Image Compression based on Neuroscience Models: Rate-Distortion performance of the Neural Code

Effrosyni Doutsi* and Panagiotis Tsakalides*†

*Foundation for Research and Technology - Hellas
100 Nikolaou Plastira str., Vassilika Vouton
Heraklion, Crete GR-70013, Greece
edoutsi@ics.forth.gr

†University of Crete
Voutes University Campus
Heraklion, Crete GR-70013, Greece
tsakalid@ics.forth.gr

Abstract

Image compression still remains one of the most challenging scientific fields as it is widely used in almost every kind of application. To design groundbreaking architectures the image processing community concentrated at first on visual perception. However, understanding how the visual system senses the semantics of the visual scene is still a big issue. Recently, a neuro-inspired compression (NICE) system was introduced, trying to encode the visual information by mimicking the firing mechanisms of neurons. The performance of this system was promising but inefficient when opposed to the JPEG standard. In this work, we aim at improving the architecture design of the NICE system. Thus, we propose 3 different architectures and we show that the quality of the reconstruction is substantially enhanced. In addition, the NICE system outperforms the JPEG standard according to the assessment of two full-reference metrics; the peak signal-to-noise ratio (PSNR) and the structure similarity index measure (SSIM), except for very low bits per pixel values.

Introduction

During the last decades, there was a significant technological revolution concerning the development of high-efficiency devices (e.g., 4K and 8K cameras, 360° cameras, drones, smartphones, tablets, etc.), scientific instruments (e.g., MRI scanners, microscopes, telescopes, satellites, etc.) and sensor-rich platforms (e.g., social networks, smart homes, self-driven cars, google glasses, etc.) characterized by the intense collection, processing and communication of massive streams of digital information. The progress of these technologies ushered in the new Big Data era, bringing not only serious challenges [1], such as the tremendous increase in volume and variety of measurements, but also a strong motivation to develop new computational infrastructures and data storage methods.

The great majority of big data concerns images and videos that can be found in different spatial and temporal resolutions depending on the electronic sensor that produces them. The spatio-temporal resolution is associated with the statistical redundancy of the signal, which is, in most cases, desirable to be eliminated before storage and/or transmission. Image (video) compression standards aim at representing an image (video) signal with the smallest possible number of bits without any significant loss of information, thereby speeding up transmission and minimizing storage requirements [2]. However, all these systems suffer from several drawbacks that have yet to be addressed such as (a) the "blind" processing of the visual content which ignores the semantics of the input and (b) the embedded mechanisms (e.g. macroblock splitting, zero-padding, etc.) that cause unpleasing and distracting artifacts to the human perception, especially for low bits per pixel values.

The latest signal processing trends are highly motivated by the computational efficiency of the brain. The brain is an intelligent "device" capable of processing and manipulating huge amount of information at ultra high speed. Neuroscientists try to understand and model some functions of the brain concentrating on the visual system. The fundamental role of each layer along the visual pathway is hidden in the structure and connectivity between neurons [3]. Neuroscience models that approximate this connectivity have been widely used in (a) machine learning, where convolutional and spiking neural networks [4] are motivated by the real neural networks, (b) image processing, where weighted difference of Gaussians (WDoG) [5] or Gabor filters [6] approximate the retina and the V1 transforms, respectively, and (c) compression algorithms, where bio-inspired architectures [7–9] and dynamic quantizers [10] have been motivated by the spike generation mechanisms of neurons.

We have recently introduced a neuro-inspired compression mechanism of RGB images, also known as NICE [11]. NICE consists of two neuroscience models; the Retina-Inspired Filter (RIF) [5] and the neuro-inspired quantizer [12] (see Fig. 1) that transform and compress an image into spikes mimicking the firing of neurons. Using these spikes, NICE enables to retrieve the original image with some loss. The performance of this system was really promising since three interesting conclusions were drawn; the NICE algorithm (a) avoids the block-effect artifacts, (b) preserves semantics of the visual scene, and (c) outperforms the JPEG standard according to no-reference image quality assessment (IQA) metrics such as BRISQUE [13]. However, the performance of NICE was unable to compete JPEG with respect to well-known full-reference IQA metrics like the peak signal-to-noise ratio (PSNR) and the structure similarity index measure (SSIM). The purpose of this work is to improve the design of NICE architecture to be more consistent to the state-of-the-art compression principles. For that reason, we propose 3 different implementation schema that substantially improve the rate-distortion performance of NICE which finally outperforms JPEG except for very low bits per pixel values.

Basics of Neuro-Inspired Compression

This section briefly describes the main principles of the retina-inspired filter and the neuro-inspired quantizer which are the components of the proposed image compression system.

Retina-inspired Filter

The retina-inspired filter (RIF) [5], which is motivated by retina-related tools [14], has been proposed to approximate precisely the dynamic properties of the outer plexiform layer (OPL) of retina. OPL is responsible for capturing and transforming the visual stimulus into electrical current. The RIF is a bunch of time-varying difference of Gaussian filters, the so-called weighted difference of Gaussians (WDoGs), defined by

$$\phi(x,t) = a(t)G_c(x) - b(t)G_s(x) \ , \tag{1}$$

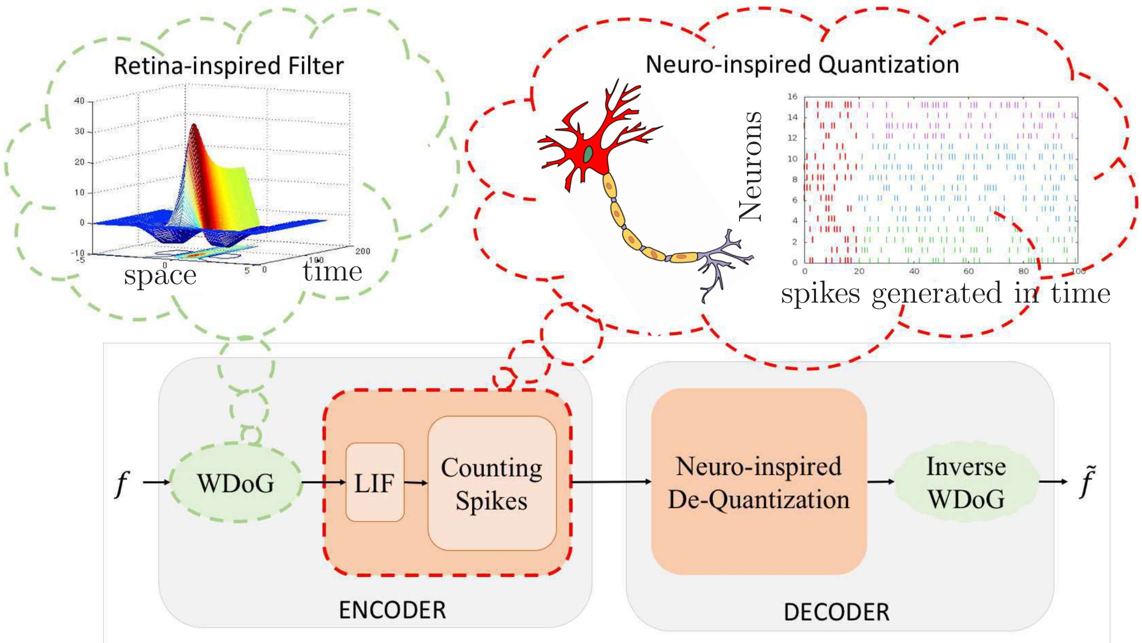

Figure 1: NICE Architecture. This system consists of two image processing tools; the retina-inspired filter and the neuro-inspired quantizer, which are designed according to the structure and functions of the visual system. The retina-inspired filter is a Difference of Gaussians that evolves time (left cloud). The neuro-inspired quantization approximates the spike generation mechanisms of neurons (right cloud).

where $G_c(x)$, $G_s(x)$ are the center and surround Gaussian filter, respectively, and $a(t)$, $b(t)$ are two time-varying weights. Assuming a uniform temporal sampling, $t_1, \ldots, t_m$, a different DoG filter $\phi_j(x) = \phi(x, t_j)$ is defined for each time instance t_j, $j = 1, \ldots, m$. Let a temporally constant input defined by

$$f(x, t_j) = f(x)\mathbb{1}_{[0,T]}(t_j)\ ,\ j = 1, \ldots, m\ , \tag{2}$$

where $f(x) = \{f(x_1), \ldots, f(x_n)\}$ is the input signal that consists of n pixels, $\mathbb{1}$ is the indicator function that equals 1, if $0 \leq t \leq T$, and 0 otherwise, and T is the time for which the signal is flashed. Then, as it is proven in [5], for each time instance the RIF yields

$$A_j(x) = A(x, t_j) = \phi_j(x) \overset{x}{*} f(x)\ ,\ j = 1, \ldots, m\ , \tag{3}$$

where $A_j(x)$ is the transformed signal at time j and $\overset{x}{*}$ denotes a spatial convolution. In addition, the RIF has been proven to be an invertible transform according to the frame theory [15], meaning that it perfectly reconstructs the input signal.

Neuro-inspired Quantization

The electrical properties of a neuron have been approximated by a parallel RC circuit (see Fig. 2) which is known as leaky integrate-and-fire model [16]. Lets suppose that a neuron is injected with some current $I(t)$ coming from its neighbors. Then, the capacitor C describes the ability of a neuron to collect and store the input current, while the resistor R represents that a neuron tries to oppose the current flow through its membrane. If the injected current is strong enough, the membrane capacitance of the cell is saturated and the neuron fires a spike, otherwise it remains silent. The sequence of spikes (hereafter refers to as neural code) is very compact and with high informative value of the input. Under the assumption of a constant input current during a period of time, $I(t) = I\mathbf{1}_{[0\leq t\leq T]}(t)$ the LIF model displace the current magnitude to its corresponding firing time (also knows as spike delay),

$$d = \begin{cases} +\infty, & RI < \theta, \\ h(RI;\theta) = -\tau_m \ln\left[1 - \dfrac{\theta}{RI}\right], & RI > \theta, \end{cases} \tag{4}$$

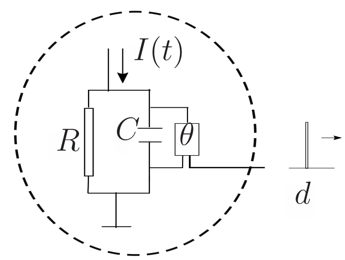

Figure 2: A neuron as a parallel RC circuit.

where R is the resistance, C is the capacitor of the electrical circuit, $\tau_m = RC$ is the leaky integrator term and θ is the membrane threshold of the neuron.

In was proposed in [12] that the intensity range of an image can be compressed if we assume that each pixel intensity is a temporally constant input current, compute the spike delay d according to the LIF model and count the number of spikes N within the observation window T as follows,

$$N = \begin{cases} 0, & d > T, \\ \left\lfloor \dfrac{T}{d} \right\rfloor, & d \leq T\ . \end{cases} \tag{5}$$

Doing so, it is possible to approximate the spike arrival delays $\tilde{d} = T/N$ and reconstruct the best possible values,

$$\tilde{I} = \begin{cases} 0, & N = 0, \\ h^{-1}(\tilde{d};\theta) = h^{-1}\left(\dfrac{T}{N};\theta\right), & N \neq 0\ . \end{cases} \tag{6}$$

Design of the Proposed Architecture

The goal of this section is to improve the design of the neuro-inspired compression architecture and optimize its rate-distortion performance. The neuro-inspired compression was first introduced in [11] in order to reduce the redundancy of RGB images. The image was first decomposed in its respective channels (red, green, and blue). The retina-inspired filter $\phi(x,t)$ was applied to each color channel f separately, generating a pile of transformed images $A(x,t)$, each one with a different frequence content. It is worth mentioning that due to the shape of the retina-inspired filter the intensity values of the output layers contain also negative values. For that reason, the authors decided in [11] to feed the neuro-inspired quantizer with the absolute values of the filtered signal reducing the range of the values as illustrated in Fig. 4(a). However, processing the absolute values changes the distribution of the input signal resulting in a strong color degradation of the reconstructed image since most of the transformed pixel intensities are too weak to generate a sequence of spikes.

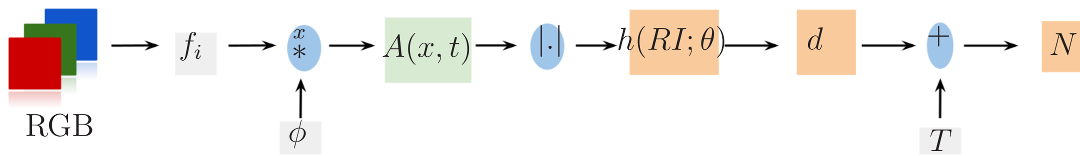

Figure 3: The NICE architecture as it was introduced in [11].

We introduce below 3 difference ways to improve the performance of the initial NICE architecture:

Architecture No 1

The first and the simplest way to improve the initially proposed architecture is to use an amplitude shifting with respect to the minimum value of the the full range of the filtered intensities $A(x,t) = A(x,t) + |\min(A(x,t))|$ as illustrated in Fig. 4(b). As a result, the histogram of the filtered signal will remain the same and the neuro-inspired quantizer will discard the redundancy more accurately.

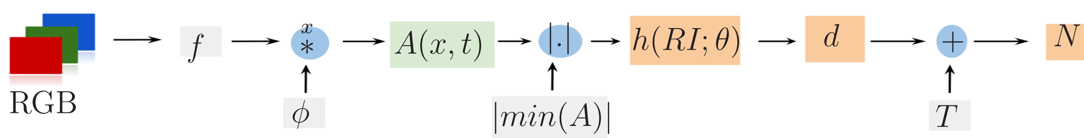

Figure 4: NICE Architecture No 1.

Architecture No 2

Although, architecture No 1 improves the performance of NICE system, the rate-distortion curve remains below the one of JPEG. For that reason, we decided to introduce a second case study with a linearly transform of the RGB color mode of the input signal into the YCbCr color mode (see Fig. 5). The YCbCr color map is widely used in image compression since it allows to separate the luminance and chrominance components and treat them in a different way. The benefit of this process is related to the visual perception which has been proven to be more sensitive with respect to the brightness than the color of a visual scene.

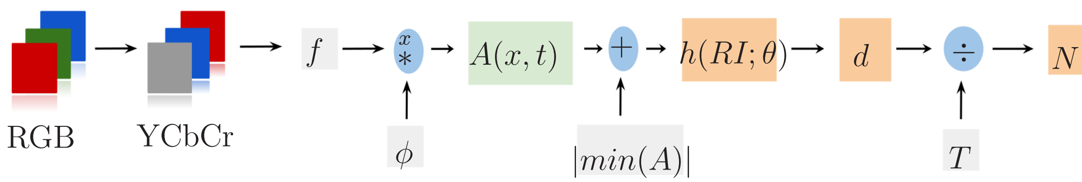

Figure 5: NICE Architecture No 2.

Architecture No 3

Although, NICE is highly motivated by neuroscience models we believe that a great breakthrough could be only succeeded if it is combined with conventional image processing and compression tools. As a result, the last schema we propose is based on the 4:2:0 sampling format (also referred as YV12) where the chrominance components, Cb and Cr, have half the horizontal and vertical resolution of Y [17]. According to the sampling formal the luminance will generate higher number of spikes than the chrominance (see Fig. 6).

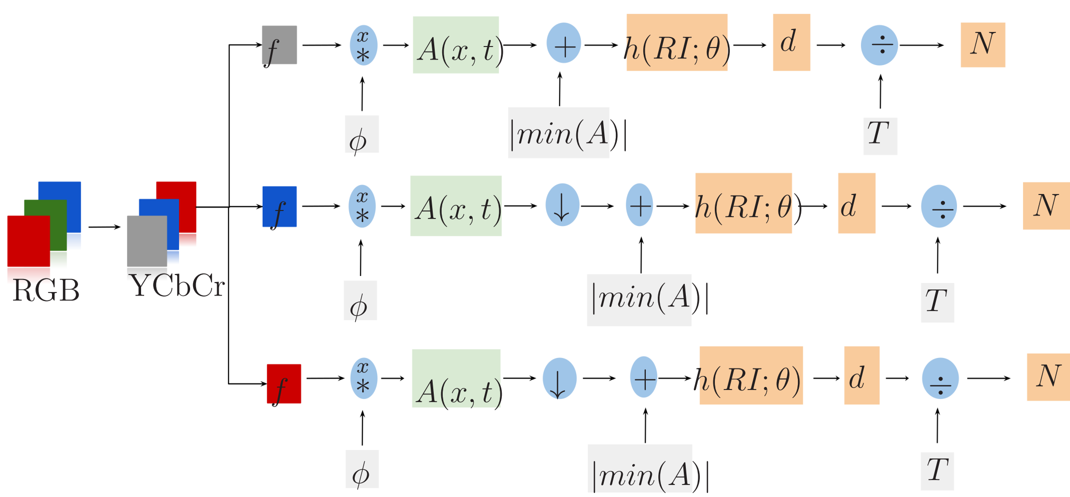

Figure 6: NICE Architecture No 3.

Experimental Results

For our experiments we used 100 color images of a size 230×230 that were randomly selected from ImageNet database [18]. The set of the filter parameters are explicitly given in [5]. We assume that the input image is constant during some time $T = 150$ ms as a result the RIF will generate 150 decomposition layers. Each layer is fed to the neuro-inspired quantizer. All the fixed parameters of the quantizer and their values are listed here $T = 150$ ms, $R = 1000\ \Omega$ and $C = 1$ F. The threshold was selected as a percentage of the total range of the values. The higher the percentage is the smaller the amount of intensities that will fire some spikes.

Rate Calculation

To evaluate the performance of our proposed NICE compression system we employ the Shannon entropy, H, for calculating the number of bits per pixel,

$$H = \frac{1}{m} \sum_{j=1}^{m} H_j \ , \tag{7}$$

$$H_j = - \sum_{k=1}^{N_j} P(k) \log_2 P(k) \ , \ j = 1, \ldots, m \ , \tag{8}$$

where P is the probability mass function of the number of spikes k that correspond to each layer. Throughout this paper, the entropy is given in bits per pixel.

Quality Metrics

Two famous full reference IQA metrics are used to evaluate the quality of the reconstructed image; the peak signal-to-noise ratio (PSNR) (10) and the structural similarity index (SSIM). For an RGB image, the PSNR is given by

$$\text{PSNR}(f, \tilde{f}) = \frac{1}{3} \left(\text{PSNR}_R + \text{PSNR}_G + \text{PSNR}_B \right) \ , \tag{9}$$

where the PSNR of each color channel $c \in \{R, G, B\}$, is defined by

$$\text{PSNR}_c = 10 \log_{10} \frac{255^2}{\text{MSE}(f_c, \tilde{f}_c)}, \tag{10}$$

$$\text{MSE}(f_c, \tilde{f}_c) = \frac{1}{n} \sum_{i=1}^{n} \| f_c(x_i) - \tilde{f}_c(x_i) \|^2 \ . \tag{11}$$

Since the proposed architecture is based on neuroscience models, we also employ the SSIM (12), which is considered to be a more accurate visual perception quality metric, defined by

$$\text{SSIM}(f, \tilde{f}) = \left(\frac{2\mu_f \mu_{\tilde{f}} + c_1}{\mu_f^2 + \mu_{\tilde{f}}^2 + c_1} \right) \left(\frac{2\sigma_f \sigma_{\tilde{f}} + c_2}{\sigma_f^2 + \sigma_{\tilde{f}}^2 + c_2} \right) \left(\frac{\sigma_{f,\tilde{f}} + c_3}{\sigma_f \sigma_{\tilde{f}} + c_3} \right), \tag{12}$$

where μ_f and $\mu_{\tilde{f}}$ denote the average of f and $\tilde{f}$, respectively, σ_f^2 and $\sigma_{\tilde{f}}^2$ are the variances of f and $\tilde{f}$, and $\sigma_{f,\tilde{f}}$ is the covariance of f and $\tilde{f}$. The constants $c_1 = k_1 L^2$, $c_2 = k_2 L^2$ and $c_3 = c_2/2$ are positive numbers used to stabilize the division with a weak denominator, where L is the dynamic range of the pixel values and $k_1 = 0.01$, $k_2 = 0.03$.

A fair comparison to JPEG standard was achieved using the "imwrite" Matlab function that allows to generate 100 different reconstruction qualities. We calculated

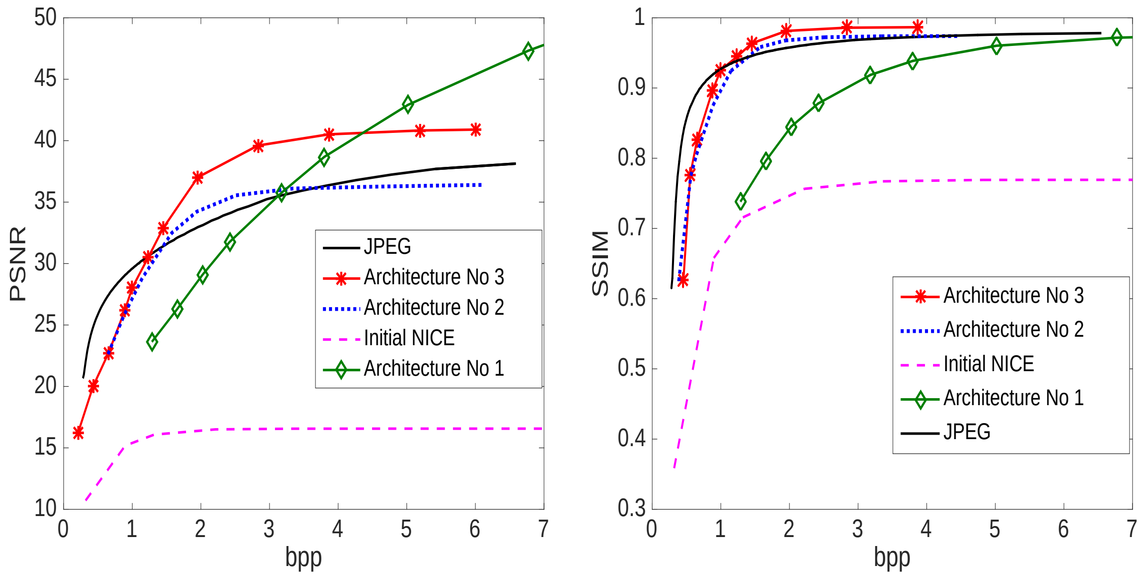

Figure 7: Comparison of the rate-distortion performance between the all the different NICE architectures and the JPEG standard.

the bits per pixel according to the size each image required to be stored on the disk driver while the quality was measured by PSNR and SSIM. Figure 7 shows the average rate-distortion performance of the whole dataset. It compares the performance of the proposed NICE architectures against NICE. It is worth mentioning that the YCbCr mode is much more efficient than the RGB color map. In addition, the PSNR metric shows that for very low bits per pixel values the JPEG standards outperforms NICE. However, NICE obtains better reconstruction quality for values greater than 1bpp.

Conclusion

In this work, we studied 3 different scenarios of how to improve the performance of the neuro-inspired compression system. According to the experimental results, the reconstruction quality was remarkably enhanced when the RGB input image was transformed to the YCbCr color map. In addition, according to the PSNR and the SSIM metrics the proposed NICE system outperforms the JPEG standard for > 1 bit per pixel values.

References

[1] Ishwarappa and J. Anuradha, "A brief introduction on big data 5vs characteristics and hadoop technology," *Procedia Computer Science*, vol. 48, pp. 319 – 324, 2015, International Conference on Computer, Communication and Convergence (ICCC 2015).

[2] Jerry D. Gibson and Al Bovik, Eds., *Handbook of Image and Video Processing*, Academic Press, Inc., Orlando, FL, USA, 1st edition, 2000.

[3] D. Field and B. A. Olshausen, "Vision and the Coding of Natural Images - The human brain may hold the secrets to the best image-compression algortihms," *American Scientist*, vol. 88, pp. 238–245, 2000.

[4] G. C. Cardarilli, A. Cristini, L. D. Nunzio, M. Re, M. Salerno, and G. Susi, "Spiking neural networks based on LIF with latency: Simulation and synchronization effects," *Asilomar Conference on Signals, Systems and Computers*, 2013.

[5] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired Filter," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3484–3499, 2018.

[6] D. H. Hubel and T. N. Wiesel, "Receptive Fields of Single Neurones in the Cats Striate Cortex," *J. Physiol.*, vol. 148, pp. 574–591, 1959.

[7] S. J. Thorpe, A. Delorme, and R. VanRullen, "Spike-based strategies for rapid processing," *Neural Networks*, vol. 14, no. 6-7, pp. 715–725, 2001.

[8] A. Lazar and E. Pnevmatikakis, "Video Time Encoding Machines," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 461–473, 2011.

[9] K Masmoudi, M. Antonini, and P. Kornprobst, "Exact Reconstruction of the Rank Order Coding using Frames Theory," 2011.

[10] E. Doutsi, L. Fillatre, M. Antonini, and P. Tsakalides, "Dynamic Quantization using Spike Generation Mechanisms," hal-02351180, 2019.

[11] E. Doutsi, G. Tzagkarakis, and P. Tsakalides, "Neuro-inspired Compression of RGB Images," 2019.

[12] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Neuro-inspired Quantization," in *IEEE International Conference on Image Processing (ICIP)*, Athens, 2018.

[13] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec 2012.

[14] B. Cessac, P. Kornprobst, S. Kraria, H. Nasser, D. Pamplona, G. Portelli, and T. Viville, "Pranas: A new platform for retinal analysis and simulation," *Frontiers in Neuroinformatics*, vol. 11, pp. 49, 09 2017.

[15] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired filtering for dynamic image coding," *IEEE International Conference in Image Processing (ICIP)*, pp. 3505–3509, 2015.

[16] W. Gerstner, *A framework for spiking neuron models: The spike response model*, vol. 4, North-Holland, 2001.

[17] Iain E. Richardson, *The H.264 Advanced Video Compression Standard*, Wiley Publishing, 2nd edition, 2010.

[18] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.