Piotr JÓŹWIK-WABIK[1], Bartłomiej GŁADYS[1], Marek HERMANSA[1], Dawid MACHA[1], Seweryn KALISZ[1], Tomasz STRZODA[1], Paweł FOSZNER[2], Adam POPOWICZ[3], Michał MARCZYK[*4]

Chapter 7. REMOVING COMPRESSION ARTIFACTS ON WHOLE SLIDE HE-STAINED HISTOPATHOLOGICAL IMAGES

7.1. Introduction

Histopathological tissue slides stained with hematoxylin and eosin (HE) are commonly used in medical diagnosis. The hematoxylin stains cell nuclei blue, and eosin stains the extracellular matrix and cytoplasm pink. Other structures are seen with different combinations of these colors. A single slide contains a wealth of information that can be used to monitor the mechanisms contributing to disease progression or patient survival [1]. Since whole-slide imaging (WSI) scanners were introduced, it is possible to digitize large fragment of tissue section in a cost-effective and efficient manner. One of the biggest advantages of WSI compared to conventional glass slides is that they enable the use of automatic image analysis methods. This allows to perform diagnosis-related tasks like segmentation of Region of Interest (ROI) such as tumor region detection [2], lymphocyte detection [3] and many others.

Most WSI scanners perform scanning at 20× or 40× magnification with a spatial resolution in the order of 0.50 μm/pixel and 0.25 μm/pixel, respectively [4]. WSIs are often saved as SVS files. In this format the images are stored as a pyramid structure with diffrent resolution at each level of the pyramid. Because of the large size (highest-

[1] Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology.
[2] Department of Graphics, Computer Vision and Digital Systems, Silesian University of Technology.
[3] Department of Electronics,Electrical Engineering and Microelectronics, Silesian University of Technology.
[*] Corresponding author: michal.marczyk@polsl.pl, ul. Akademicka 16, 44-100 Gliwice, PL
[4] Department of Data Science and Engineering, Silesian University of Technology.

resolution images in SVS can have hundreds of thousands of pixels), the captured RGB images are compressed most commonly with JPEG or JPEG 2000 methods.

Applying JPEG compression to WSI may introduce image artifacts. When the quality parameter Q=30 is used, the resulting 8x8 pixel blocks can be visually seen. These artifacts may have a negative impact on the segmentation of small structures observed in the tissue image. In image analysis a large number of methods designed to reduce compression artifacts exist, ranging from relatively simple hand-designed filters to machine learning-based image restoration methods. Among common filtering methods, there are adaptive filtering with locally adjusted filter kernels [5], trilateral filters [6] or block-shift filtering-based algorithm [7]. Learning-based methods are built on large datasets to approximate inversion of the compression function. Some examples are multi-layer perceptron for JPEG deblocking [8], piecewise linear regression in the space of DCT [9] or maximum a posteriori (MAP) framework build upon experts scores [10].

To illustrate the problem, we applied a segmentation-based method for lymphocyte identification (LISI) [11, 12] on whole-slide HE stained tissue images with a different JPEG compression quality parameter Q (Figure 7.1). As can be seen, identyfing lymphocytes (represented by blue cluster) and other tissue structures under existence of compression artifacts is problematic and the segmented regions might be improper (Figure 7.1). In this paper, we introduce an edge-preserving algorithm for removing compression artifacts. First, using dataset containing smaller fragments of uncompressed image, we presented visually the results of proposed method and evaluated its performance using several image quality measures. Then, using whole-slide image data with compressed images, we showed the impact of artifacts removal on the segmentation obtained by the LISI algorithm.

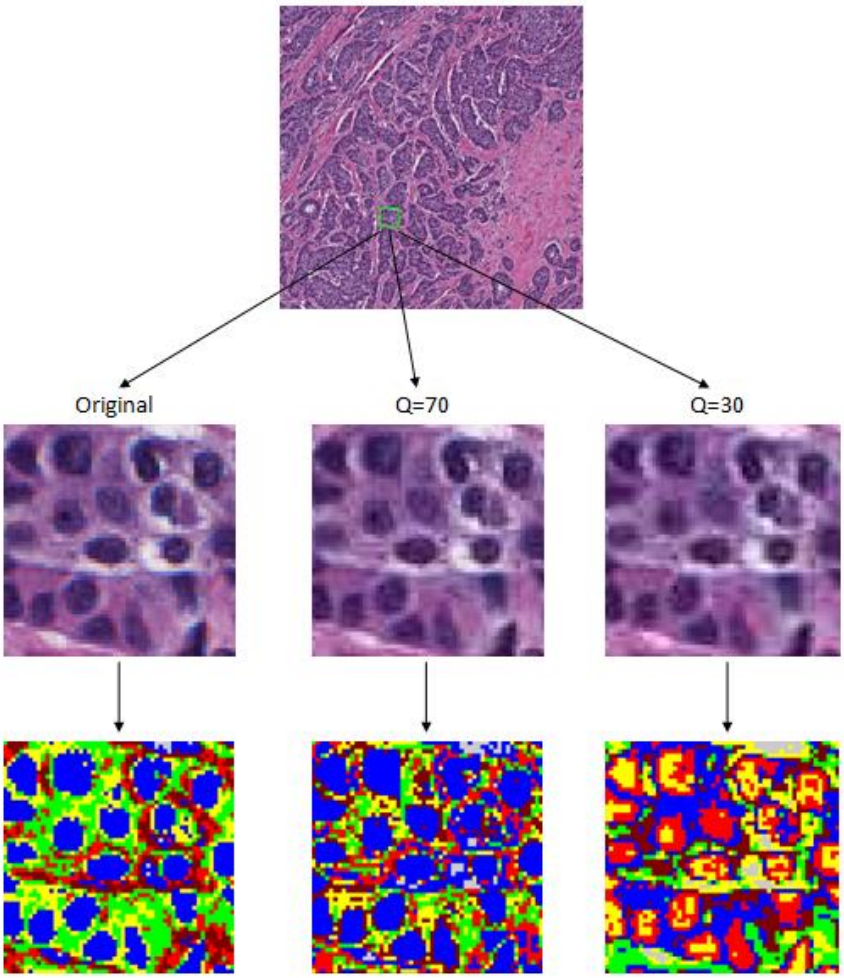

Fig. 7.1. Results of the image segmentation using LISI method on the same part of a whole-slide image with different compression quality parameter
Rys. 7.1. Wyniki segmentacji obrazu metodą LISI na tej samej części obrazu całej tkanki dla różnych wartości parametru jakości kompresji

7.2. Materials and methods

7.2.1. Data

BreCaHAD dataset contains 162 histopathological images of breast cancer divided into 17 different cases [13]. The images were obtained from archival exemplary cases of surgical pathology that were archived for educational purposes. Each image was taken with 40x magnification and saved as 1360x1024 uncompressed TIF file.

TCGA-BRCA dataset contains 1 987 whole slide images of breast cancer patients [14]. The images were scanned with usually 40x magnification and saved in SVS format using lossy JPEG compression with quality Q between 30 and 70. We randomly selected

a single image with Q equal 30 to show the results of segmentation after using proposed compression artifacts removal method.

7.2.2. Image quality metrics

Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are the most widely used metrics to evaluate the quality of reconstructed images after artifact removal. MSE is computed by averaging squared intensity differences of the distorted image and the original image. PSNR is the most versatile measure for image restoration tasks and can be represented as follows:

$$PSNR = 10 \cdot log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{1}$$

where MAX_I is the maximum possible pixel value of the image.
These two methods have been criticized because they do not correlate well with perceived quality measurement. To face these issues, Structural Similarity index (SSIM) has been proposed [15], which operates on the principle of comparison of local patterns of pixel intensities. Given images *I* and *J*, SSIM is defined as follows:

$$SSIM(I,J) = \frac{(2\mu_I\mu_J + C_1)(2\sigma_{IJ} + C_2)}{(\mu_I^2 + \mu_I^2 + C_1)(\sigma_J^2 + \sigma_J^2 + C_2)} \tag{2}$$

where μ is the average and σ is the variance of pixel intensity.
As the last measure, we calculated spatial correlation coefficient (SCC) between the high-pass filtered original and processed images as an index of the spatial quality. This measurement is based on the fact that the most important spatial information in original image is mostly concentrated in the high frequency domain.

7.2.3. Compression artifacts removal method

The algorithm starts by counting gradient magnitudes for each color channel of an image *I* and then binarizes them using Otsu method. Next, the edge map *M* is created using the following assumption: if a pixel is annotated as an edge in all channels, we assign the value 0, otherwise 1. Additionally, all pixels of the compression grid get also value 1. Filtering is then performed on *M* to get the weight matrix *W*. Subsequently, each channel of *I* is multiplied with corresponding values of *M* and filtered after that. As a result, we obtain the matrix *R*. At the end, we divide each value of *R* by corresponding values of *W* to obtain the smoothed image *I'*.

Proposed algorithm is inpired by the method described in [16], however we improved it significantly to make it relevant for processing of histopathological images. In our implementation we do not filter color channels separately, but we use combined information from all channels. Moreover, we decided to use pixels annotated as real edges, but we introduced an edge map. Last difference is that for filtering we use a 7x7 gaussian filter with standard deviation σ=1.1.

7.3. Results and discussion

We used BreCaHAD dataset that includes uncompressed images to evaluate performance of proposed compression artifacts removal method under different compression rates. We tested three JPEG compression quality values Q: 30, 70 and 100. We compared the original images with the compressed one and the one after removing artifacts using three image quality metrics described in the previous section (Figure 7.2). Calculating different metrics on 162 images revealed mutual results that can be used before deciding about running compression artifacts removal. Application of proposed algorithm increased image quality (p=3.45e-06 for PSNR, p=2.54e-12 for SSIM, p=3.44e-06 for SCC, measured using Wilcoxon test) for slides with low compression quality (Q=30). When Q was set to 70, in most cases the quality was reduced after removing artifacts. For Q=100, it seems that compression artifacts removal is not necessary. Results obtained by SCC metric seem to be most stable, as the results on different images are similar.

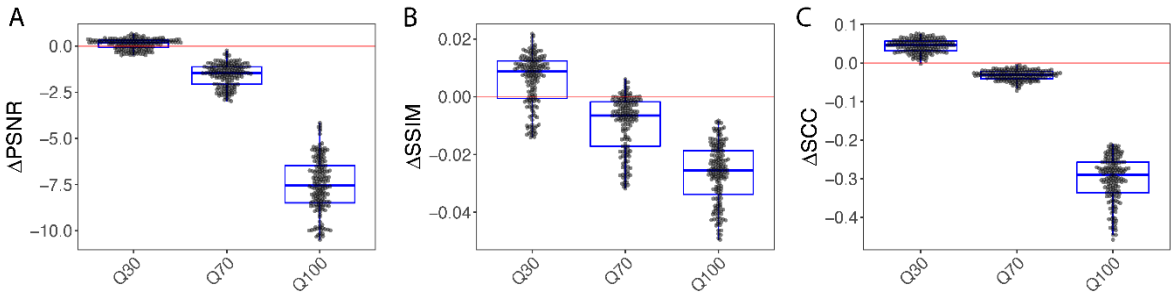

Fig. 7.2. Comparison of image quality before and after compression artifacts removal using PSNR (a), SSIM (b) and SCC (c) measures. Delta value shown in y-axis is the measure difference calculated between filtered and compressed image (the higher the value above 0, the better). X-axis shows results after compression with different quality values Q

Rys. 7.2. Porównanie jakości obrazu przed i po usunięciu artefaktów kompresji przy użyciu miar PSNR (a), SSIM (b) i SCC (c). Wartość delta pokazana na osi y to różnica miar obliczona między obrazem filtrowanym a skompresowanym (im wyższa wartość powyżej 0, tym lepiej). Oś X pokazuje wyniki po kompresji dla różnych wartości jakości Q

To visually show the effect of compression artifacts removal, we selected two images based on the image quality score values. We chose images compressed with Q=30 value and plotted images with the best and the worst scores (Figure 7.3). Direct comparison of original images shows that there were probably some technical problems with staining for worst-score image, since we observe a large color variance within each individual cell nucleus (Figure 7.3 A, D). This might explain the problem of the proposed method to efficiently detect edges on this image, which leaded to worse result. Nevertheless, in both cases 8x8 pixel blocks introduced during compression (Figure 7.3 B, E) are efficiently removed (Figure 7.3 C, F). Observed image blur is an effect of Gaussian smoothing. Despite the fact that two extreme examples were selected, visually it is hard to find the difference between processed images. Further research is required in order to obtain higher correlation between results after compression artifacts removal and image quality metrics outcome.

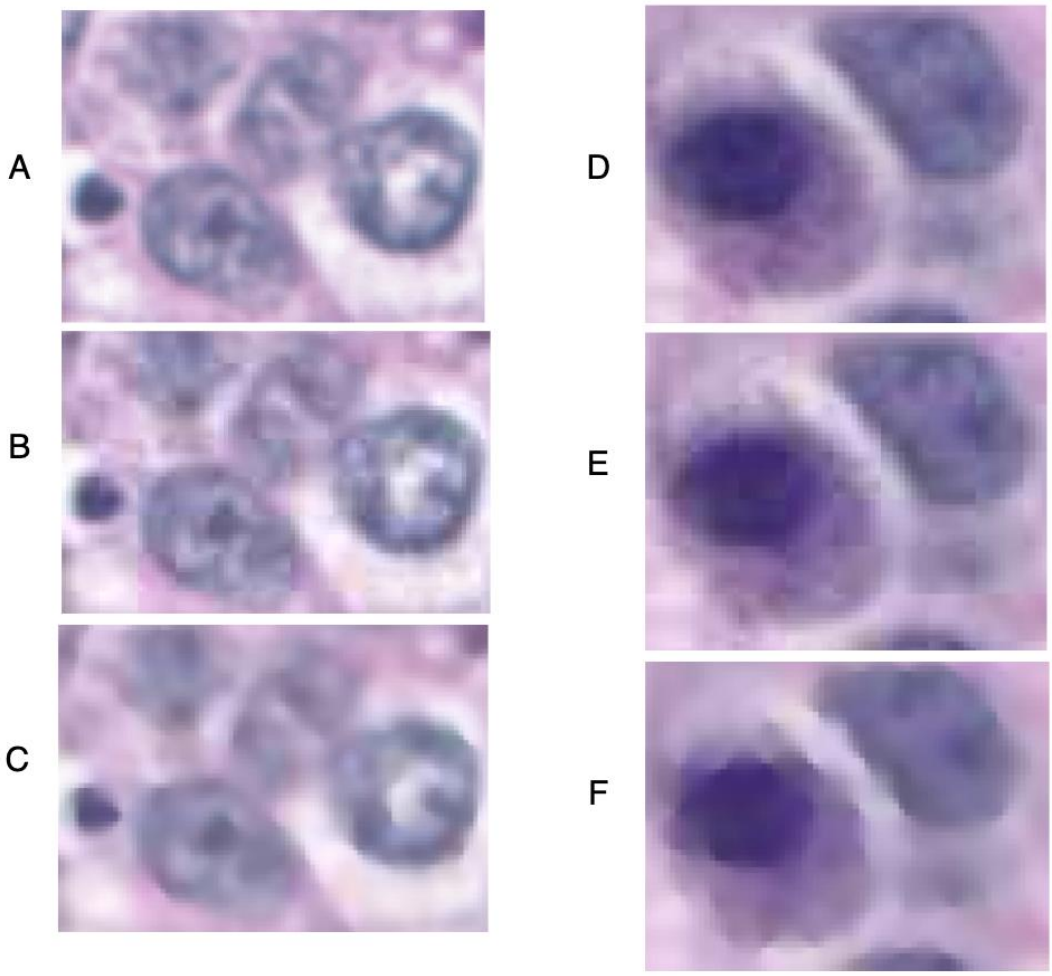

Fig. 7.3. Zoomed sections of the images with the worst (left) and the best (right) image quality scores after removing compression artifacts. Individual panels show original image (A, D), image compressed with Q=30 (B, E) and image after compression artifacts removal (C, F)

Rys. 7.3. Powiększone fragmenty obrazów z najgorszą (po lewej) i najlepszą (po prawej) jakością obrazu po usunięciu artefaktów kompresji. Poszczególne panele pokazują oryginalny obraz (A, D), obraz skompresowany dla Q=30 (B, E) oraz obraz po usunięciu artefaktów kompresji (C, F)

We used TCGA-BRCA dataset that includes compressed images to visually evaluate influence of the proposed compression artifacts removal method on the image segmentation using LISI algorithm. Selected image presented on Figure 7.4 was scanned with 40x magnification and saved in SVS format using lossy JPEG compression with parameter Q=30. The color coding in LISI algorithm of the most important clusters is as follows: blue for nuclei, green for stroma and yellow for cytoplasm. It can be seen that the clusters obtained after removing the compression artifacts reflect the actual biological structures much better than in case with no processing. We visually compared all other images in the TCGA-BRCA dataset. For a significant number of images, the use of the compression artifact removal algorithm had a positive effect on the segmentation results obtained with the LISI algorithm. Unfortunately, it is impossible to quantify how much the use of the described algorithm affects obtained results, since the ground truth is not known.

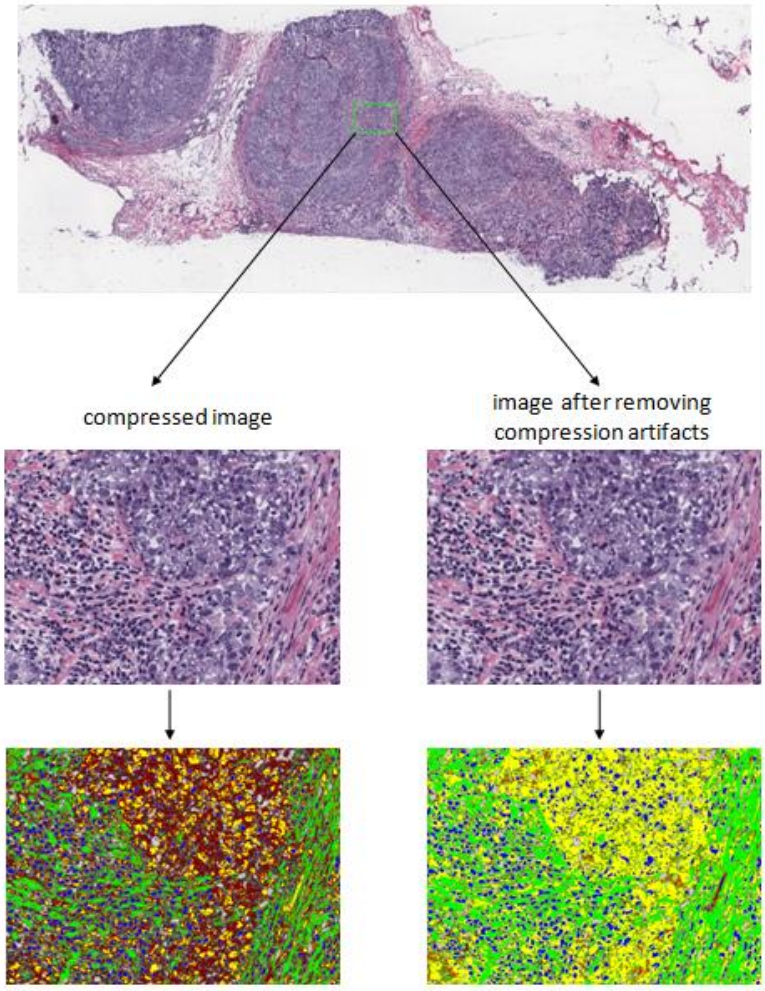

Fig. 7.4. The influence of the compression artifacts removal algorithm on the segmentation of whole-slide image using LISI method. Left side: segmentation results on original image, right side: segmentation results after removing the compression artifacts

Rys. 7.4. Wpływ algorytmu usuwania artefaktów kompresji na segmentację obrazu całej tkanki metodą LISI. Lewa strona: wyniki segmentacji na oryginalnym obrazie, prawa strona: wyniki segmentacji po usunięciu artefaktów kompresji

7.4. Conclusions

JPEG compression of HE-stained images is currently a standard in histopathology. Our approach to remove compression artifacts was successful, mostly for images compressed with low quality parameter Q, which is a commonly used setting. When Q value was set higher, like Q=70 in our case, performance of the proposed method was slightly worse. Nevertheless, the main goal of removing 8x8 compression grid has been achieved. We showed that existing of compression artifacts might affect image segmentation and their removal is crucial in that task. We plan to tune the algorithm to work better with images compressed with higher quality parameter, to develop an universal compression artifacts removal tool for HE stained WSI images.

Acknowledgements

The research reported in this paper was co-financed by the European Union from the European Social Fund in the framework of the project "Silesian University of Technology as a Center of Modern Education based on research and innovation" POWR.03.05.00-00-Z098/17.

Bibliography

1. M. Aswathy, M. Jagannath, Detection of breast cancer on digital histopathology images: Present status and future possibilities. *Informatics in Medicine Unlocked* **8**, 74-79 (2017).
2. F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, in *2016 international joint conference on neural networks (IJCNN)*. (IEEE, 2016), pp. 2560-2567.
3. J. Saltz *et al.*, Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Reports* **23**, 181-193.e187 (2018).
4. S. Nam *et al.*, Introduction to digital pathology and computer-aided pathology. *J Pathol Transl Med* **54**, 125-134 (2020).
5. D. Tschumperlé, R. Deriche, Vector-valued image regularization with PDEs: A common framework for different applications. *IEEE transactions on pattern analysis and machine intelligence* **27**, 506-517 (2005).
6. T. Wang, G. Zhai, JPEG2000 image postprocessing with novel trilateral deringing filter. *Optical Engineering* **47**, 027005 (2008).

7. G. Zhai, W. Lin, J. Cai, X. Yang, W. Zhang, Efficient quadtree based block-shift filtering for deblocking and deringing. *Journal of Visual Communication and Image Representation* **20**, 595-607 (2009).
8. G. Qiu, MLP for adaptive postprocessing block-coded images. *IEEE transactions on circuits and systems for video technology* **10**, 1450-1454 (2000).
9. K. Lee, D.S. Kim, T. Kim, Regression-based prediction for blocking artifact reduction in JPEG-compressed images. *IEEE Transactions on Image Processing* **14**, 36-48 (2004).
10. D. Sun, W.-K. Cham, Postprocessing of low bit-rate block DCT coded images based on a fields of experts prior. *IEEE Transactions on Image Processing* **16**, 2743-2751 (2007).
11. F. Binczyk, M. Marczyk, J. Polanska, in *MAQC Society 2nd Annual Meeting*. (Shanghai, China, 2018), pp. 33.
12. F. Binczyk, M. Marczyk, J. Polanska, in *Computational Approaches in Precision Medicine*. (Vienna, Austria, 2017), pp. poster.
13. A. Aksac, D.J. Demetrick, T. Ozyer, R. Alhajj, BreCaHAD: a dataset for breast cancer histopathological annotation and diagnosis. *BMC Research Notes* **12**, 82 (2019).
14. K. Clark *et al.*, The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging* **26**, 1045-1057 (2013).
15. Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**, 600-612 (2004).
16. B. Oztan, A. Malik, Z. Fan, R. Eschbach, in *Color Imaging XII: Processing, Hardcopy, and Applications*. (International Society for Optics and Photonics, 2007), vol. 6493, pp. 649306.

REMOVING COMPRESSION ARTIFACTS ON WHOLE SLIDE HE-STAINED HISTOPATHOLOGICAL IMAGES

Abstract

Whole-slide imaging in histopathology produces high-resolution digital files that may take up an enormous amount of disk space, therefore a wide range of compression methods is used in practice. The most common is JPEG compression algorithm that splits an image into 8x8 pixel blocks with adjustable compression quality. In most cases,

the compression quality parameter is set to around 30, which decreases the size of a single image significantly, but also leads to the appearance of an artificial grid. The grid significantly impedes the work of segmentation algorithms that are used to explore the complex structure of the scanned tissue. In this paper, we introduced an edge-preserving algorithm for removing compression artifacts and showed how using it affects the image quality and image segmentation. We conclude that removing artifacts might be crucial, mostly in case where the compression quality parameter is low.

Keywords: histopathological images, HE-staining, JPEG lossy compression, artifacts removal.