Image Prefiltering in DeepFake Detection

Szymon Motłoch[1], Mateusz Szczygielski[1] and Grzegorz Sarwas[2][a]
[1]*Faculty of Electrical Engineering, Warsaw University of Technology, Warsaw, Poland*
[2]*Institute of Control and Industrial Electronics, Warsaw University of Technology, Warsaw, Poland*

Keywords: Deepfake, Fractional Order Derivative, Image Preprocessing, SRM Filter.

Abstract: Artificial intelligence, becoming common technology, creates a lot of new possibilities and dangers. An example can be open source applications that enable swapping faces on images or videos with other faces delivered from other sources. This type of modification is named DeepFake. Since the human eye cannot detect DeepFake, it is crucial to possess a mechanism that would detect such changes.

This paper analyses solution based on Spatial Rich Models (SRM) for image prefiltering connecting convolutional neural network VGG16 to increase DeepFake detection with neural networks. For DeepFake detection, a fractional order spatial rich model (FoSRM) is proposed, which was compared with classical SRM filter and integer order derivative operators. In the experiment, we used two different approximation fractional order derivative methods: first based on the mask and second used Fast Fourier Transform (FFT). Achieved results we also compare with the original ones and the VGG16 network with an additional layer added to select the parameters of the prefiltering mask automatically.

As a result of the work, we questioned the legitimacy of using additional image enrichment by prefiltering when using the convolutional neural network. Additional network layer gave us the best results from the performed experiments.

1 INTRODUCTION

The emergence of convolutional neural networks allowed their use in the field of image processing and computer vision. Along with the convolution operation structures, a completely new range of possibilities appeared, mainly supporting analyzing the vision scene. Not only the classification of images, but most of all the detection of objects on them has become a problem that neural networks are able to cope with even better than humans. The development of autoencoders (Garcia et al., 2017; Chen et al., 2017) and the Generative Adversarial Network (GAN) (Yi et al., 2019; Ledig et al., 2017) technology caused that the neural network became capable to transform the styles of a given image or replace part of the content of an image or video sequence. Like professional painters, the developed methods can change the entire structure of the image so that the result is beyond recognition for an average observer. The further development of these methods resulted in creating a DeepFake modification, enabling the replacement of a face from the original photo or video with a face provided from another source. Swapped faces look very realistic, often showing emotions or behaviors, making the obtained results practically undetectable to the human eye. Consequently, this technology has become very dangerous. For example, stock exchange quotations are heavily dependent on statements and events attended by important and influential people. The creation of these types of manipulated videos can cause a sudden change in stock prices. Another example of a threat is discrediting other people. One possibility here is to substitute the person's face in a pornographic film. For example, the preparation of such a video may affect the results of national elections.

The described situations show that the detection of DeepFake modifications has become a crucial task (Mirsky and Lee, 2021). As a result, there is an increase in interest in this, which affects the emergence of many solutions aimed at detecting changes in images or video sequences made by algorithms. Deep neural networks have been trained for this problem that perform well, but there is no solution for 100% reliability. It is evident when trying to detect modifications in images, where the achieved effectiveness

[a] https://orcid.org/0000-0003-4113-2387

Motłoch, S., Szczygielski, M. and Sarwas, G.
Image Prefiltering in DeepFake Detection.
DOI: 10.5220/0010841200003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 4: VISAPP*, pages 476-483
ISBN: 978-989-758-555-5; ISSN: 2184-4321

is lower than operating on video sequences. For this reason, algorithms are constantly being improved that allow for the best solution to this problem.

One of the methods for increasing the effectiveness of the neural network presented in the literature is preliminary filtering of the input data, aimed at strengthening the relevant information. Some papers confirm the efficacy of such action (Chang et al., 2020; Han et al., 2021; Liu et al., 2020). However, keep in mind that a significant portion of your filters or operations can backfire. Incorrectly used prefiltering may reduce the resolution of photos or remove information necessary from the point of view of the neural network.

Based on the solution presented by (Chang et al., 2020), who used Spatial Rich Models (SRM) connecting with VGG16 for the detection DeepFake modifications in photos, we analyze various methods of prefiltering images to determine their usefulness. Apart from the classic SRM known from the literature, we propose a novel solution based on the fractional order operator called Fractional order Spatial Rich Model (FoSRM). Two different types of fractional order derivative approximations were examined to choose the best order of fractional operator and the most efficient approximator. Achieved results we also compare with the original ones and the VGG16 network with an additional layer added to automatically select the parameters of the prefiltering mask.

The organization of the document is as follows. In the next chapter, the solutions and results obtained in other works on this issue are described. Section 3 discusses the algorithms used in the research. In section 4 described performed experiments and obtained results. The last chapter summarizes our research.

2 RELATED WORK

In (Rana and Sung, 2020) and (Wang et al., 2018) authors took several steps to improve the quality of the input data. The first operation was to excise only the face from the analyzed images. This step uses a face detection algorithm for cutting out only the indicated area to focus only on an essential region of interest. After applying these operations, the cut image is scaled to the same size.

In (Zi et al., 2021) as part of data preprocessing, authors extract face landmarks to align all the faces in a face sequence, which allows being robust on changing face orientations.

In (Chang et al., 2020) used the SRM filter for image preprocessing. Such a procedure amplifies the noise, which contains information that can improve the effectiveness of the network. The applied filtering operation made it possible to detect information that was not visible in the RGB channels before use.

In (Han et al., 2021) also used the SRM filter as a data preprocessing method in this problem. However, in this case, the direct application of this operation did not give satisfactory results. The authors of the publication say that this filter can significantly impact the detection of pasted objects, but not the detection of DeepFake manipulation. The reason is the complicated nature of such a transformation. As a result, is achieved the opposite effect to that obtained in the previously described work. This publication also proposes a filter called ”learnable SRM”. It was shown that applying such a method in a two-channel network improved the obtained results. According to the authors of the publication, this method increased the effectiveness of the neural network by 2% compared to other proposed methods.

In (Zhang et al., 2018) was shown that using the SRM filter is very useful for detecting pasted objects in photos. Besides, this operation performed well in the steganalysis task.

In (Younus and Hasan, 2020) confirmed that detection of DeepFake modifications based on the sharpening of edges is an effective method. The Haar transform was used to strengthen the edges.

In (Liu et al., 2020) investigated several aspects of input preparation for this problem. Their research checked whether the excision of fragments containing only facial skin is sufficient to detect modifications. It was also verified whether the algorithm's effectiveness would change for the black and white versions of the input photos. In addition, the effect of applying the L0 filter to minimize the gradient was investigated. The results obtained showed that the regions containing the skin contain enough information to detect facial modifications efficiently. The results were similar to those obtained with standard full-face excision methods. It was also shown that for black and white images, the effectiveness was very similar and only slightly decreased in detecting modifications in color picture. Using the L0 filter as a method of initial filtering of images resulted in a significant deterioration of the results, reducing the AUC by 0.2 compared to the tests carried out without using this filter.

In (Guo et al., 2020) proposed the Adaptive Manipulation Traces Extraction Network (AMTEN) as a method to enhance the modifications found in the input data. The results obtained in the research showed that the applied data preprocessing method improved the effectiveness of DeepFake transformation detection.

In this paper, the influence of SRM filters on the operation of the convolutional neural network used to detect DeepFake modifications in the photos will be tested. The impact of using non-integer order derivatives as a data preprocessing method will also be investigated. It was decided to use such an operation due to its ability to sharpen the edges of the image, which are described as key in detecting this type of modification. This type of mechanism also implies the enhancement of high-frequency information, which may contain information essential for this problem.

3 ALGORITHMS

This section describes algorithms used for preprocessing input to VGG16 (Simonyan and Zisserman, 2015). The reason for choosing this network model was a possibility for comparing the results of the experiment described in the article (Chang et al., 2020). This study examined the effect of image enhancement and one of the SRM filters on the detection performance of DeepFake modifications. This paper will extend the analysis of the influence of image preprocessing algorithms on the neural network's effectivity in the analyzed problem.

3.1 SRM Filter

SRM filter can extract local noise from image (Zhou et al., 2018). This method uses various types of high pass filters. Before using the convolutional neural network for steganalysis, this was the best method used for solving this problem (Kang et al., 2019).

In this paper various types of SRM filters were examined. Firstly, following directional masks of sizes 3×3 were analysed: horizontal (3), vertical (1), left-diagonal (2) and right-diagonal (4) (Reinel et al., 2021). In experiments another type of SRM filter, proposed in (Reinel et al., 2021) (5) was used.

$$\frac{1}{2}\begin{bmatrix}0 & 1 & 0\\ 0 & -2 & 0\\ 0 & 1 & 0\end{bmatrix} \tag{1}$$

$$\frac{1}{2}\begin{bmatrix}0 & 0 & 1\\ 0 & -2 & 0\\ 1 & 0 & 0\end{bmatrix} \tag{2}$$

$$\frac{1}{2}\begin{bmatrix}0 & 0 & 0\\ 1 & -2 & 1\\ 0 & 0 & 0\end{bmatrix} \tag{3}$$

$$\frac{1}{2}\begin{bmatrix}1 & 0 & 0\\ 0 & -2 & 0\\ 0 & 0 & 1\end{bmatrix} \tag{4}$$

$$\frac{1}{4}\begin{bmatrix}-1 & 2 & -1\\ 2 & -4 & 2\\ -1 & 2 & -1\end{bmatrix} \tag{5}$$

Another mask used in experiments is 5×5 proposed in (Chang et al., 2020):

$$\frac{1}{12}\begin{bmatrix}-1 & 2 & -2 & 2 & -1\\ 2 & -6 & 8 & -6 & 2\\ -2 & 8 & -12 & 8 & -2\\ 2 & -6 & 8 & -6 & 2\\ -1 & 2 & -2 & 2 & -1\end{bmatrix}. \tag{6}$$

It was also decided to use several masks at the same time. This solution is described in (Zhou et al., 2018), where authors tested 30 different SRM filters and showed that only three of them were satisfactory. These were the filters presented in points 6, 5 and 3.

Figure 1 shows an example of the result of applying filters. Two photos from the left show the original images (unmodified), and the other two were modified with DeepFake. The transformed form of the images resembles a collection of points that is not very readable for the human eye. You can see the facial features from the original photos, but the representation of these images is not convenient for a human.

3.2 Masks for Approximation of Fractional Order Derivatives

In (Amoako-Yirenkyi, P. Appati, J.K. Dontwi, 2016) proposed a mask based on Riemann-Liouville fractional-order derivative definition. The 5×5 gradient mask looks as follows:

$$\begin{bmatrix}
\frac{2\alpha\sqrt{8^\alpha}}{8\Gamma(1-\alpha)} & \frac{\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & 0 & \frac{-\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{8^\alpha}}{8\Gamma(1-\alpha)}\\
\frac{2\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{\alpha\sqrt{2^\alpha}}{2\Gamma(1-\alpha)} & 0 & \frac{-\alpha\sqrt{2^\alpha}}{2\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)}\\
\frac{2\alpha\sqrt{4^\alpha}}{4\Gamma(1-\alpha)} & \frac{\alpha}{\Gamma(1-\alpha)} & 0 & \frac{-\alpha}{\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{4^\alpha}}{4\Gamma(1-\alpha)}\\
\frac{2\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{\alpha\sqrt{2^\alpha}}{2\Gamma(1-\alpha)} & 0 & \frac{-\alpha\sqrt{2^\alpha}}{2\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)}\\
\frac{2\alpha\sqrt{8^\alpha}}{8\Gamma(1-\alpha)} & \frac{\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & 0 & \frac{-\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{8^\alpha}}{8\Gamma(1-\alpha)}
\end{bmatrix} \tag{7}$$

for the derivative in respect to x and

$$\begin{bmatrix}
\frac{2\alpha\sqrt{8^\alpha}}{8\Gamma(1-\alpha)} & \frac{2\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{2\alpha\sqrt{4^\alpha}}{4\Gamma(1-\alpha)} & \frac{2\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{2\alpha\sqrt{8^\alpha}}{8\Gamma(1-\alpha)}\\
\frac{\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{\alpha\sqrt{2^\alpha}}{2\Gamma(1-\alpha)} & \frac{\alpha}{\Gamma(1-\alpha)} & \frac{\alpha\sqrt{2^\alpha}}{2\Gamma(1-\alpha)} & \frac{\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)}\\
0 & 0 & 0 & 0 & 0\\
\frac{-\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{-\alpha\sqrt{2^\alpha}}{2\Gamma(1-\alpha)} & \frac{-\alpha}{\Gamma(1-\alpha)} & \frac{-\alpha\sqrt{2^\alpha}}{2\Gamma(1-\alpha)} & \frac{-\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)}\\
\frac{-2\alpha\sqrt{8^\alpha}}{8\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{4^\alpha}}{4\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{5^\alpha}}{5\Gamma(1-\alpha)} & \frac{-2\alpha\sqrt{8^\alpha}}{8\Gamma(1-\alpha)}
\end{bmatrix} \tag{8}$$

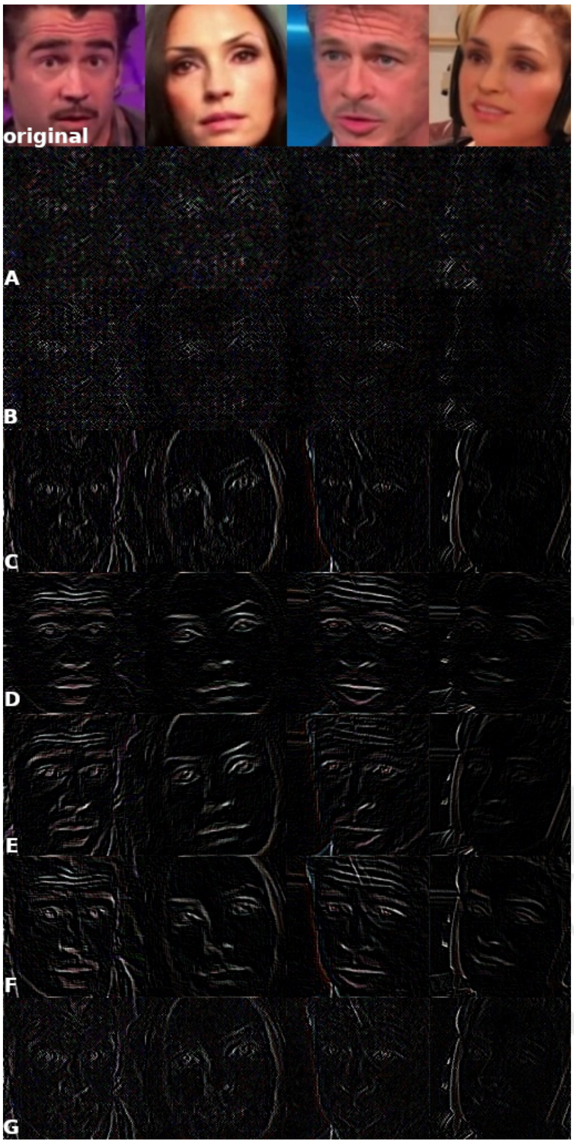

Figure 1: Results of applying SRM filters. The series of photos marked with letters represent: A) 5×5 mask B) a vertical directional mask C) left diagonal directional mask D) horizontal directional mask E) right diagonal directional mask F) 3×3 mask G) combination of 5×5 masks, 3×3 and the horizontal directional for the grayscale image.A constant value enhanced the obtained images for better visualization.

for the derivative in respect to y, where $\Gamma(z)$ is a gamma function which is an extension of the factorial function to complex numbers, a and α is a fractional order of approximated derivative.

The composition of the final result is determined by the dependence: $z \to \sqrt{x^2 + y^2}$. In general, mask elements are expressed by the following formulas:

$$\Theta_x(x_i, y_i) = -\frac{\alpha \cdot x_i}{\Gamma(1-\alpha)} \left(x_i^2 + y_i^2\right)^{-\alpha/2-1}, \tag{9}$$

Figure 2: Filtering results using an approximating mask that approximates the fractional order derivative.

$$\Theta_y(x_i, y_i) = -\frac{\alpha \cdot y_i}{\Gamma(1-\alpha)} \left(x_i^2 + y_i^2\right)^{-\alpha/2-1}, \tag{10}$$

where $-m \leq i \leq m$ or $-n \leq j \leq n$ for $(2m+1) \times (2n+1)$ is the mask size for all $m, n \geq 1$ and α is a constant parameter that specifies the order of the derivative.

Figure 2 shows examples of the transformation performed by this method for different derived orders. The resulting photos obtained for different sizes of derivative orders look very similar to each other. However, there are some minor differences like edge reinforcement.

3.3 Calculating Fractional Order Derivatives with FFT

In (Sarwas and Skoneczny, 2019) was shown how to calculate image fractional order derivatives based on the Riemann-Liouville definition. This method uses

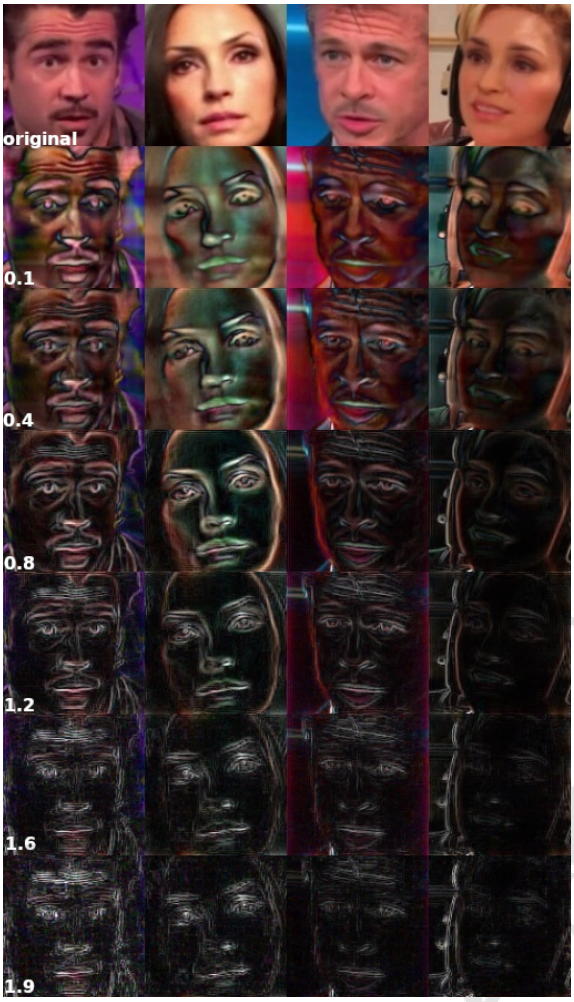

Figure 3: Results of fractional order derivative method based on FFT.

the Fast Fourier Transform. The derivative formulas for the x and y coordinates are as follows:

$$D_x^\alpha g = \mathcal{F}^{-1}((j\omega_1)^\alpha G(\omega_1, \omega_2)), \tag{11}$$

$$D_y^\alpha g = \mathcal{F}^{-1}((j\omega_2)^\alpha G(\omega_1, \omega_2)), \tag{12}$$

where $\mathcal{F}^{-1}$ is an inverse two-dimensional continuous Fourier transform operator, and G is a Fourier transform. The final result was combined for both directions, producing and calculating as the dependence: $z \to \sqrt{x^2 + y^2}$.

Figure 3 shows an example of the results obtained using this method. The same arrangement as in the previous examples was adopted. For each derivative row the edges are enhanced and the resulting photos become darker as the derivative order increases.

3.4 Integer Order Derivatives

As part of the experiment, classical derivatives of the integer order will also be tested. It was decided to use the most popular approximations in the form of the Sobel, Scharra, Prewitt, and Laplacian operators. In this way, first and second order derivatives will be calculated.

3.5 Additional Layers

Adding any filter before using the neural network is similar to extending the model with additional convolutional layers. The difference is in the selection of the filter weights. In the first case, they are predetermined and cannot be changed. In the second, the initial value is random, and only in the learning process, the network optimizes them for the best results. It was decided to extend the study of adding to the model additional convolutional layers because it gives a possibility for better evaluation of other algorithms. Experiments tested the extra layer of the following sizes: 3×3, 5×5, 7×7, and 9×9.

4 EXPERIMENTS

This section presents the description of performed experiments. At the beginning is presented the factors guided by selecting the data set, its description, and preparation for use. Then, the detailed course of the experiment and the learning chosen parameters are described. In the end, the obtained results are listed in the table and analyzed.

4.1 Dataset

One of the most extensive DeepFake datasets is Celeb-DF, developed by Yuezun Li et al. (Li et al., 2019). The second version of this collection contains 590 original videos collected from YouTube and 5,639 videos with DeepFake modification. The quality of this data set is due not only to the number of samples, but also to the realistic face manipulation, which is confirmed by the relatively poor results of the popular DeepFake detection networks.

The research focused on single images, which required prior processing of the dataset containing videos to extract individual frames. The Haar algorithm was used to detect faces on images based on the model for frontal view detection. It let cut them out of those pictures. First, the films were divided into three sets so that the faces in the test, validation, and training groups were unique. Then, the same number of frames from each movie was randomly selected, and faces were cut from them to obtain the following number of sets: 9103 images in the training set, 3145 in the validation set, and 3079 in the test set.

4.2 Parameters

The initial parameter values were taken from (Chang et al., 2020). Images with a size of 128×128 pixels were provided as input to the models and processed using the described methods (SRM filters, fractional order derivative approximation using a mask, fast Fourier transform, and classical forms of approximation of integral derivatives). Additionally, random mirror images of the image concerning the horizontal and vertical axis for data augmentation were used.

For the VGG16 model training process, the SGD optimizer with a constant decay coefficient of $1e{-}6$ and a Nesterov moment of 0.9 was used. The initial learning rate was adjusted experimentally to achieve the best performance. The values ranged from 0.001 to 0.0001. Categorical_crossentropy was chosen as the loss function. Before starting the experiment, the influence of the random factor was reduced by defining a constant seed value of the pseudorandom number generator. An early stop mechanism was used during the training of the neural network. The process was interrupted if the obtained results were not improved on the validation set for a specified number of epochs.

4.3 Results

The results of the experiment are presented in four tables. Each of them in the first line contains the result obtained for the VGG16 model for a more convenient comparison with the tested preliminary filtering algorithm. The area presented the measure of the effectiveness of a given method under the ROC curve (Receiver Operating Characteristics) marked as AUC (Area Under the Curve). This measure was calculated for the validation and test data set.

Table 1: Results of SRM filters in DeepFake detection.

Method	AUC	
	Val	Test
VGG16	0.876	0.870
SRM 3×3 + VGG16	0.856	0.844
SRM 5×5 + VGG16	0.880	0.869
SRM - vertical + VGG16	0.898	0.885
SRM - right-diagonal + VGG16	**0.910**	0.905
SRM - horizontal + VGG16	0.895	0.894
SRM - left-diagonal + VGG16	0.906	**0.910**
SRM - mix filter + VGG16	0.891	0.888

Table 1 contains the results of model detection with preliminary filtering in the form of SRM filters. Only one mask (SRM 3×3) decreased the detection efficiency of DeepFake modifications. This shows that the local noise, enhanced by the filter, had a positive effect on the performance of the model. The best results were obtained for masks with diagonal directions. The AUC was increased by 0.04 compared with detection without pre-filtering.

Table 2 shows the results obtained for the fractional-order derived methods. Before starting the experiment, similar results of both algorithms were

Table 2: Results of fractional order derivative in DeepFake detection.

Method		AUC	
		Val	Test
VGG16		0.876	0.870
Derivative approximation mask + VGG16	0.1	0.882	0.885
	0.2	**0.895**	**0.892**
	0.3	0.886	0.880
	0.4	0.891	0.881
	0.5	0.880	0.883
	0.6	0.879	0.879
	0.7	0.884	0.884
	0.8	0.882	0.872
	0.9	0.886	0.889
	1.1	0.889	0.883
	1.2	0.835	0.826
	1.3	0.884	0.885
	1.4	0.880	0.878
	1.5	0.879	0.871
	1.6	0.884	0.866
	1.7	0.881	0.868
	1.8	0.872	0.882
	1.9	0.882	0.882
Derivative FFT + VGG16	0.1	0.829	0.845
	0.2	0.875	0.877
	0.3	0.866	0.869
	0.4	0.884	0.881
	0.5	**0.888**	**0.885**
	0.6	0.876	0.878
	0.7	0.879	0.876
	0.8	0.861	0.854
	0.9	0.848	0.854
	1.1	0.824	0.820
	1.2	0.807	0.807
	1.3	0.779	0.774
	1.4	0.782	0.779
	1.5	0.783	0.761
	1.6	0.680	0.693
	1.7	0.669	0.676
	1.8	0.744	0.737
	1.9	0.658	0.676

expected. In the case of using the FFT, a downward trend can be seen with the increase in the order of the derivative. The derivative in the range of 0.2-0.7 improved the detection results, and for the remaining rows, it deteriorated. The worst efficiency was obtained for derivatives larger than 1. This is probably related to the noise that appears above the order of 1 for the FFT approximation (see Figure 3). The second method of calculating the incomplete order showed very similar results in the studied range. Single drops in effectiveness can be seen, but these were not greater than 0.045 for the measure AUC. Ultimately, the two algorithms did not bring many benefits in detecting DeepFake modifications.

Table 3: AUC results for proposed integer derivatives DeepFake detection methods.

Method		AUC	
		Val	Test
VGG16		0.876	0.870
First derivative	Sobel	0.881	0.876
	Scharr	0.910	0.897
	Prewitt	0.881	0.878
Second derivative	**Laplacian**	**0.925**	**0.907**

The results of the next experiment were placed in Table 3. Pre-filtering was tested in the form of a classical integer-order derivative. The approximations using the Sobel and Prewitt operators slightly improved the detection efficiency. The Scharr operator fared much better, increasing the AUC score by 0.027. The second derivative achieved the most significant increase in the accuracy of DeepFake modification detection compared to the experiment without using any filter. The AUC results were 0.925 and 0.907 for the validation and test sets, respectively.

Table 4: AUC results for additional convolutional layer in proposed DeepFake detection methods.

Method	AUC	
	Val	Test
VGG16	0.876	0.870
3×3 + VGG16	0.908	0.888
5×5 + VGG16	0.912	0.900
7×7 + VGG16	**0.922**	**0.923**
9×9 + VGG16	0.877	0.887

The best results have been obtained by adding an additional convolutional layer. They were placed in Table 4. Each tested filter size improves network performance. An extra 7×7 convolution layer turned out to be better than all the previously tested prefiltering algorithms.

5 CONCLUSION

This paper addresses the issue of DeepFake image modification detection. As part of the work carried out, solutions based on deep neural networks for detecting this type of disturbance were analyzed, focusing on suggestions for preliminary photo filtering.

As part of the research, various SRM filters were compared with the proposed FoSRM based on the fractional order derivatives implemented in two forms: an approximation mask and a fast Fourier transform. The operation of classical methods of determining derivatives of the integer order were also tested. To reliably evaluate the filters, it was also examined what results could be obtained if different sized of one convolution layer will be added to the neural network. Prefiltering of the images were carried out on the entire data set before starting the process of training the neural networks. Optimal training parameters were selected, and the results obtained for a set of validation and test data were compared.

The analysis of the results showed that using the fractional order derivative can rich input image as well as other SRM filters. From the linear filers the best efficiency had diagonal SRM and Laplacian. Surprisingly, the mix of SRM masks degraded the obtained results compared to the individual masks. On the other hand, adding one layer to the neural network provide the best solution. It can be concluded that in the case of using convolutional networks, it makes no sense to use additional linear and even nonlinear image prefiltering. In the learning process, the neural network automatically selects linear filters and examines their influence on the detection based on the convolution operation. On the other hand, nonlinear operators are performed using the ReLU type activation function or the MaxPooling operator, which is also analyzed during the training process.

The use of a defined filter may prove effective in more complex network architectures, where the network is forced to extract multiple different sets of features and then merge all together. An example of such an architecture is a multi-stream network. These conclusions may motivate further research on the influence of prefiltering on DeepFake modification detection results.

REFERENCES

Amoako-Yirenkyi, P. Appati, J.K. Dontwi (2016). A new construction of a fractional derivative mask for image edge analysis based on riemann-liouville fractional derivative. *Advances in Difference Equations*.

Chang, X., Wu, J., Yang, T., and Feng, G. (2020). Deepfake face image detection based on improved vgg convolutional neural network. In *2020 39th Chinese Control Conference (CCC)*, pages 7252–7256.

Chen, M., Shi, X., Zhang, Y., Wu, D., and Guizani, M. (2017). Deep features learning for medical image analysis with convolutional autoencoder neural network. *IEEE Transactions on Big Data*, pages 1–1.

Garcia, F. C. C., Creayla, C. M. C., and Macabebe, E. Q. B. (2017). Development of an intelligent system for smart home energy disaggregation using stacked denoising autoencoders. *Procedia Computer Science*, 105:248–255. 2016 IEEE International Symposium on Robotics and Intelligent Sensors, IRIS 2016, 17-20 December 2016, Tokyo, Japan.

Guo, Z., Yang, G., Chen, J., and Sun, X. (2020). Fake face detection via adaptive residuals extraction network. *CoRR*, abs/2005.04945.

Han, B., Han, X., Zhang, H., Li, J., and Cao, X. (2021). Fighting fake news: Two stream network for deepfake detection via learnable srm. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, pages 1–1.

Kang, S., Park, H., and Park, J.-I. (2019). Cnn-based ternary classification for image steganalysis. *Electronics*, 8(11).

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Li, Y., Yang, X., Sun, P., Qi, H., and Lyu, S. (2019). Celeb-df: A new dataset for deepfake forensics. *CoRR*, abs/1909.12962.

Liu, Z., Qi, X., Jia, J., and Torr, P. H. S. (2020). Global texture enhancement for fake face detection in the wild. *CoRR*, abs/2002.00133.

Mirsky, Y. and Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Comput. Surv.*, 54(1).

Rana, M. S. and Sung, A. H. (2020). Deepfakestack: A deep ensemble-based learning technique for deepfake detection. In *2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, pages 70–75.

Reinel, T.-S., Brayan, A.-A. H., Alejandro, B.-O. M., Alejandro, M.-R., Daniel, A.-G., Alejandro, A.-G. J., Buenaventura, B.-J. A., Simon, O.-A., Gustavo, I., and Raúl, R.-P. (2021). Gbras-net: A convolutional neural network architecture for spatial image steganalysis. *IEEE Access*, 9:14340–14350.

Sarwas, G. and Skoneczny, S. (2019). Half profile face image clustering based on feature points. In Choraś, M. and Choraś, R. S., editors, *Image Processing and Communications Challenges 10*, pages 140–147. Springer International Publishing.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.

Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., and Liu, W. (2018). Cosface: Large margin cosine loss for deep face recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5265–5274.

Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 58:101552.

Younus, M. A. and Hasan, T. M. (2020). Effective and fast deepfake detection method based on haar wavelet transform. In *2020 International Conference on Computer Science and Software Engineering (CSASE)*, pages 186–190.

Zhang, R., Zhu, F., Liu, J., and Liu, G. (2018). Efficient feature learning and multi-size image steganalysis based on cnn. *ArXiv*, abs/1807.11428.

Zhou, P., Han, X., Morariu, V. I., and Davis, L. S. (2018). Learning rich features for image manipulation detection. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1053–1061.

Zi, B., Chang, M., Chen, J., Ma, X., and Jiang, Y.-G. (2021). Wilddeepfake: A challenging real-world dataset for deepfake detection.