https://doi.org/10.1109/TMI.2018.2814058

Improving the Automated Detection of Calcifications using Adaptive Variance Stabilization

Alessandro Bria, *Member, IEEE*, Claudio Marrocco, Lucas R. Borges, Mario Molinara, Agnese Marchesi, Jan-Jurre Mordang, Nico Karssemeijer, *Senior Member, IEEE* and Francesco Tortorella, *Senior Member, IEEE*

***Abstract*—In this work, we analyze how stabilizing the variance of intensity-dependent quantum noise in digital mammograms can significantly improve the computerized detection of microcalcifications (MCs). These lesions appear on mammograms as tiny deposits of calcium smaller than 20 pixels in diameter. At this scale, high frequency image noise is dominated by quantum noise, which in raw mammograms can be described with a square-root noise model. Under this assumption, we derive an adaptive variance stabilizing transform (VST) that stabilizes the noise to unitary standard deviation in all the images. This is achieved by estimating the noise characteristics from the image at hand. We tested the adaptive VST as a preprocessing stage for four existing computerized MC detection methods on three datasets acquired with mammographic units from different manufacturers. In all the test cases considered, MC detection performance on transformed mammograms was statistically significantly higher than on unprocessed mammograms. Results were also superior in comparison with a 'fixed' (nonparametric) VST previously proposed for digital mammograms.**

***Index Terms*—digital mammography, quantum noise, variance stabilizing transform, microcalcification detection.**

I. Introduction

BREAST cancer is the first leading cause of death for cancer among the female population [1]. Early detection of breast cancer increases the survival rate and the treatment options for the patients [2]. Therefore, screening programs are implemented worldwide in which asymptomatic women are periodically invited for a mammographic exam. Although the effect on mortality reduction has been controversial for several years, recent studies proved that most deaths from breast cancer occur in unscreened women and encourage initiation of regular screening before age 50 [3]. To assist radiologists in reading mammograms, Computer-Aided Detection (CAD) systems have been developed that automatically find and mark suspicious lesions on the image [4]. It has been shown that the detection rate of individual radiologists increases with a CAD system [5]. However, to increase the role of CAD for screening, further improvements are required [6].

Among the earliest signs of breast cancer, microcalcifications (MCs) are usually associated with Ductal Carcinoma In Situ but can also be present in invasive cancers [7]. They are tiny deposits of calcium that appear on the mammograms as spots smaller than 1 mm in diameter corresponding to 10-20 pixels [8]. At this scale, the detectability of MCs is hampered by the presence of high frequency image noise [9]. In raw mammograms, image noise is dominated by intensity-dependent quantum noise that is produced by the random distribution of the photons within the image [10] and can be modeled by a Poisson distribution [11].

Several methods have been proposed in the literature for noise reduction or equalization in mammograms. In [12], an approach for noise detection and filtering through a neural network and three weighted-average filters was presented, while in [13] an adaptive frost filter was used to remove the noise. Other denoising approaches based on multiresolution analysis such as the wavelet transform [14], the undecimated multidirectional wavelet transform [15], and the curvelet transform [16] were analyzed, and recently, the Poisson Unbiased Risk Estimation-Linear Expansion of Thresholds (PURE-LET) noise removal algorithm has been applied to mammographic images [17]. As to noise equalization, it was originally proposed in the seminal work of [18] to eliminate the noise dependence on the image intensity in screen-film mammograms. In [19], noise equalization was embedded in a CAD scheme for the normalization of the local contrast features extracted from screen-film mammograms. An extension of this approach to digital mammograms was presented in [20] and [21].

When dealing with Poisson noise, another solution is to preprocess the image by applying a variance stabilizing transform (VST) so that the transformed noise is approximately Gaussian and can be more easily removed [22]. VSTs have been at the core of several works in statistics [23], [24], [25] and, quite recently, for noise removal in image [26], [27] and biomedical processing [28], [29], [30]. As to mammography, a denoising based on the Wiener filter was applied to synthetic [31] and screen-film mammograms [9] transformed with the Anscombe VST [24]. This approach was extended in [32] for effective denoising of digital mammography and tomosynthesis. However, the role of variance stabilization for CAD systems in clinical digital mammography has never been investigated.

In this paper we show how to improve the performance of

This research was funded by grant KUN 2012-5577 of the Dutch Cancer Society and supported by the Foundation of Population Screening Mid West. The support of NVIDIA Corporation, that donated the Titan X Pascal GPUs used in this work, is also gratefully acknowledged.

A. Bria, C. Marrocco, M. Molinara, A. Marchesi, and F. Tortorella are with the Department of Electrical and Information Engineering, University of Cassino and Southern Latium, Cassino, Italy (emails: {a.bria,c.marrocco,m.molinara,a.marchesi,tortorella}@unicas.it).

Lucas R. Borges is with the Department of Electrical and Computer Engineering, University of São Paulo, São Carlos, Brazil (email: lucas.rodrigues.borges@usp.br).

J-J. Mordang and N. Karssemeijer are with the Diagnostic Image Analysis Group, Radboud University Medical Centre, Nijmegen, The Netherlands (emails: {jan-jurre.mordang,nico.karssemeijer}@radboudumc.nl).

automated MC detection in digital mammography by stabilizing the variance of the intensity-dependent Poisson quantum noise. In [33] we already introduced a nonparametric VST that effectively stabilizes the noise variance but yields diverse noise levels for different images. Here, we design an adaptive VST that stabilizes the quantum noise to unitary standard deviation in all the images. The name 'adaptive' originates from the noise characteristics that are estimated from the image at hand. The adaptive VST can be used as preprocessing step for any MC detection scheme. We evaluated the performance of this VST when applied to four different CAD techniques on three multivendor mammographic databases.

II. Datasets

We collected three datasets consisting respectively of 252 mammograms acquired with Hologic digital mammography systems (Hologic, Bedford, Massachusetts, United States), $1,066$ mammograms acquired with GE Senographe systems (GE, Fairfield, Connecticut, United States), and 72 mammograms acquired with Siemens Mammomat Inspiration systems (Siemens, Erlangen, Germany). The images in the Hologic dataset were obtained from women who participated in a national screening program (Bevolkings Onderzoek Midden-West, The Netherlands) and were referred for diagnostic follow up. The other mammograms were acquired in Radboud University Medical Center (Nijmegen, The Netherlands) after referral in screening. In all mammograms, a total of 7,758, 7,579, and 1,787 individual MCs were annotated in the Hologic, GE, and Siemens dataset, respectively. Annotations were made on 'for-presentation' mammograms[1] by experienced readers who marked the center of each MC based on the diagnostic reports. All mammograms were acquired with standard clinical settings and only raw 'for-processing' images, traditionally used by CAD systems [35], were considered for this study.

III. Noise Model

Noise in a digital mammogram is assumed to consist of three major components: (i) structural noise due to spatial variations in detector sensitivity; (ii) quantum noise inherent to the counting nature of the x-rays; and (iii) electronic noise [34]. The noise standard deviation σ_i at spatial coordinate i from a raw digital mammogram can be described by [36]:

$$\sigma_i(y_i) = c_i\sqrt{y_i} + s_i y_i + d \tag{1}$$

where s_i and c_i are scaling factors of the structural and quantum noise, respectively, d is the standard deviation of the additive electronic noise, and y_i is the underlying signal.

For clinical applications, quantum noise can be considered as the dominant noise source in the image [10]. In this regard, McLoughlin et al. [20] reported that the noise in the compressed portion of the breast can be estimated using a single parameter model without the structural and additive noise components s_i and d. Since malignant MCs rarely occur in the peripheral zone of the breast, adopting this simplification would not affect the performance of automated MC detection.

[1]We recall that 'for-presentation' mammograms result from the postprocessing applied by manufacturers and are mainly intended for viewing [34].

In addition, according to Schie and Karssemeijer [21], even though c may depend on pixel location i due to the anode heel effect and other sources of variation, the authors have shown that modeling c as a nonuniform parameter yielded minimal improvement to the automated detection of MC.

In this work, we adopt the abovementioned approximations ($c_i = c, \forall i$; $d = 0$; $s_i = 0$) with the goal of enhancing the MC detection performance of a CAD system. This yields the following simplified square-root noise model:

$$\sigma_i(y_i) = c\sqrt{y}_i. \tag{2}$$

In clinical systems, the scaling factor c is mostly influenced by the quantum gain generated during the x-ray production and detection processes. The parameter c is also influenced by the detector gain, which is calibrated regularly during the flat-fielding process. Thus, c remains to be estimated from each mammogram individually as it is highly dependent on the anode material and angle. To this end, we require that raw 'for-processing' mammograms are used, since the post-processing applied by the manufacturers in 'for-presentation' mammograms may invalidate the noise model.

IV. Adaptive noise variance stabilization

Let z_i be the pixel of a raw digital mammogram at location i. Dropping the subscript i for simplicity, and under the assumptions of the noise model in Eq. 2, z is a random variable distributed according to a scaled Poisson process with mean and standard deviation given by:

$$\begin{aligned} \mu_z(y) &= y \\ \sigma_z(y) &= c\sqrt{y}. \end{aligned} \tag{3}$$

We aim at finding a transformation T in order to stabilize the signal-dependent standard deviation of the observations z to a desired constant S. By applying a widely used approach known as 'delta-method' [37], we can approximate the standard deviation $\sigma_{\tilde{z}}$ of the transformed signal $\tilde{z}$ with:

$$\sigma_{\tilde{z}} \approx \frac{dT(z)}{d\mu_z(y)}\sigma_z(y). \tag{4}$$

Since we want $\sigma_{\tilde{z}} \equiv S$, we obtain the differential equation:

$$dT(z) = \frac{S}{\sigma_z(y)}d\mu_z(y). \tag{5}$$

that has, by separation of variables, the following solution:

$$T(z) = \int^z \frac{S}{\sigma_z(y)}d\mu_z(y). \tag{6}$$

This expression appeared in early works (e.g. [23]) and it has been widely used in the field of signal processing including recent papers (e.g. [38], [26]). Although Eq. 6 does not follow from a rigorous mathematical derivation, it enjoys good asymptotic properties [39] and was empirically shown to provide reasonable stabilization in various applications [40]. In our case, by replacing μ_z and σ_z with Eq. 3, and after solving the integral, we obtain:

$$T(z) = \frac{2S}{c}\sqrt{z} + \kappa. \tag{7}$$

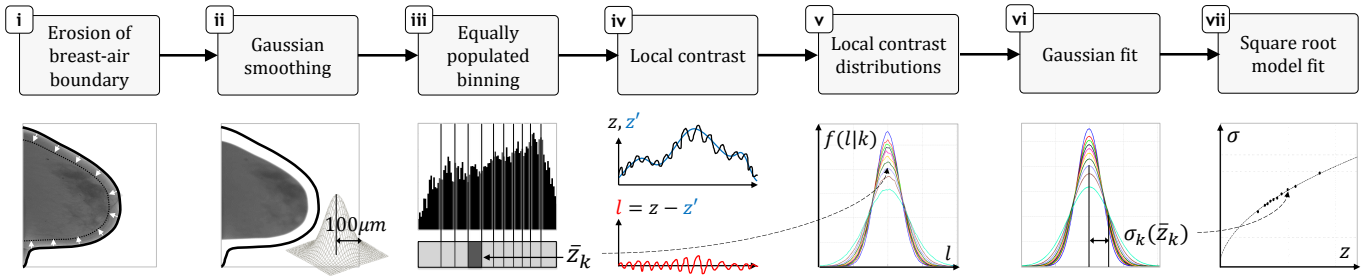

Fig. 1. Schematic overview of the procedure to estimate the noise parameter from a raw mammogram. The image shown in steps (i-ii) is contrast enhanced for better visualization. The plots shown in steps (iii-iv) are for illustration purposes only. The plots shown in steps (v-vii) are taken from a real test case.

We require that the transformation does not shift the grayscale, hence $T(0) = 0$ from which it follows that $\kappa = 0$. Further, we want to stabilize the noise to unitary standard deviation in all images, thus $S = 1$. This leads to the final form of the VST:

$$T(z) = \frac{2}{c}\sqrt{z}. \tag{8}$$

which requires c to be estimated from the image at hand (hence the name 'adaptive'). This is obtained through the following steps (see Fig. 1):

(i) the breast-air boundary is eroded by half of the breast thickness to discard the inner peripheral zone where the assumptions made for the noise model do not hold.
(ii) the image is convolved with a low-pass Gaussian filter of standard deviation 100 μm, which corresponds to 1-2 pixels for typical mammogram resolutions. This very small smoothing filter allows a good separation of the true signal from the noise by subtracting the smoothed image from the original image (see step (iv)).
(iii) the grayscale of the smoothed image is divided into K equally populated bins, each associated to its weighted mean intensity $\overline{z}_k$, $k = 1, ..., K$.
(iv) the local contrast $l_i = z_i - z'_i$ is computed at each location i, with z'_i being the pixel value of the smoothed image.
(v) local contrast distributions $f(l|k)$ are obtained by mapping l_i to the bin of z'_i.
(vi) Gaussian least-squares fitting with truncation of the values below the 20% of the maximum is applied to each local contrast distribution to estimate the noise level σ_k.
(vii) c is computed by least squares fitting of the noise samples $(\sigma_k, \overline{z}_k)_{k=1,...,K}$ to the square-root noise model of Eq. 2.

The number of bins K should be chosen small enough to allow an accurate estimation of the distribution of the data in each bin (see steps (v-vi)). On the other hand, if K is too small, there are not enough samples to characterize the noise dependency from the intensity (see step (vii)). In our experiments, we set $K = 10$ which was found to be a good trade-off.

The procedure described above to estimate the scaling parameter c is a slightly modified version of the truncated distribution method proposed by McLoughlin et al. [20], and later improved by Schie and Karssemeijer [21]. Specifically, the gaussian-smoothed image is used here not just to compute the local contrast, but also to generate the histogram bins (step (iii)) and to map the local contrast values to the corresponding bin (step (v)). This ensures regular local contrast distributions which fit well to Gaussians (see Fig. 1v). Moreover, the standard deviation of these distributions is estimated by means of Gaussian least-squares fitting (step (vi)), which is less sensitive to outliers than calculating the standard deviation directly from the distribution.

TABLE I
PER-DATASET AVERAGES OF c PARAMETERS AND CORRESPONDING R^2 STATISTICS CALCULATED FROM EACH IMAGE.

Dataset	$\overline{c}$	$\overline{R^2}$
Hologic	0.397 ± 0.037	0.974 ± 0.025
GE	0.183 ± 0.031	0.968 ± 0.199
Siemens	0.281 ± 0.019	0.956 ± 0.005

In Table I we report per-dataset averages of c parameters along with the corresponding R^2 statistics as a measure of goodness of fit to the square-root noise model of Eq. 2. The high R^2 values, close to 1 (perfect fit), indicate that the adopted model is a good approximation of the quantum noise in the inner breast area.

The C++ implementation of the adaptive VST is publicly available at https://github.com/abria and it already embeds the estimation of c from the mammogram provided as input.

V. EXPERIMENTS

We provide an extensive experimental study to evaluate the effectiveness of the adaptive VST and the influence it has on the computerized detection of MCs. Results are compared with those obtained using (i) unprocessed raw mammograms, and (ii) mammograms rescaled with a 'fixed' (nonparametric) VST that we proposed in [33] and used in a full CAD system [41], [42]:

$$T_{fixed}(z) = \sqrt{z(2^b - 1)} \tag{9}$$

where b is the number of bits of the input image. It is important to emphasize that, as demonstrated in [33], this transform does not depend on the c parameter and rescales the noise to image-dependent levels.

A. Variance stabilization

To evaluate the effectiveness of the adaptive VST, we ran steps (i)-(vi) of the procedure described in Section IV to

TABLE II
PER-DATASET AVERAGES OF PMSE VALUES CALCULATED FROM EACH IMAGE. THE LOWEST PMSE VALUES ARE LISTED IN BOLD.

Dataset	Fixed VST [33]	Adaptive VST
Hologic	2.07 ± 4.23	$\mathbf{0.25 \pm 0.21}$
GE	0.66 ± 2.89	$\mathbf{0.15 \pm 0.14}$
Siemens	1.21 ± 1.13	$\mathbf{0.32 \pm 0.27}$

obtain the noise samples $(\sigma_k, \overline{z}_k)$, and calculated the Percentage Mean Squared Error (PMSE) as a measure of the relative difference between the noise samples and the expected constant noise standard deviation S. Average and standard deviation values of PMSE are reported in Table II for the three datasets considered. All PMSE values were well within the 5%, suggesting that both the VSTs were effective in stabilizing the noise variance. In all cases, the adaptive VST largely outperformed the fixed VST, yielding up to one order of magnitude improvements in average and standard deviation values of PMSE. Scatter plots of noise samples from all the mammograms in each dataset are reported in Fig. 2 for completeness.

B. Microcalcification Detection

We selected, implemented, trained, and tested four among the most relevant CAD techniques available in the literature.

1) MC detectors: The following methods for MC detection were selected for this study: (i) the cascade of ranking-based boosting classifiers (*Cascade*) of Bria et al. [41], [43] that was the first CAD system to be favorably compared with a commercial CAD; (ii) the Convolutional Neural Network (*CNN*) approach of Mordang et al. [44] that was shown to be competitive with the previously mentioned Cascade; (iii) the Support Vector Machine (*SVM*) approach of El Naqa et al. [45] that is one of the most cited works in the field; and (iv) the weighted Difference-of-Gaussian (*DoG*) of Dengler et al. [46] that is a well-known MC detector often used as a baseline method [45]. All methods were implemented in C++ with the help of OpenCV [47] for *Cascade* and *DoG*, LIBSVM [48] for *SVM*, and Caffe [49] for *CNN*. Care was taken to ensure implementations were as faithful as possible to their original description in the literature. However, the following differences must be noted: (i) *SVM* did not include the background suppression of El Naqa et al. [45]; (ii) *CNN* consisted of a single CNN instead of a sequence of two CNNs as in Mordang et al. [44]; and (iii) *DoG* was adapted to yield a soft output for each pixel by taking the maximum DoG filter response in the detector window.

2) Training and test sets: All MC detection methods were trained and tested on patches of size 1×1 mm so that each MC is contained in one patch [8]. MC patches were extracted by centering the window on the annotated MC centers, yielding 7,758, 7,579, and 1,787 MC patches for the Hologic, GE, and Siemens dataset, respectively. The background patches were extracted from the remaining regions of the image with overlapping sliding windows, totaling 25,190,476, 27,017,503, and 18,539,324 patches for Hologic, GE, and Siemens datasets, respectively. We applied case-based 10-fold cross validation for each dataset in all experiments[2]. In each cross validation step, the MC detector was trained on 90% of the cases, and tested on the other 10%. When splitting the data into a training and test set, the patches belonging to the same case were assigned to the same set.

3) Training parameters: Cascade detectors were trained using the detection rate $d = 0.99$ and false positive rate $f = 0.30$ for each stage classifier. *SVM* detectors were trained using Gaussian RBF kernel, with Gaussian kernel variance and regularization parameter C that were automatically selected by means of a coarse-to-fine grid search in a nested 5-fold cross validation performed within each training set. The *CNN*s were trained using backpropagation and Stochastic Gradient Descent with weight updates that proceeded in batches of 32 patches. The learning rate was initially set to 10^{-3} and decreased by a factor of 10 every 6 epochs. In total, the learning rate was decreased 5 times, and the learning was stopped after 30 epochs. Momentum and weight decay were set respectively to 0.9 and 5×10^{-4}. The *DoG* detectors were optimized on the training set performing a grid search over the three parameters (positive/negative kernel width, and positive kernel weight), and the configuration yielding the smallest classification error was used on the test set.

4) Results: We performed a total of 36 experiments, resulting from four MC detectors, three datasets, and three methods (no rescaling, fixed VST, and adaptive VST). MC detectors have been evaluated in terms of Receiver Operating Characteristics (ROC) curve by plotting True Positive Rate (TPR) against False Positive Rate (FPR) for a series of thresholds on the detector output associated to each sample. Furthermore, the mean sensitivity of the ROC curve in the specificity range on a logarithmic scale was calculated and compared. The mean sensitivity is defined as in [44]:

$$\overline{S}(a,b) = \frac{1}{ln(b) - ln(a)} \int_a^b \frac{s(f)}{f} df \qquad (10)$$

where a and b are the lower and upper bound of the false positive fraction and $s(f)$ is the sensitivity at the false positive fraction f. The range $[a, b]$ was set to $\left[10^{-6}, 10^{-1}\right]$ corresponding to a wide range of operating points that can be used for further analysis by the CAD system. Statistical comparison was performed by means of bootstrapping [50]. On the test set, average ROC curves were calculated over 1,000 bootstraps, and are reported in Fig. 3. Additionally, the mean sensitivity was calculated for each bootstrap and p-values were computed for testing significance. The statistical significance level was chosen as $\alpha = 0.05$, but performance differences were considered statistically significant if $p < 0.025$ due to the Bonferroni correction[3] [51]. Comparative results are reported in Table III. In all test cases, results with the adaptive VST were statistically significantly better than without variance stabilization. The improvements in mean sensitivity $\overline{S}$ were large, on average by 2.67%, 10.97%, 7.03%, and 11.22% for

[2]except those with *CNN* which required an extensive amount of training time, so we opted for 2-fold cross validation instead.

[3]the significance level was obtained as α divided by the number of comparisons

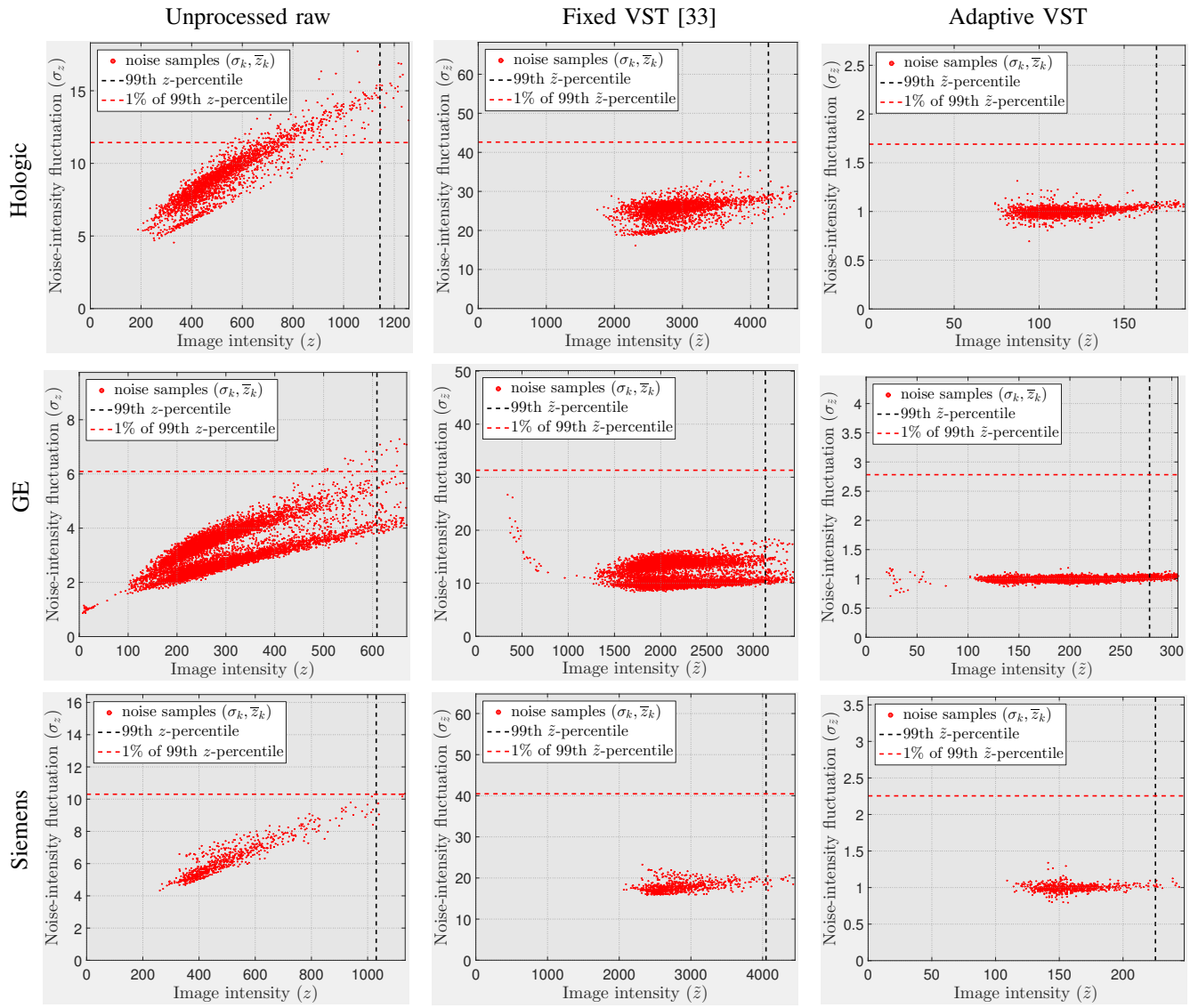

Fig. 2. Scatter plots of noise level on mammogram datasets (rows 1-3) before (column 1) and after variance stabilization (columns 2-3). Note that the numbers reported on the y-axis are not directly comparable due to the different intensity scaling. We provide vertical and horizontal dashed lines as scale-independent reference values for image intensity and noise fluctuation, respectively.

the *Cascade*, *CNN*, *SVM*, and *DoG* detectors, respectively. The adaptive VST statistically significantly outperformed the fixed VST in all but one case, with average mean sensitivity improvements of 0.81%, 0.82%, 4.52%, and 1.60% for the *Cascade*, *CNN*, *SVM*, and *DoG* detectors, respectively.

VI. Discussion and Conclusions

In this study, we adopted a one-parameter square-root noise model from which we derived an adaptive VST that stabilizes the noise variance to unitary standard deviation. It is important to emphasize that this noise model does not carry any particular assumption about the noise power spectral density. Moreover, because the VST is a pointwise transformation, the noise retains much of the spatial correlation [52], [53]. Thus, having correlated noise (as in indirect digital mammography, for example) does not cause any negative impact to the validity of our approach.

We compared the adaptive VST with a previously proposed fixed VST [33] on three datasets of digital mammograms acquired with units from different vendors. In the Hologic and GE datasets, the scatter plots of noise levels before variance stabilization (see Fig. 2, column 1) reveal the presence of two clusters which may be due to the different anode materials used. Specifically, rhodium was used in place of molybdenum to account for thicker breasts in 15% and 51% of the images for Hologic and GE, respectively. Despite this, the adaptive VST was effective in stabilizing the noise ($\overline{\text{PMSE}} \ll 1\%$), whereas the fixed transform did not accommodate for the different noise levels of the images, as indicated by the high standard deviations of the PMSE. This can also be seen by observing the scatter plots of noise levels after variance stabilization (see Fig. 2, columns 2-3), where it is noticeable that there are still two clusters for the fixed VST, whereas there is only one cluster for the adaptive VST. This result was expected, since the fixed VST is designed to rescale the

TABLE III
COMPARATIVE RESULTS OF MEAN MC DETECTION SENSITIVITY $\overline{S}$ IN THE FPR RANGE $[10^{-6}, 10^{-1}]$ FOR DIFFERENT VARIANCE STABILIZING TRANSFORMS (RAW = NONE, FIX = FIXED, ADA = ADAPTIVE). STATISTICALLY SIGNIFICANTLY DIFFERENCES ARE LISTED IN BOLD.

Dataset	**MC detector**	$\overline{S}_{RAW}$	$\overline{S}_{FIX}$	$\overline{S}_{ADA}$	$\overline{S}_{ADA} - \overline{S}_{RAW}$	$\overline{S}_{ADA} - \overline{S}_{FIX}$
Hologic	*Cascade*	67.05 (*)	69.30 (**)	69.28	$+\mathbf{2.23}$ ($p < 0.001$)	-0.03 ($p = 0.543$)
Hologic	*CNN*	60.96 (*)	67.15	67.69 (**)	$+\mathbf{6.73}$ ($p < 0.001$)	$+\mathbf{0.53}$ ($p = 0.012$)
Hologic	*SVM*	63.84 (*)	65.47	67.64 (**)	$+\mathbf{3.80}$ ($p < 0.001$)	$+\mathbf{2.17}$ ($p < 0.001$)
Hologic	*DoG*	13.76 (*)	18.99	21.41 (**)	$+\mathbf{7.65}$ ($p < 0.001$)	$+\mathbf{2.42}$ ($p < 0.001$)
GE	*Cascade*	69.27 (*)	72.30	73.47 (**)	$+\mathbf{4.20}$ ($p < 0.001$)	$+\mathbf{1.17}$ ($p < 0.001$)
GE	*CNN*	63.25 (*)	77.50	78.93 (**)	$+\mathbf{15.68}$ ($p < 0.001$)	$+\mathbf{1.43}$ ($p < 0.001$)
GE	*SVM*	63.99 (*)	64.57	66.76 (**)	$+\mathbf{2.77}$ ($p < 0.001$)	$+\mathbf{2.19}$ ($p < 0.001$)
GE	*DoG*	18.71 (*)	32.95	34.19 (**)	$+\mathbf{15.47}$ ($p < 0.001$)	$+\mathbf{1.24}$ ($p < 0.001$)
Siemens	*Cascade*	72.21 (*)	72.49	73.79 (**)	$+\mathbf{1.58}$ ($p < 0.001$)	$+\mathbf{1.30}$ ($p = 0.001$)
Siemens	*CNN*	62.48 (*)	72.49	72.98 (**)	$+\mathbf{10.50}$ ($p < 0.001$)	$+\mathbf{0.49}$ ($p = 0.015$)
Siemens	*SVM*	57.89 (*)	63.19	72.41 (**)	$+\mathbf{14.52}$ ($p < 0.001$)	$+\mathbf{9.21}$ ($p < 0.001$)
Siemens	*DoG*	21.71 (*)	31.10	32.25 (**)	$+\mathbf{10.54}$ ($p < 0.001$)	$+\mathbf{1.15}$ ($p < 0.001$)

(*) lowest $\overline{S}$ achieved in the corresponding table row
(**) highest $\overline{S}$ achieved in the corresponding table row

noise to a level that depends on the characteristics of the input image [33] whereas, in the adaptive VST, the parameter c in the denominator acts as a normalization factor that we estimate directly from the image.

The two VSTs were used as preprocessing stage of four MC detectors that rely on different training strategies and different features extracted from the images. In all test cases, MC detection performance with the adaptive VST was statistically significantly higher than without variance stabilization. The improvements in mean sensitivity were large, especially for the detectors that take in input pixel intensities (*CNN*, *SVM*, and *DoG*) and thus are more influenced by the noise, whereas *Cascade* is based on Haar-like features engineered for MC detection. In particular, *CNN* automatically learns and extracts low-level features in the first layers. We believe that stabilizing the noise variance in the input data was beneficial for those layers that captured contrast and spatial information in the 1×1 mm image patches. Consequently, this positively influenced the learning task of the subsequent layers that capture more complex features. This deserves further analysis, e.g., by testing other CNN architectures.

The results of MC detection with the adaptive VST were also statistically significantly higher than with the fixed VST. This suggests that, in addition to variance stabilization, MC detectors can also benefit from having the same noise level in all the images. Of note, the largest improvement was obtained with *SVM* which is more sensitive to variations in the data since it is trained with feature vectors directly built from patch intensities.

In this work, we have focused on the positive influence of variance stabilization on the MC detection performance. Further improvements could be obtained with a pure denoising approach, e.g., by filtering out the Gaussian noise resulting after noise stabilization [9]. However, it could require more effort to determine the correct setup of the parameters involved, including a more sophisticated analysis of the physical properties of the detector using phantom images [54]. On the contrary, the adaptive VST can be directly implemented as preprocessing modules in a full CAD scheme, and has the potential to significantly impact the detection of malignant clustered MCs.

REFERENCES

[1] S. McGuire, "World cancer report 2014," Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, 2015.

[2] American Cancer Society, "Cancer facts & figures 2016," Atlanta: American Cancer Society, 2016.

[3] J. P. Spinosa, C. Riva, P. Autier, P. Nicot, and B. Junod, "A failure analysis of invasive breast cancer: Most deaths from disease occur in women not regularly screened," *Cancer*, vol. 120, no. 18, pp. 2936–2937, 2014.

[4] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 236–251, 2009.

[5] L. H. Eadie, P. Taylor, and A. P. Gibson, "A systematic review of computer-assisted diagnosis in diagnostic cancer imaging," *European Journal of Radiology*, vol. 81, no. 1, pp. e70 – e76, 2012.

[6] N. Karssemeijer, A. M. Bluekens, D. Beijerinck, J. J. Deurenberg, M. Beekman, R. Visser, R. van Engen, A. Bartels-Kortland, and M. J. Broeders, "Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program," *Radiology*, vol. 253, no. 2, pp. 353–358, 2009.

[7] P. C. Stomper, J. Geradts, S. B. Edge, and E. G. Levine, "Mammographic predictors of the presence and size of invasive carcinomas associated with malignant microcalcification lesions without a mass," *American Journal of Roentgenology*, vol. 181, no. 6, pp. 1679–1684, 2003.

[8] H. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern Recognition*, vol. 36, no. 12, pp. 2967 – 2991, 2003.

[9] L. C. S. Romualdo, M. A. C. Vieira, H. Schiabel, N. D. A. Mascarenhas, and L. R. Borges, "Mammographic Image Denoising and Enhancement Using the Anscombe Transformation, Adaptive Wiener Filtering, and the Modulation Transfer Function," *Journal of Digital Imaging*, vol. 26, no. 2, pp. 183–197, 2013.

[10] X. Liu, C.-J. Lai, G. J. Whitman, W. R. Geiser, Y. Shen, Y. Yi, and C. C. Shaw, "Effects of exposure equalization on image signal-to-noise ratios in digital mammography: a simulation study with an anthropomorphic breast phantom," *Med. Phys.*, vol. 38, no. 12, pp. 6489–6501, 2011.

[11] M. J. Yaffe, "Digital Mammography," in *Handbook of Medical Imaging: Physics and Psychophysics*, J. Beutel, H. Kundel, and R. VanMetter, Eds. Bellingham, WA: SPIE, 2000, pp. 329–372.

[12] N. Naveed, A. Hussain, M. A. Jaffar, and T.-S. Choi, "Quantum and impulse noise filtering from breast mammogram images," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 3, pp. 1062–1069, 2012.

[13] M. T. Naseem, G. B. Sulong, and M. A. Jaffar, "MRT letter: Quantum noise removal and classification of breast mammogram images," *Microscopy research and technique*, vol. 75, no. 12, pp. 1609–1612, 2012.

[14] P. Gorgel, A. Sertbas, and O. N. Ucan, "A wavelet-based mammographic image denoising and enhancement with homomorphic filtering," *Journal of medical systems*, vol. 34, no. 6, pp. 993–1002, 2010.

[15] E. Matsuyama, D.-Y. Tsai, Y. Lee, M. Tsurumaki, N. Takahashi, H. Watanabe, and H.-M. Chen, "A modified undecimated discrete wavelet transform based approach to mammographic image denoising," *Journal of digital imaging*, vol. 26, no. 4, pp. 748–758, 2013.

[16] M. Saha, M. K. Naskar, and B. N. Chatterji, "Mammogram denoising by curvelet transform based on the information of neighbouring coefficients," in *Computer, Communication, Control and Information Technology (C3IT), 2015 Third International Conference on*. IEEE, 2015, pp. 1–6.

[17] ——, "Poisson noise removal from mammogram using poisson unbiased risk estimation technique," in *Information Systems Design and Intelligent Applications*. Springer, 2015, pp. 327–335.

[18] N. Karssemeijer, "Adaptive noise equalization and detection of microcalcification clusters in mammography," *Int. J. Pattern Recogn. Artificial Intell.*, vol. 7, no. 6, pp. 1357–1375, 1993.

[19] W. J. H. Veldkamp and N. Karssemeijer, "Normalization of local contrast in mammograms," *IEEE Trans. Med. Imaging*, vol. 19, no. 7, pp. 731–738, 2000.

[20] K. J. McLoughlin, P. J. Bones, and N. Karssemeijer, "Noise equalization for detection of microcalcification clusters in direct digital mammogram images," *IEEE Trans. Med. Imaging*, vol. 23, no. 3, pp. 313–320, 2004.

[21] G. Schie and N. Karssemeijer, "Detection of microcalcifications using a nonuniform noise model," *Proc. of the 9th International Workshop on Digital Mammography*, pp. 378–384, 2008.

[22] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, ridgelets, and curvelets for poisson noise removal," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1093–1108, 2008.

[23] M. Bartlett, "The square root transformation in analysis of variance," *Supplement to the Journal of the Royal Statistical Society*, vol. 3, no. 1, pp. 68–78, 1936.

[24] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[25] M. F. Freeman and J. W. Tukey, "Transformations related to the angular and the square root," *The Annals of Mathematical Statistics*, pp. 607–611, 1950.

[26] A. Foi, "Clipped noisy images: Heteroskedastic modeling and practical denoising," *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, 2009.

[27] M. Makitalo and A. Foi, "Optimal inversion of the anscombe transformation in low-count poisson image denoising," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 99–109, 2011.

[28] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast haar-wavelet denoising of multidimensional fluorescence microscopy data," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. IEEE, 2009, pp. 310–313.

[29] U. Bagci and D. J. Mollura, "Denoising pet images using singular value thresholding and steins unbiased risk estimate," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 115–122.

[30] A. Zhang, H. Jiang, L. Ma, Y. Liu, and X. Yang, "A shearlet-based algorithm for quantum noise removal in low-dose ct images," International Society for Optics and Photonics, pp. 9784 – 9784 – 7, 2016.

[31] M. A. Vieira, P. R. Bakic, A. D. Maidment, H. Schiabel, and N. D. Mascarenhas, "Filtering of poisson noise in digital mammography using local statistics and adaptive wiener filter," in *International Workshop on Digital Mammography*. Springer, 2012, pp. 268–275.

[32] L. R. Borges, P. R. Bakic, A. Foi, A. D. Maidment, and M. A. Vieira, "Pipeline for effective denoising of digital mammography and digital breast tomosynthesis," International Society for Optics and Photonics, pp. 10 132 – 10 132 – 11, 2017.

[33] A. Bria, C. Marrocco, J.-J. Mordang, N. Karssemeijer, M. Molinara, and F. Tortorella, "LUT-QNE: Look-Up-Table Quantum Noise Equalization in Digital Mammograms," in *International Workshop on Digital Mammography*. Springer, 2016, pp. 27–34.

[34] U. Bick and F. Diekmann, *Digital mammography*. Springer Science & Business Media, 2010.

[35] J. Wang, Y. Yang, and R. M. Nishikawa, "Quantitative study of image features of clustered microcalcifications in for-presentation mammograms," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3404–3408.

[36] A. Burgess, "On the noise variance of a digital mammography system," *Med. Phys.*, vol. 31, no. 7, pp. 1987–1995, 2004.

[37] W. H. Green, "Econometric analysis 5th edition," p. 913, 2003.

[38] K. Hirakawa and T. W. Parks, "Image denoising using total least squares," *IEEE Transactions on image processing*, vol. 15, no. 9, pp. 2730–2742, 2006.

[39] J. H. Curtiss, "On transformations used in the analysis of variance," *The Annals of Mathematical Statistics*, vol. 14, no. 2, pp. 107–122, 1943.

[40] A. Foi, "Optimization of variance-stabilizing transformations," 2009. [Online]. Available: http://www.cs.tut.fi/~foi/papers/Foi-OptimizationVST-preprint.pdf

[41] A. Bria, N. Karssemeijer, and F. Tortorella, "Learning from unbalanced data: A cascade-based approach for detecting clustered microcalcifications," *Medical Image Analysis*, vol. 18, no. 2, pp. 241–252, 2014.

[42] A. Bria, C. Marrocco, N. Karssemeijer, M. Molinara, and F. Tortorella, "Deep cascade classifiers to detect clusters of microcalcifications," in *International Workshop on Digital Mammography*. Springer, 2016, pp. 415–422.

[43] A. Bria, C. Marrocco, M. Molinara, and F. Tortorella, "An effective learning strategy for cascaded object detection," *Information Sciences*, vol. 340, pp. 17–26, 2016.

[44] J.-J. Mordang, T. Janssen, A. Bria, T. Kooi, A. Gubern-Mérida, and N. Karssemeijer, "Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks," in *International Workshop on Digital Mammography*. Springer, 2016, pp. 35–42.

[45] I. El Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Med. Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[46] J. Dengler, S. Behrens, and J. Desaga, "Segmentation of microcalcifications in mammograms," *Medical Imaging, IEEE Transactions on*, vol. 12, no. 4, pp. 634–642, Dec 1993.

[47] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[48] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[49] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[50] F. W. Samuelson and N. Petrick, "Comparing image detection algorithms using resampling," in *IEEE Int. Symp. Biomed. Imag.*, 2006, pp. 1312–1315.

[51] O. J. Dunn, "Multiple Comparisons Among Means," *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52–64, 1961.

[52] L. R. Borges, I. Guerrero, P. R. Bakic, A. Foi, A. D. A. Maidment, and M. A. C. Vieira, "Method for simulating dose reduction in digital breast tomosynthesis," *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, pp. 2331–2342, 2017.

[53] L. Azzari and A. Foi, "Variance stabilization in poisson image deblurring," *Proc. 2017 IEEE Int. Sym. Biomedical Imaging (ISBI), Melbourne, Australia*, 2017.

[54] C. E. Tromans, M. R. Cocker, and M. Brady, "Quantification and normalization of x-ray mammograms," *Physics in Medicine and Biology*, vol. 57, no. 20, p. 6519, 2012.