Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images

Farhad Ghazvinian Zanjani
Svitlana Zinger
Bastian Piepers
Saeed Mahmoudpour
Peter Schelkens
Peter H. N. de With

SPIE.

Farhad Ghazvinian Zanjani, Svitlana Zinger, Bastian Piepers, Saeed Mahmoudpour, Peter Schelkens, Peter H. N. de With, "Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images," *J. Med. Imag.* **6**(2), 027501 (2019), doi: 10.1117/1.JMI.6.2.027501.

Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images

Farhad Ghazvinian Zanjani,[a,*] Svitlana Zinger,[a] Bastian Piepers,[b] Saeed Mahmoudpour,[c,d] Peter Schelkens,[c,d] and Peter H. N. de With[a]
[a]Eindhoven University of Technology, SPS-VCA, Eindhoven, The Netherlands
[b]Barco NV, Healthcare Division, Kortrijk, Belgium
[c]Vrije Universiteit Brussel, Department of Electronics and Informatics, Brussels, Belgium
[d]IMEC, Leuven, Belgium

Abstract. The availability of massive amounts of data in histopathological whole-slide images (WSIs) has enabled the application of deep learning models and especially convolutional neural networks (CNNs), which have shown a high potential for improvement in cancer diagnosis. However, storage and transmission of large amounts of data such as gigapixel histopathological WSIs are challenging. Exploiting lossy compression algorithms for medical images is controversial but, as long as the clinical diagnosis is not affected, is acceptable. We study the impact of JPEG 2000 compression on our proposed CNN-based algorithm, which has produced performance comparable to that of pathologists and which was ranked second place in the CAMELYON17 challenge. Detecting tumor metastases in hematoxylin and eosin-stained tissue sections of breast lymph nodes is evaluated and compared with the pathologists' diagnoses in three different experimental setups. Our experiments show that the CNN model is robust against compression ratios up to 24:1 when it is trained on uncompressed high-quality images. We demonstrate that a model trained on lower quality images—i.e., lossy compressed images—depicts a classification performance that is significantly improved for the corresponding compression ratio. Moreover, it is also observed that the model performs equally well on all higher-quality images. These properties will help to design cloud-based computer-aided diagnosis (CAD) systems, e.g., telemedicine that employ deep CNN models that are more robust to image quality variations due to compression required to address data storage and transmission constraints. However, the results presented are specific to the CAD system and application described, and further work is needed to examine whether they generalize to other systems and applications. [DOI: 10.1117/1.JMI.6.2.027501]

Keywords: convolutional neural networks; tumor detection; JPEG 2000 compression; image quality; digital pathology.

Paper 18191RR received Sep. 2, 2018; accepted for publication Apr. 1, 2019; published online Apr. 24, 2019.

1 Introduction

Computational histopathology involves computer-aided diagnosis (CAD) for microscopic analysis of stained histopathological whole-slide images (WSIs) to study the presence, localization, or grading of diseases. Emerging new scanners for digital microscopic imaging make it possible to acquire gigapixel histopathological images at a large scale.[1,2] These large-scale digital datasets make digital pathology a perfect use case to deploy data-greedy, deep-learning models. The availability of these massive amounts of data in combination with recent advances in artificial intelligence, based on state-of-the-art deep-learning models and more specifically convolutional neural networks (CNNs), results in a situation where for many clinical image-analysis tasks, computational pathology solutions have a comparable performance to that of humans.[3] For example in pathology, recent deep learning-based techniques are comparable or even outperform humans in detecting and localizing breast cancer metastases in lymph node WSIs.[4]

Although increasing the number of image samples boosts the performance of a deep CNN by better learning of the image-content diversity,[5] the intrinsic image quality of the used samples will also impact the CNN's performance. Furthermore, dealing with a large database for storage and the associated transmission for cloud-based computing is challenging and for the design of a CAD system even critical. For example, working on big data in the cloud requires reconciling two contradictory design principles. On the one hand, cloud computing is based on the concepts of consolidation and resource pooling, while on the other hand, big data systems (such as Hadoop) are built on the shared nothing principle, where each node is independent and self-sufficient.[6] These issues are more crucial in telemedicine and cloud-based computation, regarding privacy and security issues. For example, in the CAMELYON17 challenge,[2,7] which is an international competition on designing the best CAD algorithm for automated breast cancer metastases detection, about 1000 histopathological WSIs (>3 terabyte image data) have been made publicly available. Downloading the whole dataset on a local machine for training a CAD model was cumbersome and requires a significant amount of time and network bandwidth. Given the large size of WSIs, the use of compression algorithms is a very appealing solution. Particularly, lossy compression that can support larger compression ratios is interesting. Luckily, it is generally not prohibited by the main regulatory bodies in the European Union,

*Address all correspondence to Farhad Ghazvinian Zanjani, E-mail: f.ghazvinian.zanjani@tue.nl

United States, Canada, and Australia, provided that it does not impair the diagnostic quality and does not cause new risks compared with conventional practice.[8] Hence, it is important to define a strategy or protocol for an efficient parameterization of the deployed compression techniques to yield a high compression ratio without jeopardizing the classification performance. The issue of higher compression ratio with lower encoding time has been recognized as well in recent efforts for creating the DICOM standard in the field of digital pathology.[9]

For studying the impact of lossy compression on the diagnostic performance of human experts, several studies have been reported.[8,10–16] Mostly, they reported that the human visual perception is to some extent robust against image quality degradation. However, there is not a generally accepted tolerance level with respect to the diagnostic accuracy. In addition, because clinical evaluations can be subjective and have a bias regarding the task at hand and the skill of experts, different studies have suggested different compression ratios corresponding to the addressed clinical task. For example, Kalinski et al.[8] reported that the impact of a JPEG 2000 compression ratio up to 20 did not show significant influence on the detection of Helicobacter pylori gastritis in gastric histopathological images, performed by three pathologists. In another work by Krupinski et al.,[11] by involving six pathologists, a compression factor of up to 32 did not cause noticeable difference in distinguishing benign from malignant cancer in breast tissue. At the same time, they reported that increasing the compression ratio to 64:1 affected the diagnostic performance significantly. Marcelo et al.[13] studied the accuracy of diagnosis and confidence level of 10 pathologists between noncompressed and JPEG compressed (reduced 90% in file size) pathology images. They reported no statistically significant difference in diagnostic accuracy at 95% confidence interval (CI). Johnson et al.[14] reported a threshold of about 13:1 compression ratio for human observer to discriminant JPEG 2000 compressed versus uncompressed breast histopathological images. In work of Pantanowitz et al.,[12] a compression ratio of 200:1 was reported as an acceptable threshold for measuring the HER2 score in immunohistochemical images of breast carcinoma evaluated by a conventional image processing algorithm.[17] Lopez et al. used a cell counting CAD system as a reference for statistical evaluation of cell counting error in the uncompressed and JPEG-compressed histopathological images. They involved three different compression ratios of 3, 23, and 46 and concluded that increasing the compression ratio deteriorates the performance of cell counting in images. They concluded that the significant factors influencing the classification-performance degradation of a CAD system are the compression ratio and the intrinsic image complexity. According to their study, a more complex image is known as an image with a higher number of nuclei.

Although all the above works study the impact of lossy compression on the diagnostic performance, they do not involve a complex model such as a deep CNN as a model observer. Furthermore, their experiments with a CAD observer are limited to training on the high-quality input data and evaluating on both high- and low-quality image data, while they have not considered the performance of a model observer that has been adapted (trained) on low-quality input data.

JPEG 2000[18] was introduced as a follow-up standard for JPEG (ISO/IEC 10918-1 to ITU-T Rec. T.81) bringing improved rate-distortion performance and additional functionality, such as resolution and quality scalability.[19] One of the main differences between JPEG 2000 and the JPEG algorithm is the exploitation of the discrete wavelet transform instead of a block-based discrete cosine transform. In terms of visual artifacts, JPEG 2000 produces "ringing" and "blocking" artifacts at high compression ratios, whereas JPEG produces both particularly blocking artifacts.[20] Nonetheless, although their performance for higher bitrates is comparable at mid and lower bitrates, JPEG 2000 outperforms JPEG in terms of rate-distortion performance.[21] With the use of the JPEG 2000 algorithm, it also becomes possible to store different parts of the same picture with different qualities, which makes it attractive for the compression of WSIs,[22] since ∼80% of a WSI area contains an empty (white) background region[23] that does not contain any tissue. Helin et al.[24] showed that applying a very high degree of JPEG 2000 compression on the background part of WSIs and applying a conventional amount of compression (e.g., 35:1) on the tissue-containing part results in a high overall compression ratio. Compression gains of up to a factor 3 are reported compared to classical, nonadaptive compression with JPEG 2000.

A number of studies have assessed the impact of the quality of natural images in terms of compression on the performance of a deep CNN.[25,26] In work of Dodge and Karam,[25] a VGG-16 network, which has been trained on the ImageNet 2012 dataset,[27] was found resilient to JPEG and JPEG 2000 compression up to a compression factor of 10 and down to 30-dB peak signal-to-noise ratio, respectively. In similar work by Dejean-Servières et al.,[26] again an experiment on the ImageNet dataset is performed, where a CNN showed only a drop of one unit on the classification ranking for object categorization after applying the compression (with rate of 16:1). Here the classification ranking is defined by sorting (in descending order) the output probabilities of the assigned class labels, given an input image by the network. So ideally the true class should be recognized with rank one while categorizing in any lower ranking can be considered as a greater error in the classification performance.

Although these studies involve a CNN for the assessment of classification performance on compressed natural images, to the best of our knowledge, no similar study has been carried out on the histopathological images/WSIs. The outcome of such a study can be different from the obtained results for the natural images, since the histopathological image contents have a high intercomponent correlation and can be processed differently.

In this paper, we investigate the impact of the compression ratio to evaluate the performance of a deep CNN applied to JPEG 2000 compressed, histopathological WSI data. We employ a recently proposed CAD model that produces comparable performance to that of pathologists in detecting cancer metastases in breast lymph nodes. Since our CAD system exploits a learning model, we study the impact of degradation in image quality due to compression, both by varying the quality (i.e., compression ratio) of the training data in the training phase as by varying the quality of the test data during the testing phase. Such a study reveals the adaptivity of the CNN model for preserving its high performance on lower quality images when its parameters are adapted.

In the following, we first introduce the data that has been used in this study. Afterward, the CNN model is briefly explained and our experimental setup and the obtained results are detailed. Finally, the results are discussed and conclusions are drawn.

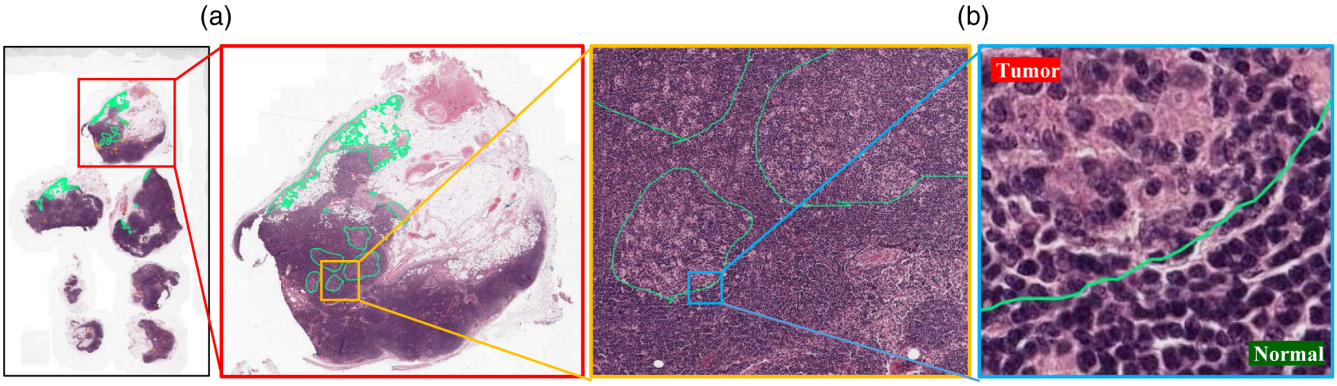

Fig. 1 An example WSI from the CAMELYON16 dataset. (a) A histopathology WSI of size 220k × 98k pixels is shown with multiple zoom levels. (b) An image patch of size 300 × 300 pixels. The metastases have been delineated by the pathologists (green contours).

2 Materials and Methods

2.1 *Dataset*

We use the CAMELYON16 dataset[4] for the experiments. This dataset, which is the preceding version of CAMELYON17, contains tumor annotations at pixel level (Fig. 1). The task in CAMELYON16 was detecting and segmenting the metastases in WSIs, while in CAMELYON17 the task was changed to categorization of each detected metastatic region into four types (i.e., grades), according to their area. This can be considered as a postprocessing stage to what was defined in CAMELYON16. In this study, we base our evaluation of the CAD performance on the task that was defined in the CAMELYON16 challenge, as it obtains more accurate quantified measures in comparison with the slide-level categorization task of CAMELYON17.

The CAMELYON16 dataset consists of WSIs having pixels acquired with a resolution equal to 0.243 μm, collected from two clinical centers in the Netherlands. Originally, the dataset was split into a training set and a testing set. The training set consists of 111 WSIs with and 159 without metastases. The testing set consists of 129 WSIs, 49 with and 80 without metastases. We removed one slide (namely tumor slide number 114) from the testing set because it does not have an exhaustive annotation for all its mestastasis regions as also mentioned by the data provider. For ground truth, the pixel-level annotation for the positive (i.e., containing tumor) WSIs was provided by a group of pathologists. The original WSIs were stored in the TIFF format that was already compressed by the JPEG compression with 80% quality and 4:2:2 Y'CbCr chroma subsampling. The WSIs are stored in a pyramidal structure with different levels of magnification. Here we use the 20× magnification level, since it has shown the highest performance for tumor detection.[28] In our study, the uncompressed high-quality data (also labeled with 1:1 ratio) refers to this dataset. More details about this dataset can be found in the paper of Bejnordi et al.[4]

2.1.1 *Data sampling*

Since involving all the regions inside a WSI is redundant and inefficient for training a CNN model, a data-sampling stage is applied, which consists of two parts: region of interest (ROI) detection and patch extraction. As mentioned earlier, about 80% of a WSI area contains empty background region,[23] which can be easily detected using a conventional image processing technique such as Otsu thresholding.[1] By detecting the empty regions of each WSI, they are ignored for further analysis by the CNN model. Because of the very large image-frame size of WSIs, directly using them as input to a CNN is impractical. A common approach is, therefore, the processing of image patches and employing the CNN as a patch classifier.[29] In a patch classification approach, the input to the CNN is a patch image with predefined dimensions and the output is the predicted class of the central pixel inside the image patch. After training the network on image patches, prediction on WSIs can be performed by sliding a window over the entire WSI and consequently predicting the central pixel of the window. Training the model on all possible extracted patches is redundant, and the population of samples between two classes would be highly imbalanced because in most cases only a small portion of the examined tissue contains tumor cells. For compensating the problem of highly imbalanced data in patch sampling, we only randomly select a limited number of negative patches, while all the extracted positive patches from the training set are used.[30] Here the negative and positive patches refer to the patches that have been labeled as normal and tumor patches, respectively.

In total, 650k patches of size 300 × 300 pixels are extracted. A patch is labeled as a positive sample (tumor) if >20% of its pixels are annotated as positive, otherwise it is labeled as a negative (normal) sample. For better training of the model, the extracted patches are augmented on the fly (during training) by applying random rotation, using multiples of 90 deg (i.e., rotation angles of {0, 90, 180, 270}). Consequently, the images are also randomly chosen to be flipped or not. Flipping may be vertically, horizontally, or in both directions. For increasing the generalization ability of the classifier against minor changes in chromatic information, we apply color augmentation on the training image patches that has become a common practice in training a deep CNN.[4,7] This leads to training a wider range of color variations compared to what typically occurs in the training set. To do so, we insert some noise into the lightness and saturation channels of the HSL color coordinates, by adding a (uniformly distributed) random value to the pixels of each patch (or subtracting a random value from them). The maximum magnitude of such an additive noise is equal to 0.25% of maximal value of the channel (e.g., 0.25 × 255).

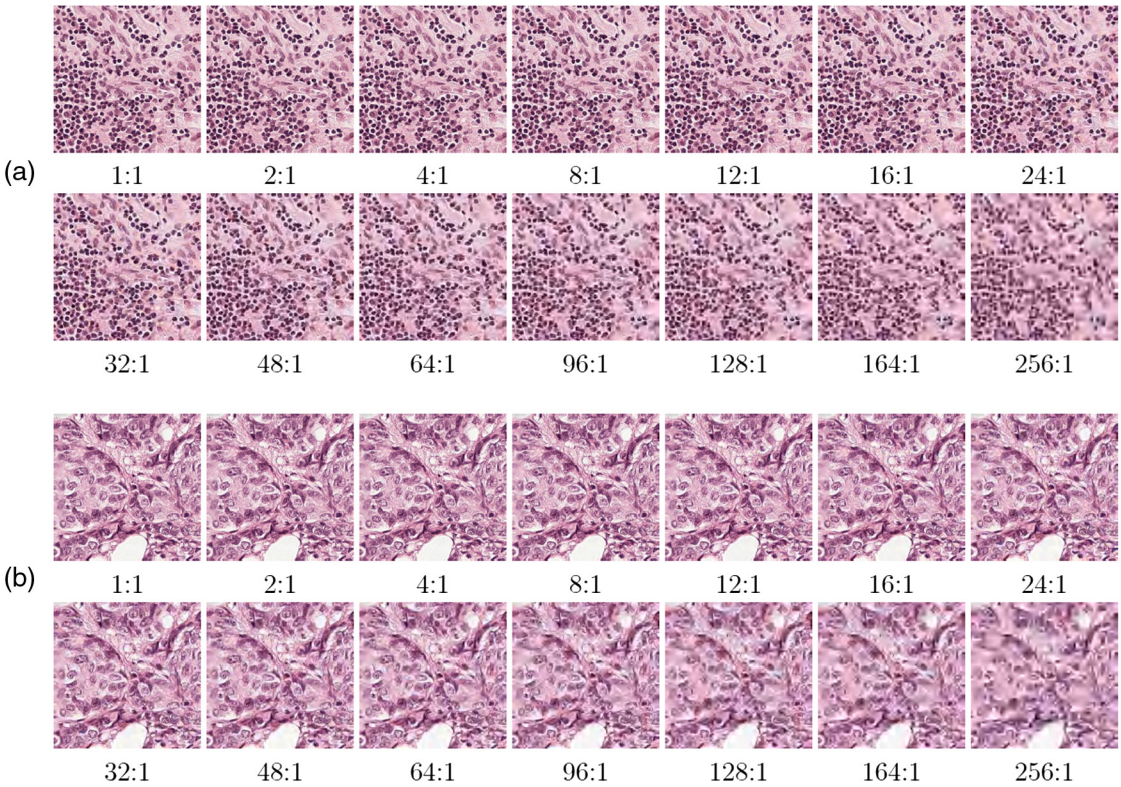

Fig 2 Examples of (a) normal and (b) tumor image patches compressed with JPEG 2000 at different compression ratios.

2.2 *JPEG 2000 and Image Quality*

The extracted patches from the data sampling stage are compressed. We deployed JPEG 2000 with 6 wavelet decomposition levels and 14 different compression ratios. Figure 2 shows a normal and a tumor patch compressed at different compression ratios.

2.3 *Automated Tumor Detection*

Our recently proposed CNN-based model[1] is adopted as an automated cancer metastases detection system in breast lymph node WSIs. This model uses the "Inception-v3" architecture,[31] a 48-layer deep CNN, as patch classifier. The input data to the model are full-color RGB image patches and its output is a 2-element vector with one-hot encoding, representing a binary classification. For speeding up the training, the parameters of CNN are initialized using the parameters, which have been trained on the ImageNet 2012 dataset.[27] The Inception-v3 architecture has shown to have a better performance for image classification with a much lower number of parameters, compared with its preceding versions, due to its convolution factorization strategy.[31] In computational pathology, this model has shown human-level performance in detecting tumor cells[28] and won second place in the CAMELYON17 challenge,[2,7] which includes the CAMELYON16 dataset used in this paper.

2.4 *Experiments and Evaluation*

2.4.1 *Experiments*

First, ROI selection and patch extraction are performed on the training and test WSIs as described earlier. About 150k positive samples were extracted from the regions with metastases in positive WSIs, according to the provided ground truth. For compensating the problem of severe class imbalance in samples, we have extracted only about 500k negative samples from the normal regions of the negative and positive WSIs. Afterward, these samples are compressed by the JPEG 2000 algorithm using 14 different compression ratios.

The impact of changing the compression ratio of JPEG 2000 on the performance of the CNN is extensively evaluated for the following three distinct scenarios.

1. The CAD system is trained on high-quality uncompressed images and is evaluated on compressed low-quality images with several compression ratios.
2. The CAD system is trained and tested on the same level of compression. For example, if the model is trained on images compressed by a factor of 32, it is also evaluated on images that are compressed by the same factor.
3. The CAD system is trained on images compressed with the maximal compression ratio that still allowed the classification performance to be above a predefined threshold (e.g., <10% drop from the maximum $F1$ score). Thereafter, it is evaluated with test images with both lower and higher compression ratio.

The first scenario is highly applicable to telemedicine and in particular telepathology,[32] where a primary diagnosis can be obtained by transmitting the compressed images to a remote CAD system. Such a remote CAD system can have already been trained on high-quality input data. The second scenario is more relevant in cloud-based computing and training,

Table 1 $F1$ scores, AUC values, and their 95% CIs for the CAD model, tested on the images with different compression ratios, when it is trained on the uncompressed (1:1) high-quality training images (scenario 1).

		JPEG 2000 compression ratios													
Metric		1:1	2:1	4:1	8:1	12:1	16:1	24:1	32:1	48:1	64:1	96:1	128:1	164:1	256:1
$F1$ score	CI upper bound	0.954	0.954	0.952	0.946	0.938	0.934	0.923	0.903	0.878	0.836	0.754	0.687	0.580	0.594
	Mean	**0.956**	**0.957**	**0.954**	**0.949**	**0.941**	**0.937**	**0.927**	**0.908**	**0.883**	**0.841**	**0.759**	**0.692**	**0.586**	**0.600**
	CI lower bound	0.959	0.959	0.957	0.952	0.944	0.940	0.931	0.913	0.887	0.845	0.764	0.698	0.592	0.605
AUC	CI upper bound	0.991	0.991	0.991	0.989	0.987	0.985	0.980	0.972	0.957	0.931	0.872	0.820	0.668	0.661
	Mean	**0.992**	**0.992**	**0.991**	**0.990**	**0.988**	**0.987**	**0.981**	**0.974**	**0.959**	**0.934**	**0.876**	**0.825**	**0.674**	**0.668**
	CI lower bound	0.993	0.992	0.992	0.991	0.989	0.988	0.983	0.976	0.961	0.936	0.879	0.829	0.680	0.675

where several pathology labs share their data to a remote server for training and evaluation. The third scenario is valid for a case where a powerful computation engine is locally available, e.g., exploiting a supercomputer in a clinical institute or large hospital, so that the transmission of high-quality images is not an issue internally, but utilizing external images from remote data sources for training still has limitations due to transmission bandwidth constraints.

2.4.2 *Evaluation method*

The performance of the binary classification between tumor and normal image patches is evaluated by reporting the $F1$ score and area under the receiver operating characteristic (ROC) curve, called AUC. Evidently, the configuration of a CAD system with a higher AUC and $F1$ score represents a better performance.

Since the discrimination threshold of the binary classification system is varied, we report the diagnostic capability of the system with a complementary measurement of the precision–recall (PR) curve. In comparison with alternative measures such as ROC curve, a PR curve can better expose the differences between algorithms, especially when highly skewed cancer detection data are studied.[33] The PR curve visualizes the performance of a classifier by ignoring the true negative samples; this property highlights well the change in classification performance when imbalanced data are processed. It is worth mentioning that even if the training set contains an equal number of patches per class, the data originally are considered imbalanced, since the area of tumor region is often smaller than the normal region in a pathology slide.

3 Results

3.1 *Scenario 1: High-Quality Training Data and Evaluation on Lower-Quality Images*

In this experiment, the CNN model is trained on uncompressed images for a fixed number of iterations equal to 10k. Afterward, we evaluate its performance on the test set, which has been compressed with 14 different compression ratios, including the original uncompressed test images (compression ratio 1:1). The obtained $F1$ score and AUC values are depicted in Table 1 and the PR curves are plotted in Fig. 3. As expected, by degrading the image quality due to increasing the compression ratio, the CNN performance was decreased. It can be observed that up to a factor of 24, the performance does not show considerable changes, but for a ratio of 32:1, the $F1$ score drops to 0.908. As the $F1$ scores and the PR curves illustrate, a factor of 24 shows a trade-off between performance and compression.

3.2 *Scenario 2: Training and Testing on Images of the Same Quality*

In this experiment, we have trained the networks multiple times, each time using training images that are compressed with a specific ratio. After training, the model is evaluated on the test set, which has been compressed with the same ratio as applied on the training set. Table 2 and Fig. 4 show the obtained results. The outcome drastically differs from the previous experiment. As can be observed, the performance of the CNN is in this case not much impacted, meaning that a CNN can be trained to handle larger compression ratios. The difference between the performance of the model under different compression ratios is minimal. In comparison with the performance of the model in the previous experiment (scenario 1), the improvement is significant per compression rate. For example, the $F1$ score for compressed images with the factor of 164 is equal to 0.934, whereas when the model is trained on high-quality images, its $F1$ score was only 0.586. This represents about 59% improvement. A possible explanation for such a strong improvement is the adaptation of the network parameters to the distortion and degradation of the image quality, which are also present in its training set.

3.3 *Scenario 3: Evaluation on Varying-Quality Images with Fixed Compressed Trained Images*

In this experiment, the performance of the model, which was trained on images compressed with a factor of 48, is evaluated on a compressed test set with various compression ratios as well as uncompressed images. The compression ratio of 48:1 was selected, as it shows a maximum compression ratio where the $F1$ score of the CNN drops <10% of its maximum, according to the previous experiment (scenario 2). As we can observe from Table 3 and Fig. 5, the results improve for the higher compressed images (lower than 48:1 factors), compared with the first experiment when the model was trained on uncompressed images. The reason may be similar to what is observed in scenario 2 because the system has learned the compression

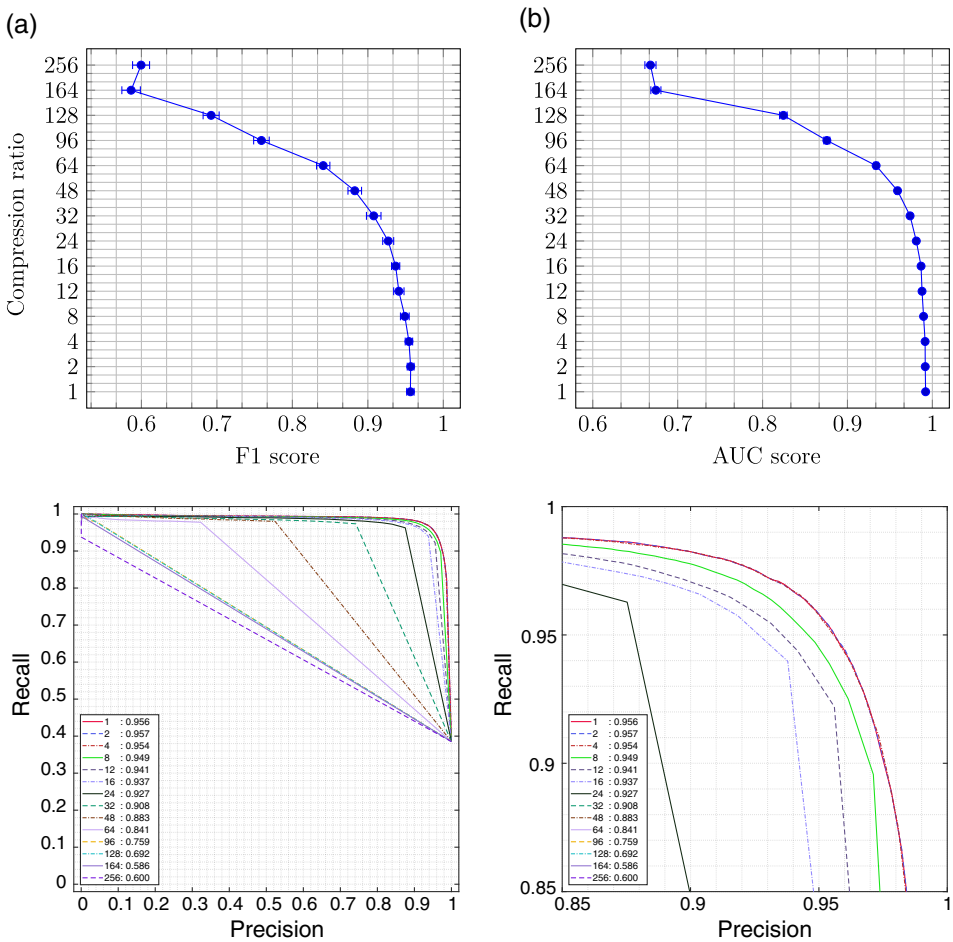

Fig. 3 Evaluation of the CAD model on compressed test images, when trained on uncompressed (1:1) high-quality training images: (a) PR curves and (b) enlarged view on the PR curves. The *F*1 scores are shown in the legend of the graphs.

Table 2 *F*1 scores, AUC values, and their 95% CIs for the CAD model, tested on the compressed images, which have been compressed with the same ratio as applied to the training images (scenario 2).

		JPEG 2000 compression ratios													
Metric		1:1	2:1	4:1	8:1	12:1	16:1	24:1	32:1	48:1	64:1	96:1	128:1	164:1	256:1
F1 score	CI upper bound	0.955	0.952	0.952	0.949	0.953	0.950	0.946	0.945	0.943	0.927	0.933	0.930	0.933	0.908
	Mean	**0.956**	**0.953**	**0.953**	**0.950**	**0.954**	**0.951**	**0.947**	**0.946**	**0.944**	**0.929**	**0.934**	**0.931**	**0.934**	**0.910**
	CI lower bound	0.957	0.954	0.955	0.952	0.956	0.953	0.948	0.947	0.946	0.930	0.936	0.933	0.936	0.912
AUC	CI upper bound	0.991	0.991	0.990	0.990	0.991	0.990	0.989	0.988	0.988	0.981	0.984	0.984	0.985	0.975
	Mean	**0.992**	**0.991**	**0.991**	**0.990**	**0.991**	**0.991**	**0.990**	**0.989**	**0.989**	**0.982**	**0.985**	**0.985**	**0.985**	**0.976**
	CI lower bound	0.992	0.992	0.991	0.991	0.992	0.991	0.990	0.989	0.989	0.982	0.986	0.985	0.986	0.977

artifacts from the training samples. In comparison with the second experiment, the performance slightly decreases on either side of the trained compression ratio. In a nutshell, from the results, we can observe that a trained CNN model on the low-quality images, e.g., with compression ratio of 48:1, can perform almost equally well on all higher-quality images and even on the slightly lower-quality samples.

4 Conclusion and Discussion

Compression of histopathology images has not yet been approved by regulatory agencies in the US for clinical applications. The contribution of this paper is that the investigation provides evidence that compression may be used from the CAD point of view, but a much larger effort is needed to accept compression

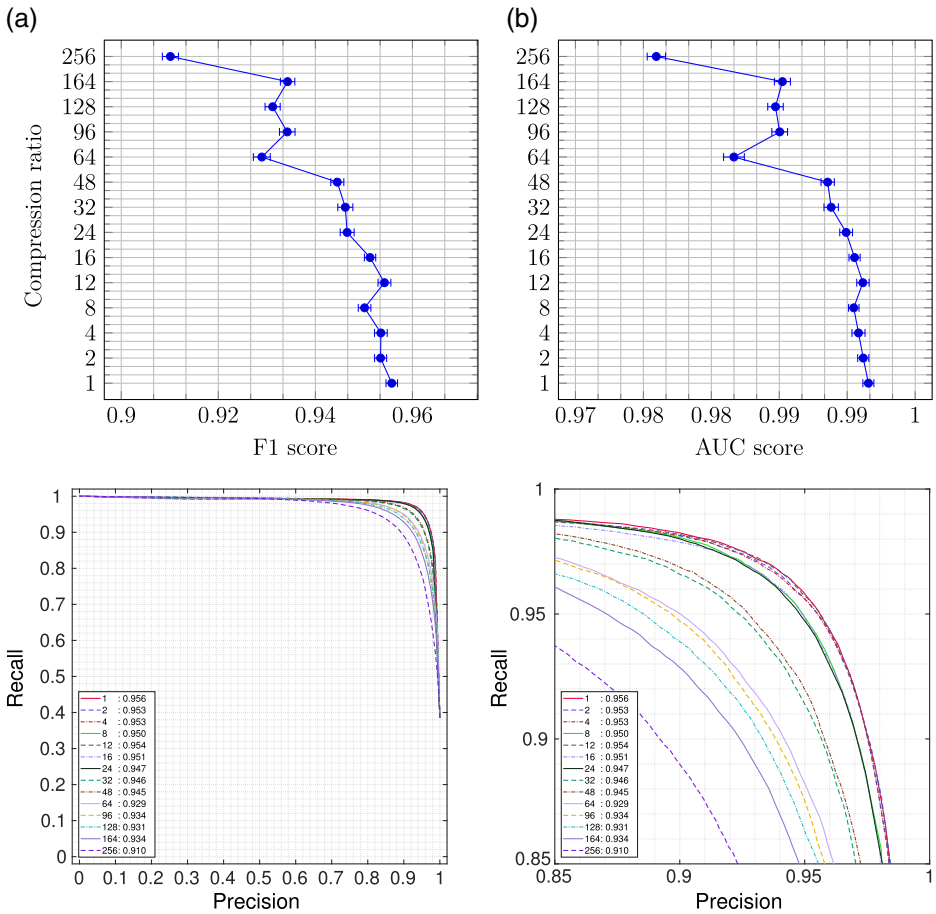

Fig. 4 Evaluation of the CAD model on compressed test images with the same compression ratio as applied to the training images: (a) PR curves and (b) enlarged view on the PR curves. The *F*1 scores are shown in the legend of the graphs.

Table 3 *F*1 scores, AUC values, and their 95% CIs for the CAD model tested on the images with different compression ratios while it was trained on 48:1 compressed training images (scenario 3).

		JPEG 2000 compression ratios													
Metric		1:1	2:1	4:1	8:1	12:1	16:1	24:1	32:1	48:1	64:1	96:1	128:1	164:1	256:1
*F*1 score	CI upper bound	0.929	0.930	0.926	0.925	0.927	0.929	0.936	0.935	0.939	0.933	0.912	0.891	0.847	0.737
	Mean	**0.932**	**0.933**	**0.930**	**0.928**	**0.930**	**0.932**	**0.939**	**0.939**	**0.943**	**0.936**	**0.915**	**0.895**	**0.852**	**0.745**
	CI lower bound	0.935	0.936	0.934	0.931	0.934	0.936	0.942	0.942	0.946	0.938	0.919	0.899	0.856	0.752
AUC	CI upper bound	0.985	0.985	0.983	0.983	0.984	0.984	0.986	0.986	0.987	0.985	0.976	0.966	0.942	0.864
	Mean	**0.986**	**0.986**	**0.984**	**0.984**	**0.985**	**0.985**	**0.987**	**0.987**	**0.988**	**0.986**	**0.978**	**0.968**	**0.944**	**0.868**
	CI lower bound	0.987	0.987	0.986	0.985	0.986	0.986	0.988	0.989	0.989	0.987	0.979	0.971	0.946	0.872

in the clinical practice. In this investigation, the aim is to access the impact of JPEG 2000 compression on the diagnostic performance of a deep CNN model for detecting metastases in breast lymph node histopathological WSIs. To do so, three series of experiments have been set up. When employing the uncompressed high-quality images for training the deep neural network model, the performance drops significantly for a compression ratio of 48:1 and higher, but it does not change significantly for the 24:1 and lower compression ratios. Our findings about the robustness of the CNN model against JPEG 2000 compression is in near agreement with the cut-off quality threshold of 32:1 ratio in the work of Krupinski et al.,[11] which involved pathologists as observers for classifying benign versus malignant breast cancer in WSIs. However, this work did not examine any effects on the human performance when reviewing such images.

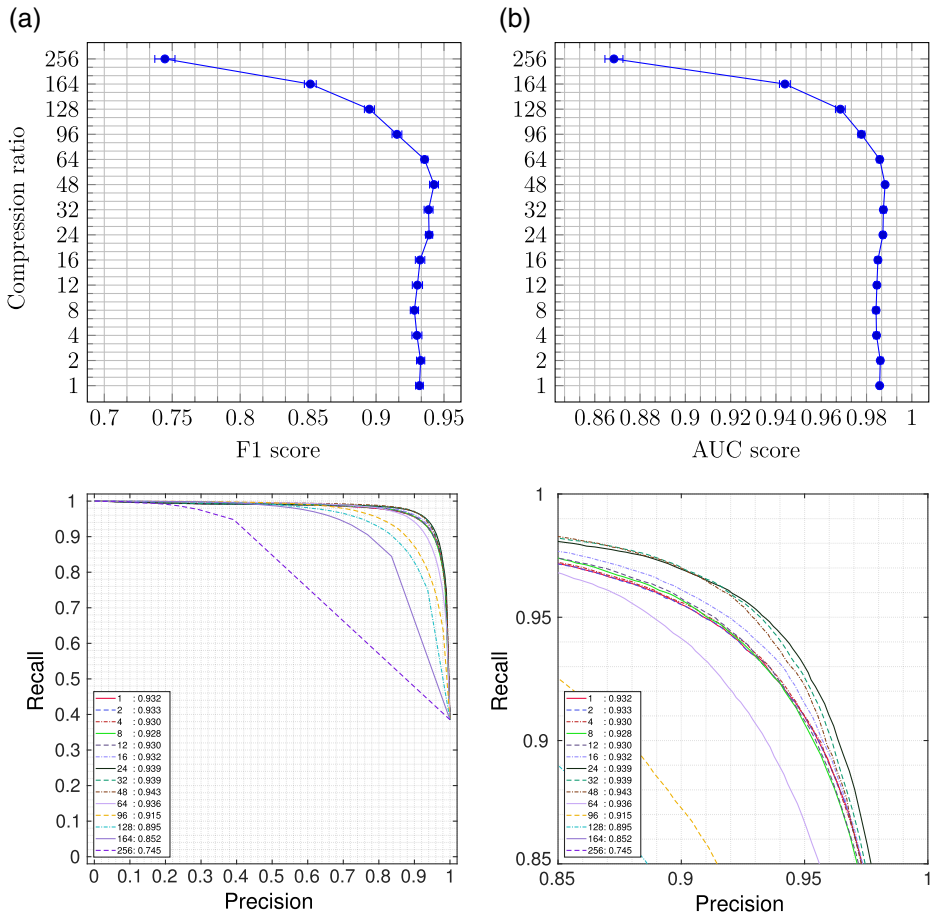

Fig. 5 Evaluation of the CAD model which was trained on 48:1 compressed images, while tested with variable compressed images: (a) PR curves and (b) enlarged view on the PR curves. The $F1$ scores are shown in the legend of the graphs.

Training and predicting on the same quality images produces drastically better results compared with the previous scenario in which the model only has been trained on uncompressed data. The outcome is remarkably improved for high compression ratios, while it does not change for low compression ratios. As an example of such an improvement, the performance of the model on compressed images with factor of 164 is on par with results of a previous experiment with factor of 24. This mainly happens because the CNN parameters have been optimized by observing the low-quality (distorted) training images. So it can be robust to some extent to the presence compression artifacts.

Finally, we have empirically shown that training the networks on 48:1 compressed images increases the performance for somewhat lower and higher compression ratios. These findings can help for designing a more efficient CAD system, mainly when a constraint exists for transmission and storage, such as in a system with a cloud-based computation or telepathology. Also we have shown that for a better training of the CNN model, the availability of high-quality uncompressed images is not a necessity.

Here we emphasize that our results presented in this study are specific to the CAD system and application described in this paper, and further work is needed to examine whether they generalize to other systems and applications.

Disclosures

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been funded by ITEA3 in the 3-D Pathology (ITEA151003) project.

References

1. F. G. Zanjani et al., "Cancer detection in histopathology whole-slide images using conditional random fields on deep embedded spaces," *Proc. SPIE* **10581**, 105810I (2018).
2. P. Bandi et al., "From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge," *IEEE Trans. Med. Imaging* **38**(2), 550–560 (2019).
3. G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.* **42**, 60–88 (2017).
4. B. E. Bejnordi et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA* **318**(22), 2199–2210 (2017).
5. J. Cho et al., "How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?" arXiv:1511.06348 (2015).
6. S. A. El-Seoud et al., "Big data and cloud computing: trends and challenges," *Int. J. Interact. Mobile Technol.* **11**(2), 34–52 (2017).
7. P. Bándi et al., "From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge," *IEEE Trans. Med. Imaging* **38**(2), 550–560 (2018).
8. T. Kalinski et al., "Lossless compression of JPEG 2000 whole slide images is not required for diagnostic virtual microscopy," *Am. J. Clin. Pathol.* **136**(6), 889–895 (2011).
9. M. D. Herrmann et al., "Implementing the DICOM standard for digital pathology," *J. Pathol. Inf.* **9**, 37 (2018).

10. L. Platiša et al., "Influence of study design on digital pathology image quality evaluation: the need to define a clinical task," *J. Med. Imaging* **4**(2), 021108 (2017).
11. E. A. Krupinski et al., "Compressing pathology whole-slide images using a human and model observer evaluation," *J. Pathol. Inf.* **3**, 17 (2012).
12. L. Pantanowitz et al., "Impact of altering various image parameters on human epidermal growth factor receptor 2 image analysis data quality," *J. Pathol. Inf.* **8**, 39 (2017).
13. A. Marcelo et al., "Effect of image compression on telepathology: a randomized clinical trial," *Arch. Pathol. Lab. Med.* **124**(11), 1653–1656 (2000).
14. J. P. Johnson et al., "Using a visual discrimination model for the detection of compression artifacts in virtual pathology images," *IEEE Trans. Med. Imaging* **30**(2), 306–314 (2011).
15. C. López et al., "Effects of image compression on automatic count of immunohistochemically stained nuclei in digital images," *J. Am. Med. Inf. Assoc.* **15**(6), 794–798 (2008).
16. A. Sharma, P. Bautista, and Y. Yagi, "Balancing image quality and compression factor for special stains whole slide images," *Anal. Cell. Pathol.* **35**(2), 101–106 (2012).
17. A. Brügmann et al., "Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains," *Breast Cancer Res. Treat.* **132**(1), 41–49 (2012).
18. C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG 2000 still image coding system: an overview," *IEEE Trans. Consum. Electron.* **46**(4), 1103–1127 (2000).
19. F. Aul-Llinas and J. Serra-Sagrista, "JPEG 2000 quality scalability without quality layers," *IEEE Trans. Circuits Syst. Video Technol.* **18**(7), 923–936 (2008).
20. G. Fan and W.-K. Cham, "Model-based edge reconstruction for low bit-rate wavelet-compressed images," *IEEE Trans. Circuits Syst. Video Technol.* **10**(1), 120–132 (2000).
21. D. Santa-Cruz, R. Grosbois, and T. Ebrahimi, "JPEG 2000 performance evaluation and assessment," *Signal Process. Image Commun.* **17**(1), 113–130 (2002).
22. A. Kanakatte et al., "Cloud solution for histopathological image analysis using region of interest based compression," in *39th Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc. (EMBC)*, IEEE, pp. 1202–1205 (2017).
23. D. Wang et al., "Deep learning for identifying metastatic breast cancer," arXiv:1606.05718 (2016).
24. H. Helin et al., "Optimized JPEG 2000 compression for efficient storage of histopathological whole-slide images," *J. Pathol. Inf.* **9**, 20 (2018).
25. S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Eighth Int. Conf. Quality of Multimedia Exp. (QoMEX)*, IEEE, pp. 1–6 (2016).
26. M. Dejean-Servières et al., "Study of the impact of standard image compression techniques on performance of image classification with a convolutional neural network," PhD thesis, INSA Rennes, Univ Rennes, IETR; Institut Pascal (2017).
27. O. Russakovsky et al., "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vision* **115**(3), 211–252 (2015).
28. Y. Liu et al., "Detecting cancer metastases on gigapixel pathology images," arXiv:1703.02442 (2017).
29. D. Ciresan et al., "Deep neural networks segment neuronal membranes in electron microscopy images," in *Adv. Neural Inf. Process. Syst.*, pp. 2843–2851 (2012).
30. M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks* **106**, 249–259 (2018).
31. C. Szegedy et al., "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2818–2826 (2016).
32. E. Kldiashvili, "Telemedicine for pathology," *Stud. Health Technol. Inf.* **131**, 227–244 (2008).
33. J. Davis and M. Goadrich, "The relationship between precision–recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, ACM, pp. 233–240 (2006).

Farhad Ghazvinian Zanjani received his master's degree in biomedical engineering from Amirkabir University of Tehran and his master degree in computer science (with honors) from Radboud University of Nijmegen, The Netherlands. Since 2016, he has been working on his PhD in signal processing at Eindhoven University of Technology. His current research interests include artificial intelligence, machine learning, and medical image analysis. He is a member of SPIE.

Svitlana Zinger received her PhD in 2004 on interpolation and resampling of 3-D data from the École nationale supérieure des télécommunications, Paris, France. In 2005, she was a postdoc fellow in the Multimedia and Multilingual Engineering Laboratory, French Atomic Agency, France. From 2006 to 2008, she was an associate researcher in the AI Department, University of Groningen, The Netherlands. She is currently an assistant professor in the Video Coding and Architectures Research Group, Eindhoven University of Technology, The Netherlands. Her research interests are image analysis for computer-aided diagnosis and prognosis.

Bastian Piepers received his master's degree in science of engineering from the Catholic University of Leuven, Belgium, in 2001. Currently, he is an active project manager in the Technology and Innovation Group of Barco Healthcare. His research activities are image processing, color calibration, and GPU acceleration.

Saeed Mahmoudpour received his PhD in computer and communications engineering from Kangwon National University, South Korea, in 2017. In 2016, he joined IMEC, Leuven, Belgium, as a research scientist and he is currently a postdoctoral fellow at IMEC and in the Department of Electronics and Informatics, Vrije Universiteit Brussel (VUB), Brussels, Belgium. He is a member of the ISO/IEC JTC1/SC29/WG1 (JPEG) Standardization Committee. His research interests include image quality assessment, machine learning, perceptual modeling, and high-dynamic range imaging.

Peter Schelkens holds a professorship in the Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium, and is a research group leader at IMEC, Leuven, Belgium. In 2013, he received an EU ERC Consolidator Grant focusing on digital holography. He authored or co-authored more than 250 journals and conference publications. He is the co-editor of the books *The JPEG 2000 Suite* (Wiley, 2009) and *Optical and Digital Image Processing* (Wiley, 2011). His current research focuses on the multidimensional signal processing while especially focusing on cross-disciplinary research. He is a member of the ISO/IEC JTC1/SC29/WG1 (JPEG) Standardization Committee. He is an associate editor of the *IEEE Transactions on Circuits and Systems for Video Technology* and *Signal Processing: Image Communications*.

Peter H. N. de With received his PhD from the University of Technology, Delft, The Netherlands. From 1984 to 1997, he worked for Philips Research Eindhoven on video compression and programmable TV architectures. From 1997 to 2000, he was a full professor of computer engineering at the University of Mannheim, Germany, and the chair of Digital Circuitry and Simulation. Since 2011, he has been a full professor at the Eindhoven University of Technology (TU/e), heading the chair on video coding and architectures. He is a fellow of the IEEE and has (co)authored over 400 papers on video coding, analysis, architectures, 3D processing, and their realization. He is a (co) recipient of multiple papers awards of the IEEE CES, ICIP, and a Eurasip Signal Processing Award.