Malaysian Journal of Computing, 7 (1): 1047-1055, 2022

eISSN: 2600-8238 online

RESTORATION OF OLD MALAY JAWI MANUSCRIPTS USING MUMFORD-SHAH AND BERTALMIO INPAINTING MODELS

Nurul Nabilah Zainal[1], Nur Fatini Mohammad Yuri[2] and Abdul Kadir Jumaat[3,4*]

Department of Mathematics, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[4]Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia

[1]2018226316@isiswa.uitm.edu.my, [2]2018200908@isiswa.uitm.edu.my, [3,4*]abdulkadir@tmsk.uitm.edu.my

ABSTRACT

Jawi is the earliest writing script of the Malay Archipelago that was derived from Arabic letters. Old Malay Jawi manuscripts may provide vital information about the legacy, cultures, and historical evolution of the Malay Archipelago over time. However, while preserved, old Malay Jawi manuscripts tend to be damaged. Image inpainting is a process of reconstructing missing parts of an image which can be used to restore the old Malay Jawi manuscripts. The Mumford-Shah and Bertalmio inpainting models are two well-known and effective methods for solving the image inpainting problem. Hence, the aim of this study is to determine which model is better at restoring corrupted input images of the old Malay Jawi manuscripts. A sample of thirty (30) old Malay Jawi manuscript images were obtained from Kumpulan Penyelidikan Etnomatematik Melayu (KUPELEMA). The corrupted images were restored using both models, implemented using the MATLAB software. The Structural Similarity Index Measure (SSIM) and Mean Absolute Error (MAE) were utilized to assess the quality of the results. The numerical experiment demonstrates that the average values of SSIM and MAE for Mumford-Shah inpainting model are 0.9380 and 0.0151 respectively, while the values for the Bertalmio inpainting model are 0.8762 and 0.0255 respectively. This indicates that the Mumford-Shah inpainting model is more effective than Bertalmio inpainting model. The algorithms used in this study can be upgraded to a software framework for commercial use and can be implemented for other kinds of digitized data.

Keywords: *Bertalmio Model, Image Inpainting, Mumford-Shah Model, Old Malay Jawi Manuscripts.*

Received for review: 20-08-2021; Accepted: 21-03-2022; Published: 31-03-2022

1. Introduction

Thousands of old manuscripts on socio-economic growth, government, and the history of the Malay archipelagos have been penned in Jawi. Jawi is the earliest writing script that was derived from Arabic letters but has extra letters in the Malay spoken language to fit certain local pronunciation. Years before the arrival of Roman letters in countries around the world, Jawi was the lingua franca in the Malay Archipelago. According to Razak (2016), old manuscripts or ancient texts often contribute to knowledge and are therefore an extremely valuable source for discovering heritage, cultures, and Malaysia's historical development.

Manuscripts, which have existed for centuries in antiquity, have considerable significance. In general, old manuscripts provide details about the culture and values of civilizations over the centuries (Rocha *et al*., 2018). According to Kaur *et al*. (2020), manuscripts came into existence from the moment human beings discovered the need to keep a record of things.

According to an article by Halid (2020), a group of national ethnomathematics researchers found 34 Malay manuscripts related to science and mathematics in the 17th century, which were believed to help improve the education system, science, technology, engineering and mathematics. The director of the Institute of Mathematical Research (INSPEM), Universiti Putra Malaysia (UPM), Associate Professor Dr Muhammad Rezal Kamel Ariffin, said that all the manuscripts were found in Acheh, Indonesia and Champa, Cambodia, as well as in several locations around the country that are now stored digitally at the institute.

Currently, these old manuscripts are stored as a resource for modern day in galleries. However, while preserved, the manuscripts tend to be damaged. Fortunately, the fragmented part can be repaired via image processing to reduce the issue of analyzing, interpreting, and transliterating information from the old manuscripts. Image processing is a technique whereby such operations are performed on an image to produce and improve the image or to derive valuable data from it.

There are many branches of image processing, such as image segmentation (Jumaat & Chen, 2018; Jumaat & Chen, 2019), image denoising (Dang & Phan, 2020) and image inpainting (Zhao *et al*., 2019; Zhang *et al*., 2020). For this research, image inpainting is the focused. Image inpainting refers to a filling-in phase of incomplete details in a visual input. The aim of the phase is to recreate incomplete parts or broken images in a way that a casual observer will not be able to detect the inpainted area.

There are many researchers had discussed on image inpainting in literature. For instance, Zhao *et al*. (2019) and Zhang *et al*. (2020) applied the learning-based approach for image inpainting. Despite having good results, however, these learning based approaches may depend too muchon the amount of data. Other past studies have also incorporated user guides to enhance image inpainting, such as boundaries line, extended directions, guide regions, as well as several image exemplars, such as Ashikhmin (2001), Hays and Efros (2007), Huang *et al*. (2013) and Yu *et al*. (2019). However, these principles are mainly structured and are not content based.

The Mumford and Shah (1989) and Bertalmio (2001) inpainting methods are found to be popular and widely used in the field of image inpainting. The Mumford-Shah model is an especially well-liked and efficient numerical approach for solving image segmentation problems as well as problems of image restoration such as inpainting (Thanh *et al*., 2019). For instance, Tsai *et al*. (2001) used the model in the implementation of curve evolution. Later, Chan and Shen (2002) applied the model to construct a mathematical model for local non-texture image inpainting. The research from Bonneel *et al*. (2018) implemented the Mumford-Shah mesh using the Ambrosio-Tortorelli functional.

The intention of using this model for image inpainting originates in the studies of Tsai *et al*. (2001), Chan and Shen (2002), and Esedoglu and Shen (2002). In the work of Tsai *et al*. (2001), the model is among the key instances and implementations for the algorithm of numerical active contour, which was then designed to total variation inpainting as an alternative in the work of Chan and Shen (2002). Esedoglu and Shen (2002) further considered the Mumford-Shah image model using the approximation of Ambrosio-Tortorelli (Ambrosio & Tortorelli, 1990) to solve the numerical solution. The approximation of Ambrosio-Tortorelli contributes to an easy inpainting algorithm with fast numerical convergence.

For the Bertalmio inpainting model, Barbu (2018) used the model using an iterative discretization algorithm that numerically approximates the solution that represents the inpainted image. Rana *et al*. (2017) applied the model to estimate 2D smoothness estimation with the Laplacian equation. According to Barbu (2018), Bertalmio introduced an influential anisotropic diffusion model for image inpainting. When an image that contains damaged or missing regions is detected, it can be interpreted as a two-dimensional partial function identified only outside those areas. Thus, during the inpainting process, the image evolves, which considers a nonlinear second-order anisotropic diffusion model with boundary conditions that recovers the original image from the image observed.

Due to the strengths of the Mumford-Shah and Bertalmio models, this study aims to investigate their performance in restoring the corrupted parts of the old Malay Jawi manuscripts obtained from *Kumpulan Penyelidikan Etnomatematik Melayu* (KUPELEMA, 2021), an active research group under INSPEM that analyzes old manuscripts. Interested researchers can access their collection of the old manuscripts, including old Jawi manuscripts, on their website. Both methods were programmed and simulated using the MATLAB software. Then, the restoration performances were evaluated and compared to identify the better approach between the two in restoring old Malay Jawi manuscripts. This is done by computing Structural Similarity Index Measure (SSIM) and Mean Absolute Error (MAE).

The remaining section of this paper is structured as follows. In Section 2, the methodology of this research is reported. In Section 3, the experimental result and discussion are presented. The conclusion and recommendations are drawn in Section 4.

2. Methodology

Figure 1 shows the flow of the methodology. Each phase is discussed in the following subsections.

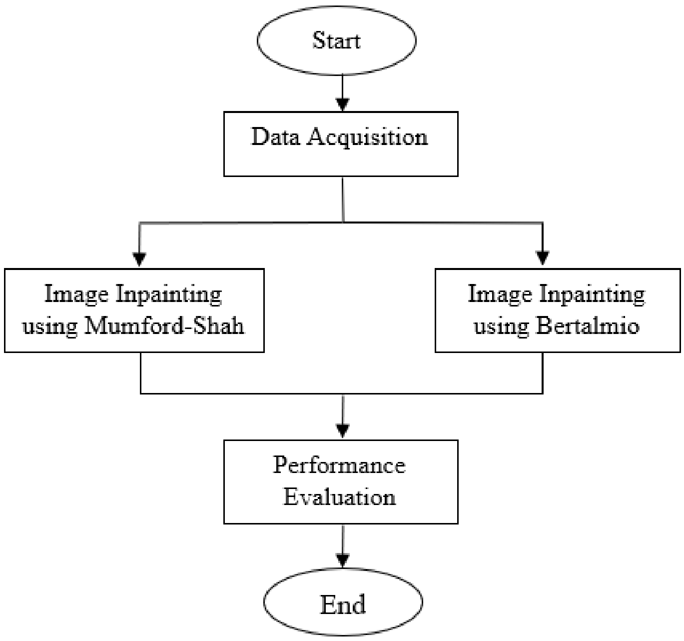


Figure 1. Research Methodology Framework.

2.1 Data Acquisition

Thirty (30) images of old Malay Jawi manuscripts data that were used in this project were obtained from *Kumpulan Penyelidikan Etnomatematik Melayu* (KUPELEMA, 2021). The images then are cropped and resized to 241×96 pixels.

2.2 Image Inpainting using Mumford-Shah Model

The Mumford-Shah inpainting model is given in Equation (1).

$$\min_{u,\Gamma} J(u,\Gamma) = \frac{\gamma}{2}\int_{\Omega\setminus\Gamma} |\nabla u|^2 \, dx + \alpha.Length(\Gamma) + \frac{\lambda}{2}\int_{\Omega\setminus D} (u-g)^2 \, dx \tag{1}$$

where γ, α, and λ are positive weight and the chosen values are 0.5, 1, and 10^9 respectively. This model produced two outputs which were the inpainted (restored) image u, and its corresponding edge image Γ from the input image g. The first two terms ensured the smooth output while the third term avoided the over smoothing effect. The model was effectively solved using elliptic solver and iterative scheme. Details of the algorithm can be found in Tsai *et al.* (2001), Chan and Shen (2001), Esedoglu and Shen (2002) and Schonlieb (2015a) and the algorithm was implemented in the MATLAB software.

2.3 Image Inpainting using Bertalmio Model

The Bertalmio inpainting model is defined as Equation (2).

$$u_t = \nabla^{\perp} u.\Delta\nabla u + \nabla.\left(g\left(|\nabla u|\right)\nabla u\right) \tag{2}$$

The first term was the transport equation that was used to ensure smoothness in the solution, while the second term was the anisotropic diffusion equation to prevent level lines from crossing. The model was solved iteratively using finite difference method. Details of the algorithm is discussed in Bertalmio (2001) and Schonlieb (2015b). The MATLAB software was used to implement the model.

2.4 Performance Evaluation

At this stage, the quality of the restoration output by the Mumford-Shah and Bertalmio methods was examined using Structural Similarity Index Measure (SSIM) and Mean Absolute Error (MAE).

To measure the similarity between the original images and restored images, SSIM method was applied. The SSIM values ranged from -1 to 1, with 1 indicating complete structural similarity between two images and -1 indicating poor structural similarity. The SSIM formula is defined as Equation (3).

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_x\sigma_y + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{3}$$

where μ_x and μ_y are the average of x and y respectively, σ_x and σ_y are the variance of x and y respectively while C_1 and C_2 are two variables to stabilize the division with weak denominator.

MAE method was used to measure the average or sum of the absolute error between two images which are the original images and restored images. The image produced will be of higher quality if the value is lower. The MAE formula is defined as Equation (4).

$$MAE(x,y)=\frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\left|x(i,j)-y(i,j)\right| \tag{4}$$

where $x(i,j)$ and $y(i,j)$ are the original and the restored image respectively.

3. Result and Discussion

In this section, the overall restoration results for both inpainting models evaluated using SSIM and MAE are discussed.

SSIM was used to measure the index similarity between the original image and the restored image. A decimal value between -1 to 1 is the corresponding SSIM index and a positive value shows the case of two data sets which are highly similar. The higher the value, the better the image produced. MAE was the absolute error between the clean image and the inpainted i.e. the restored image obtained. The lesser value shows the better quality of the image produced. We illustrate the comparison results of the Mumford-Shah and Bertalmio inpainting models for the first three data in Table 1.

Table 1. Image Inpainting using Mumford-Shah and Bertalmio Models.

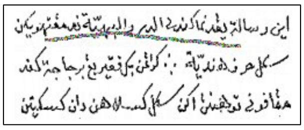
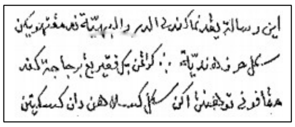
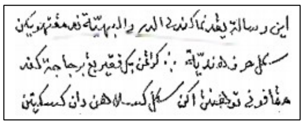
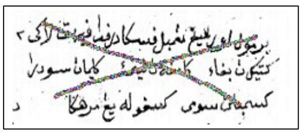
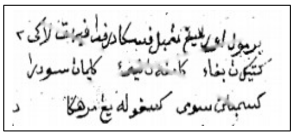
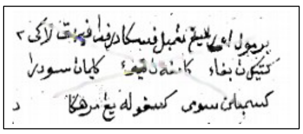
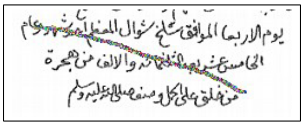
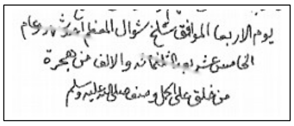
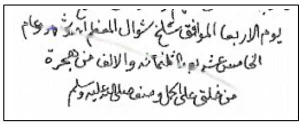
Corrupted Image	**Inpainting Model**	
	Mumford-Shah	**Bertalmio**
	 SSIM=0.9958, MAE=0.0011	 SSIM=0.9824, MAE=0.0019
	 SSIM=0.9702, MAE=0.0082	 SSIM =0.9522, MAE =0.0095
	 SSIM =0.9865, MAE=0.0040	 SSIM =0.9740, MAE =0.0049

Table 1 shows that both models were able to restore the corrupted images. By comparing the values of SSIM and MAE for both models, it shows that the value of SSIM for the Mumford-Shah model had a higher value than Bertalmio, while the value of MAE for the Mumford-Shah model was lower than the Bertalmio model for the first three illustrated images in Table 1.

Table 2: The Values of SSIM and MAE.

Corrupted Test Images	**Inpainting Model**			
	Mumford-Shah	**Bertalmio**	**Mumford-Shah**	**Bertalmio**
	SSIM		**MAE**	
Data 1	0.9958	0.9824	0.0011	0.0019
Data 2	0.9702	0.9522	0.0082	0.0095
Data 3	0.9865	0.9740	0.0040	0.0049
Data 4	0.9948	0.9929	0.0017	0.0020
Data 5	0.9943	0.9877	0.0019	0.0025
Data 6	0.9870	0.9797	0.0040	0.0042
Data 7	0.9862	0.9637	0.0047	0.0062
Data 8	0.7739	0.6175	0.0436	0.0756
Data 9	0.9438	0.9066	0.0132	0.0218
Data 10	0.9800	0.9746	0.0071	0.0076
Data 11	0.9286	0.8594	0.0168	0.0289
Data 12	0.8878	0.7587	0.0219	0.0481
Data 13	0.8821	0.7552	0.0224	0.0454
Data 14	0.8688	0.7633	0.0264	0.0532
Data 15	0.8657	0.6879	0.0287	0.0634
Data 16	0.8509	0.6693	0.0327	0.0660
Data 17	0.8818	0.7155	0.0274	0.0589
Data 18	0.9628	0.9337	0.0110	0.0156
Data 19	0.9511	0.9272	0.0139	0.0169
Data 20	0.9485	0.8896	0.0116	0.0200
Data 21	0.9783	0.9457	0.0060	0.0112
Data 22	0.9364	0.8741	0.0215	0.0286
Data 23	0.9458	0.9237	0.0138	0.0182
Data 24	0.9609	0.9298	0.0108	0.0136
Data 25	0.9544	0.8531	0.0119	0.0238
Data 26	0.9384	0.8984	0.0164	0.0226
Data 27	0.9265	0.8400	0.0255	0.0337
Data 28	0.9356	0.8719	0.0209	0.0299
Data 29	0.9701	0.9386	0.0100	0.0136
Data 30	0.9540	0.9207	0.0132	0.0164
Average	0.9380	0.8762	0.0151	0.0255

Illustrated in Table 2, the values of SSIM and MAE for both models are recorded for all tested data. In addition, the average values of SSIM and MAE are computed to show the overall performance. Mumford-Shah inpainting model produced a better image quality as it had a higher average value of SSIM and a lower average value of MAE compared to the Bertalmio inpainting model.

4. Conclusion and Recommendation

In this paper, the performance of the Mumford-Shah and Bertalmio inpainting models in restoring old Malay Jawi manuscripts are compared. Thirty (30) image datasets were used to evaluate the efficacy of both inpainting models. Furthermore, two types of performance indicators which were Structural Similarity Index Measure (SSIM) and Mean Absolute Error (MAE) were used to examine the quality of restoration results, which were the inpainted images. All the algorithms were simulated using MATLAB software.

The results of this study demonstrated that the Mumford-Shah inpainting model produced better image quality than the Bertalmio inpainting model because it has a higher SSIM and lower MAE values. This study may be beneficial to KUPELEMA to evaluate which inpainting method is effective for the digitized old manuscripts that produced the best output results. This will allow INSPEM to retrieve the manuscripts and reduce the issue of the transliteration process.

Based on the findings of this study, there are several recommendations for future research. In order to achieve better results, the output images can be subjected to a threshold process to remove the blurred part of the image before undergoing SSIM and MAE. On top of that, this study can be a foundation knowledge in applying and evaluation other inpainting methods. This model can also be implemented for other kinds of digitized data in 2D and 3D. For large data, the optimization method by Jumaat and Chen (2017) and Jumaat and Chen (2020) can be applied. In addition, it is worth to model simultaneously inpainting and segmentation model based on the works from Abdullah and Jumaat (2022), Zaman *et al*. (2015), Jumaat *et al*. (2014), Yasiran *et al*. (2012), Jumaat *et al*. (2012) and Yasiran *et al*. (2011). Besides that, the algorithms used in this study is suggested to be upgraded to a software framework for commercial use.

References

Abdullah, S. A., & Jumaat, A. K. (2022). Selective image segmentation models using three distance functions. *Journal of Information and Communication Technology,* 21, 95-116.

Ambrosio, L., & Tortorelli, V. M. (1990). Approximation of functionals depending on jumps by elliptic functionals via gamma convergence. *Communications on Pure and Applied Mathematics*, 43, 999-1036.

Ashikhmin, M. (2001). Synthesizing natural textures. In *Symposium on Interactive 3D graphics, 2001* (pp. 217-226). SIGGRAPH.

Barbu, T. (2018). Second-order anisotropic diffusion-based framework for structural inpainting. *Proceedings of the Romanian Academy, Series A*, 19(2), 329-336.

Bertalmio, M. (2001). *Processing of flat and nonflat image information on arbitrary manifolds*. [Ph.D. thesis, The University of Minnesota]. Citeseer, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.957&rep=rep1&type=pdf.

Bonneel, N., Coeurjolly, D., Gueth, P., & Lachaud, J. O. (2018). Mumford-Shah mesh processing using the Ambrosio-Tortorelli functional. *Computer Graphics Forum*, 37(7), 75-85.

Chan, T. F., & Shen, J. (2002). Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics*, 62, 1019-1043.

Dang, T., & Phan, K. (2020). A weighted total variation based image denoising model using mean curvature. *Optik,* 217, 164940.

Esedoglu, S., & Shen, J. (2002). Digital inpainting based on the Mumford-Shah-Euler image model. *European Journal of Applied Mathematics*, 13, 353-370.

Halid, S. (2020). 34 Manuskrip Simpan Rahsia Sains, Matematik. *A post at Berita Harian available at https://upm.edu.my/berita/manuskrip sains_dan _matem- atik _tamadun melayu_abad_ke_17_ditemui_metrotv-58717?L=en.*

Hays, J., & Efros, A. A. (2007, May). Scene completion using millions of photographs. In *Computer Graphics Proceedings, Annual Conference Series 2007* (pp. 1-7). SIGGRAPH.

Huang, J.B., Kopf, J., Ahuja, N., & Kang, S. B. (2013). Transformation guided image completion. In *IEEE International Conference on Computational Photography (ICCP)* (pp. 1-9). IEEE.

Jumaat, A. K., & Chen, K. (2020). Three-dimensional convex and selective variational image segmentation model. *Malaysian Journal of Mathematical Sciences*, 14(3), 437-450.

Jumaat, A. K., & Chen, K. (2019). A reformulated convex and selective variational image segmentation model and its fast multilevel algorithm. *Numerical Mathematics Theory Methods and Applications,* 12(2), 403-437. http://dx.doi.org/10.4208/nmtma.OA-2017-0143.

Jumaat, A. K., & Chen, K. (2018). Fast algorithm for selective image segmentation model. *International Journal of Engineering and Technology (UAE)*, 7(4), 41-44.

Jumaat, A. K., & Chen, K. (2017). An optimization-based multilevel algorithm for variational image segmentation models. *Electronic Transactions on Numerical Analysis*, 46, 474-504.

Jumaat, A. K., Yasiran, S. S., Malek, A. A., Rahman, W. E. Z. W. A., Badrin, N., Osman, S. H., Rafiee, S. R., & Mahmud, R. (2014). Performance comparison of Canny and Sobel edge detectors on Balloon Snake in segmenting masses. In J. Jaafar, L. A. Rahim, & E. E. Mustapha (Eds.), *International Conference on Computer and Information Sciences (ICCOINS)* (pp. 116-120). IEEE. https://doi.org/10.1109/ICCOINS.2014.6868368.

Jumaat, A. K., Rahman, W. E., Ibrahim, A., Yasiran, S. S., Mahmud, R., & Malek, A. A. (2012). Masses characterization based on angular margin measurement. In D. Al-Dabass, Z. Ibrahim, K. H. Ghazali, & S. W. Mawawi (Eds.), *International Conference on Computational Intelligence, Modelling and Simulation* (pp. 265-269). IEEE. https://doi.org/10.1109/CIMSim.2012.50.

Kaur, A., Raj, A., & Indu, S. (2020). Inpainting of Irregular Holes in a Manuscript using UNet and Partial Convolution. In *Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 778-784). IEEE.

KUPELEMA (2021). Kumpulan Penyelidikan Etnomatematik Melayu. *A post at INSPEM available at https://inspem.upm.edu.my/ laboratori/ laboratori_etnomatematik_ dan_didaktik/kumpulan_penyelidikan_etnomatematik_melayu_kupelema-54857.*

Mumford, D., & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Ap- plied Mathematics*, 42(5), 577-685.

Rana, M. S., Hassan, M. M., & Bhuiyan, T. (2017). An Enhanced Model for Inpainting on Digital Images Using Dynamic Masking. *Journal of Communications*, 12(4), 248-253.

Razak, Z. (2016). *Old Jawi manuscript: Digital recognition* [Ph.D. thesis, University of Malaya]. UM, http://studentsrepo.um.edu.my/6757/.

Rocha, C., Deborah, H., & Hardeberg, J. Y. (2018). Ink bleed-through removal of historical manuscripts based on Hyperspectral imaging. *Lecture Notes in Computer Science* (pp. 473-480). Springer.

Schonlieb, C.B. (2015a). *The Mumford-Shah Image Model for Inpainting. In Partial Differential Equation Methods for Image Inpainting*. Cambridge: Cambridge University Press.

Schonlieb, C.B. (2015b). *Transport Inpainting. In Partial Differential Equation Methods for Image Inpainting.* Cambridge: Cambridge University Press.

Thanh, D. N. H., Prasath, V. B. S., Son, N. V., & Hieu, L. M. (2019). An adaptive image inpainting method based on the modified mumford-shah model and multiscale parameter estimation. *Computer Optics*, 43(2), 251-257.

Tsai, A., Yezzi, A., Jr., & Willsky, A. S. (2001). Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Transactions on Image Processing*, 10, 1169-1186.

Yasiran, S. S., Jumaat, A. K., Abdul Malek, A., Hashim, F. H., Dhaniah Nasrir, N., Azirah Sayed Hassan, S. N., Ahmad, N., & Mahmud, R. (2012). Microcalcifications segmentation using three edge detection techniques. In M. R. Mahmood, U. M. Noor, & M. K. Hamzah (Eds.), *International Conference on Electronic Devices, Systems, and Applications* (pp. 207-211). IEEE. https://doi.org/10.1109/ICEDSA.2012.6507798.

Yasiran, S. S., Jumaat, A. K., Manaf, M., Ibrahim, A., Wan Eny Zarina, W. A. R., Malek, A., Laham, M. F., & Mahmud, R. (2011). Comparison between GVF snake and ED snake in segmenting microcalcifications. In M. K. Hamzah, W. Mansor, & N. M. Tahir (Eds.), *International Conference on Computer Applications and Industrial Electronics* (pp. 597-601). IEEE. https://doi.org/10.1109/ICCAIE.2011.6162204.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. (2019). Free-form Image Inpaint- ing with Gated Convolution. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 4470-4479). IEEE.

Zaman, N. A. K., Rahman, W. E. Z. W. A., Jumaat, A. K., & Yasiran, S. S. (2015). Classification of breast abnormalities using artificial neural network. Paper presented at the AIP Conference Proceedings,1660, 1-7. https://doi.org/10.1063/1.4915671.

Zhang, L., Chen, Q., Hu, B., & Jiang, S. (2020). Text-guided neural image inpainting. In *28th ACM International Conference on Multimedia* (pp. 1302-1310). ACM.

Zhao, Y., Price, B., Cohen, S., & Gurari, D. (2019). Guided Image Inpainting: Replacing an image region by pulling content from another image. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1514-1523). IEEE.