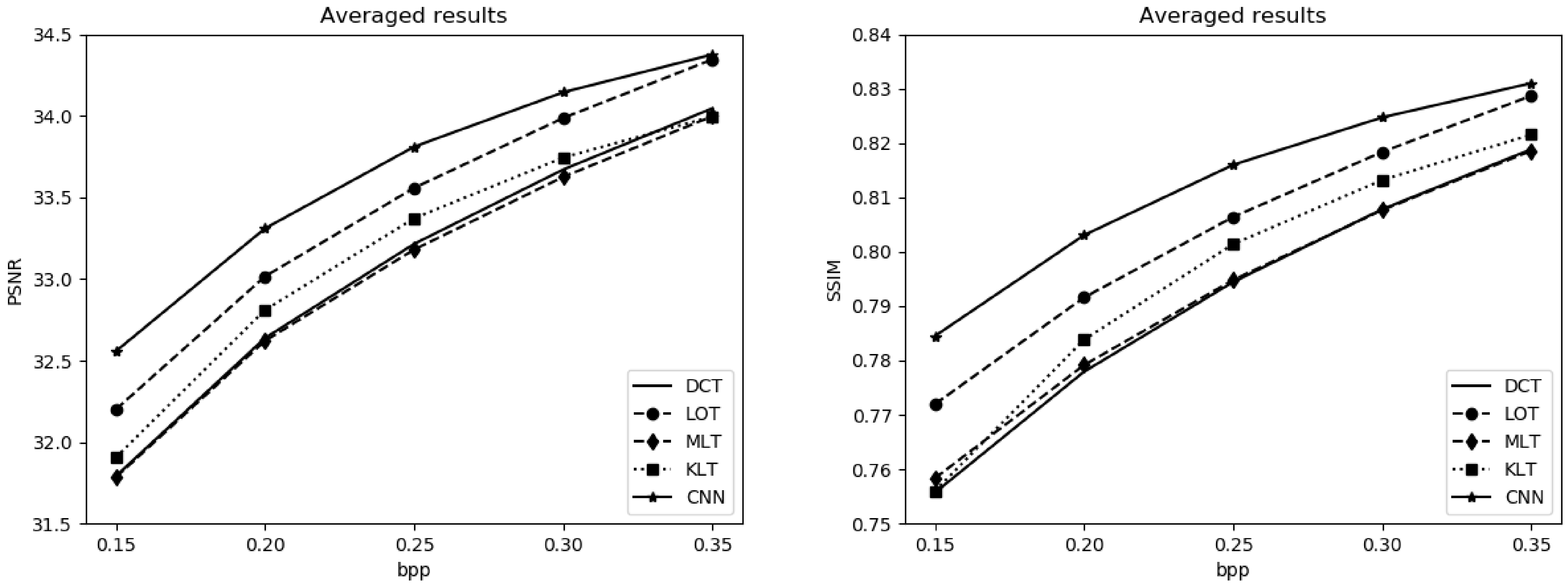


Fig. 9. The averaged results for 100 test images from *FFHQ* database, proposed CNN trained on 20 random images from the same database.

Averaged Results for *FFHQ* Database Images

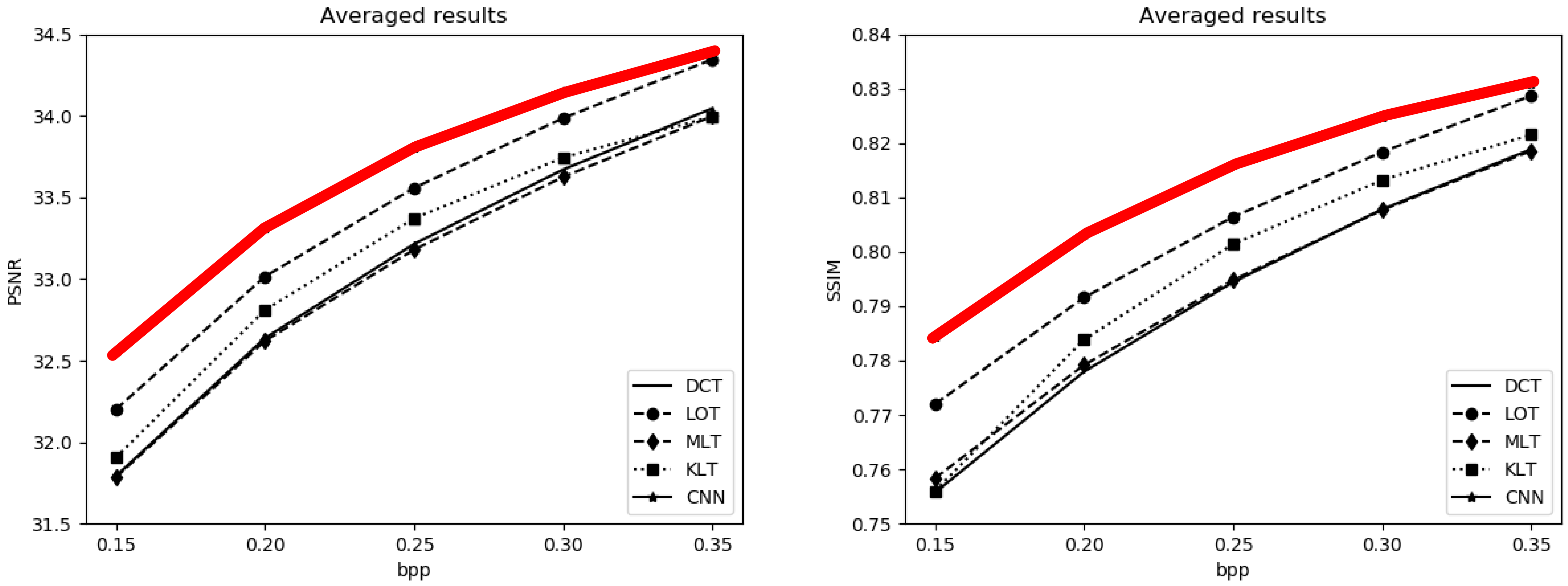


Fig. 9. The averaged results for 100 test images from *FFHQ* database, proposed CNN trained on 20 random images from the same database.

Conclusions

The main conclusions regarding the presented CNN-based JPEG compression scheme are:

- Significantly better results than all the remaining approaches for both
  - objective image quality measures PSNR and SSIM
  - subjective image quality - visible tiling effect elimination
- Simple linear two-stage structure
  - including KLT applied in the last processing stage for feature maps inter-correlation reduction
- Trained only once, applied for multiple image classes
  - no need for network's weights storage in the JPEG stream's APP merkers
- Easy incorporation into standard JPEG stream / JFIF file format
  - CNN replaces standard DCT in the JPEG's transform coding stage
  - no need for any alternation of the remaining JPEG's stream stages

Thank You For Your Attention.