Advances in Artificial Intelligence and Machine Learning; Research 4 (3) 2783-2809 Received 19-07-2024; Accepted 23-09-2024; Published 30-09-2024

Wavelet Based Inpainting Detection

Barglazan Adrian-Alin adrian.barglazan@ulbsibiu.ro
Department of Computers and Electrical Engineering,
Faculty of Engineering, University "Lucian Blaga",
Sibiu, Romania

Brad Remus remus.brad@ulbsibiu.ro
Department of Computers and Electrical Engineering,
Faculty of Engineering, University "Lucian Blaga",
Sibiu, Romania

Corresponding Author: Barglazan Adrian-Alin

Abstract

With the advancement in image editing tools, manipulating digital images has become alarmingly easy. Inpainting, which is used to remove objects or fill in parts of an image, serves as a powerful tool for both image restoration and forgery. This paper introduces a novel approach for detecting image inpainting forgeries by combining DT-CWT with Hierarchical Feature segmentation and with noise inconsistency analysis. The DT-CWT offers several advantages for this task, including inherent shift-invariance, which makes it robust to minor manipulations during the inpainting process, and directional selectivity, which helps capture subtle artifacts introduced by inpainting in specific frequency bands and orientations. By first applying colour image segmentation and then analysing for each segment, noise inconsistency obtained via DT-CW we can identify patterns indicative of inpainting forgeries. The proposed method is evaluated on a benchmark dataset created for this purpose and is compared with existing forgery detection techniques. Our approach demonstrates superior results compared with SOTA in detecting inpainted images. The proposed methodology source code is uploaded here: https://github.com/jmaba/Wavelet-based-inpainting-detection

Keywords: Image inpainting detection, Image forgery detection.

1. INTRODUCTION

Image inpainting is the process of reconstructing lost or deteriorated parts of an image. It has evolved significantly over the past decades. It is an essential task in various applications such as image restoration, content removal, and computer graphics. Traditional inpainting techniques have given way to more sophisticated methods powered by advances in deep learning, particularly Convolutional Neural Networks (CNNs), Transformer models, and diffusion-based approaches. This review provides an overview of these contemporary methods, highlighting their strengths and challenges. Before the rise of deep learning, inpainting relied on traditional approaches like patch-

Citation: Barglazan Adrian-Alin and Brad Remus. Wavelet Based Inpainting Detection. Advances in Artificial Intelligence and Machine Learning. 2024;4(3):162.

based methods and diffusion-based methods. Patch-based techniques like the ones suggested by [1], operate by copying and pasting similar patches from the known regions of the image to fill in the missing areas. Diffusion methods [2], progressively propagate information from the known regions towards the unknown, gradually refining the inpainted content. While these techniques achieved some success, they often struggled with complex missing regions, leading to blurry or repetitive results. More newer methods like [3], introduces a novel approach to image inpainting that leverages contextual attention mechanisms to fill in large missing regions in images. Traditional convolutional networks struggle with inpainting due to their limited ability to utilize information from distant spatial locations. The proposed method addresses this by explicitly borrowing information from surrounding image areas through a contextual attention layer. This layer helps synthesize new image structures and textures that are consistent with the existing parts of the image. The approach is validated on multiple datasets, including CelebA, DTD, and ImageNet, demonstrating superior results compared to existing methods. Even more recent research from 2021, like in [4], explores image inpainting by employing diffusion models in a latent space. By focusing on latent representations instead of pixel space, the model efficiently learns and restores missing parts of images while preserving the global structure and ensuring high-quality outputs. This method leverages the strengths of both latent space manipulation and diffusion processes to achieve significant improvements in the visual fidelity of inpainted images. In the same category – as diffusion methods, the paper in [5], presents a way for high-resolution image synthesis using latent diffusion models. Diffusion models have been successful in generating high-quality images by iteratively denoising samples from a Gaussian distribution. The latent diffusion approach improves efficiency and scalability by operating in a latent space rather than directly on pixel space. This allows for generating high-resolution images with lower computational costs. The method achieves state-of-the-art results in various image synthesis tasks, highlighting its potential for practical applications. While the advancements in image inpainting and object removal technologies offer numerous benefits, they also bring significant ethical challenges. The potential for misuse in the justice system underscores the need for stringent regulations, ethical guidelines, and the development of forensic tools to detect image tampering. Ensuring that these technologies are used responsibly is crucial to prevent their misuse and protect the integrity of visual evidence in legal contexts.

Detecting image inpainting forgeries is a crucial task in digital forensics and image authentication. Inpainting forgeries, created by sophisticated methods, can seamlessly remove objects or alter images, making detection challenging. One primary approach to detecting these forgeries involves analyzing inconsistencies at various levels, such as noise, texture, and blur. Noise level analysis is essential because authentic images generally exhibit consistent noise patterns throughout. Inpainted regions, however, often display noise discrepancies, as inpainting algorithms may not replicate the original noise characteristics perfectly. Techniques like the Dual-Tree Complex Wavelet Transform (DT-CWT) [6], are particularly effective in identifying these noise level inconsistencies, capturing subtle noise patterns and highlighting discrepancies between inpainted and authentic regions. Texture analysis is another crucial aspect, as inpainting algorithms might fail to maintain texture continuity, especially in complex scenes. Examining the texture features can reveal abrupt changes or unnatural patterns that suggest tampering. Techniques such as co-occurrence matrices and local binary patterns are commonly used to quantify texture and detect anomalies indicative of inpainting. Blur detection is also significant in identifying inpainted regions. Inpainted areas may exhibit different levels of blur compared to the original parts of the image, as the inpainting process might not perfectly match the sharpness or smoothness of the surrounding areas. Gradient-based methods assess the sharpness and gradient distribution across the image,

identifying areas with inconsistent blur levels. Combining these analysis techniques provides a robust framework for detecting inpainting forgeries. Leveraging multi-level feature analysis allows for the identification of subtle inconsistencies that single-level approaches might miss. This holistic approach is crucial for maintaining the integrity and authenticity of digital images in an era where advanced inpainting techniques are increasingly accessible.

Despite the improvements, significant deficiencies remain in the detection of object removal, which can be attributed to several unresolved or partially resolved challenges, such as limited generalizability across diverse image datasets, lack of realism, restricted image variability, inadequate detection accuracy, and the rapid evolution of inpainting technologies. To address these deficiencies, this paper proposes a novel approach that employs image segmentation combined with Dual-Tree Complex Wavelet Transform (DTCWT) to analyze inconsistencies in high-level details within each textured segment. Furthermore, to refine the detection of forged areas, our proposed method incorporates multiple enhancement techniques, including mean, median, and modified median Wiener filters.

2. RELATED WORK

In the world of investigations, ensuring evidence is genuine is crucial. Forensic detection methods come to the rescue, helping identify if documents, images, or videos have been tampered with. These methods aim to confirm the evidence's authenticity and integrity. There are two main approaches active and passive. Active (Watermark-Based), like a hidden security tag, these methods rely on pre-embedded data, such as digital watermarks, within the original content. Verifying the presence and condition of this embedded information helps expose any tampering attempts. Passive (Blind) methods don't require prior knowledge of the original or embedded information. Instead, it analyzes the content itself for inconsistencies that might indicate manipulation. This is particularly useful for evidence from social media or open-source platforms, where the original might not be available. Usually how passive methods work is by inconsistency analysis combined with statistical analysis. These methods are like detectives examining the scene for clues. They look for inconsistencies within the content itself, like pixel-level differences in images, lighting/shadow inconsistencies, or unnatural pattern repetitions. Statistical techniques analyze the content's properties, like noise patterns or compression artifacts.

Forgeries can introduce statistical anomalies that this analysis can detect. For e.g. different cameras and image acquisition processes introduce unique noise patterns. Forgeries involving splicing or copy-pasting from different sources might exhibit inconsistencies in noise characteristics. Statistical measures like standard deviation or kurtosis can analyze noise patterns in different image regions. Deviations from expected values within the image can indicate potential manipulation. The paper "Image Noise and Digital Image Forensics" [7], provides an in-depth exploration of the role of noise in digital image forensics. Noise, an intrinsic element in all forms of imaging, stems from various sources during the image acquisition process, such as the nature of light and optical artifacts, as well as the conversion from electrical signals to digital data. The study highlights several key aspects like sources, models and techniques. Noise Sources and Models can categorize the different types of noise that affect digital images and discusses various models used to represent these noises. Understanding these models is crucial for forensic analysis because noise patterns can be distinctive and unique to specific imaging devices. The paper details how noise can be utilized in digital image forensics to detect forgeries. Techniques like analyzing sensor pattern noise (SPN) and photo-

response non-uniformity (PRNU) are explained. These techniques help in identifying the source of an image and detecting any tampering by examining inconsistencies in the noise patterns. The authors in [8], propose using intrinsic sensor noise differences, which often vary due to different ISO settings, to identify tampered regions in images. The technique involves estimating noise levels locally using a PCA-based algorithm and then clustering the weighted noise levels with k-means. Experimental results demonstrate the method's effectiveness in localizing and detecting spliced regions, outperforming several state-of-the-art techniques. The paper [8], introduces a method to detect image splicing by leveraging inconsistencies in sensor noise levels. By analyzing noise patterns unique to camera sensors, the method identifies discrepancies indicating spliced regions. This approach proves effective in revealing forgeries that are not discernible through visual inspection alone. They first estimated locally the noise based on PCA, then estimated noise levels are clustered via k-means. This paper [9], presents an image forgery localization method utilizing a fully convolutional network (FCN) enhanced with noise features. The approach uses noise patterns to highlight subtle changes in images, thereby improving the network's ability to generalize and detect tampered regions. The FCN architecture generates pixel-wise predictions, and the inclusion of a region proposal network enhances robustness. Experimental results on standard datasets demonstrate the method's effectiveness in accurately locating tampered areas, showcasing improved generalization and robustness over existing techniques. A more recent attempt to detect forgeries based on noise was done by authors in [10]. The paper presents a dual-branch image manipulation detection technique that leverages noise and edge features to improve the accuracy of identifying manipulated images. The approach involves two primary branches: Noise Feature Branch - This branch focuses on analyzing the noise characteristics of the image. Noise patterns are intrinsic to images captured by digital sensors and are often altered when images are manipulated. By examining these noise inconsistencies, the method can highlight areas that are likely tampered with. Edge Feature Branch - this branch analyzes the edges within the image. Manipulation often introduces unnatural edges or alters existing ones. By closely examining edge patterns and inconsistencies, the method can detect signs of image tampering. The paper describes a convolutional neural network (CNN) architecture with two parallel branches: one for noise detection and one for edge detection. Each branch processes the image independently and extracts relevant features. A fusion layer combines the outputs from both branches to make a final decision about the presence of manipulation. This combined approach leverages the strengths of both noise and edge analysis, providing a more robust detection mechanism. In [11], the authors present a novel approach for detecting post-processed image forgeries by utilizing a Signal Noise Separation-based Network (SNIS). This method draws an analogy between image forgery detection and blind signal separation, treating the problem of forgery detection as one of separating signal from noise. The main core idea is to separate the image content (signal) from the noise that is inherently introduced by the image sensor. Since post-processing operations often alter these noise patterns, the method can effectively identify forgeries. The proposed network consists of two main branches: one for noise extraction and one for signal extraction. This dual-branch structure allows the network to learn and distinguish between authentic image content and manipulated regions. The authors trained their model using large datasets of both genuine and forged images, ensuring robust performance across various types of post-processed forgeries. Another automated tool was introduced in [12], designed to detect image forgeries through noise analysis. The approach leverages the inherent noise patterns that digital camera imprint on images. By analyzing these noise characteristics, the system can identify discrepancies indicative of tampering. The system extracts noise from an image using advanced signal processing techniques. This involves separating the noise components from the actual image content. A background stochastic model is defined to characterize the typical

noise patterns produced by a camera sensor. This model serves as a baseline for comparison. The extracted noise is analyzed locally across different regions of the image. The method identifies areas where the noise pattern deviates significantly from the expected model, suggesting potential tampering. Various statistical techniques are employed to quantify the degree of anomaly in noise patterns. These techniques help in distinguishing genuine parts of the image from manipulated ones. The Noisesniffer operates in a fully automatic mode, requiring minimal user intervention. The tool processes the image, extracts noise, analyzes patterns, and provides a decision on the authenticity of the image.

While significant advancements have been made in image forgery detection using noise analysis and machine learning, these methods still face challenges related to computational efficiency, generalization, and robustness against various types of forgeries. To mention a few of the current limitations:

- many methods struggle with low-resolution or highly compressed images
- noise analysis and pattern detection are computationally intensive, requiring significant processing power.
- machine learning-based methods require extensive training datasets to achieve high accuracy
- high sensitivity can lead to false positives, while less sensitivity might result in undetected forgeries
- some methods might produce false positives in images with naturally high noise variations or by different camera models and settings.

3. THEORETICAL ASPECTS

3.1 Dual Tree Complex Wavelet

The Dual Tree Complex Wavelet Transform (DT-CWT) is an image processing method that offers several advantages over traditional Discrete Wavelet Transforms (DWT). DWT decomposes an image into subbands containing information about different spatial frequencies and orientations. However, it treats horizontal and vertical details separately. DT-CWT addresses this by performing the decomposition on two separate trees simultaneously (called A and B tree in the original paper [6], (see below FIGURE 1). These filter banks share identical frequency responses but have slightly different phases, achieved through specific filter design techniques like Kingsbury Q-shift filters. This redundancy helps in achieving near-shift invariance. Each filter bank employs scaling functions ($\varphi(x)$) and wavelet functions ($\psi(x)$) to decompose the signal at different scales and orientations. Inpainting detection identifies areas within an image that have been altered or filled in.

The Dual-Tree Complex Wavelet Transform (DTCWT) is effective as a feature extractor in this context due to several properties. Improved Directionality and Texture Representation: DTCWT captures edge and texture information more effectively than traditional wavelet transforms, making it adept at detecting discrepancies in texture and edge continuity caused by inpainting. Unlike traditional wavelet transforms, DTCWT maintains consistent feature extraction despite shifts, ensuring

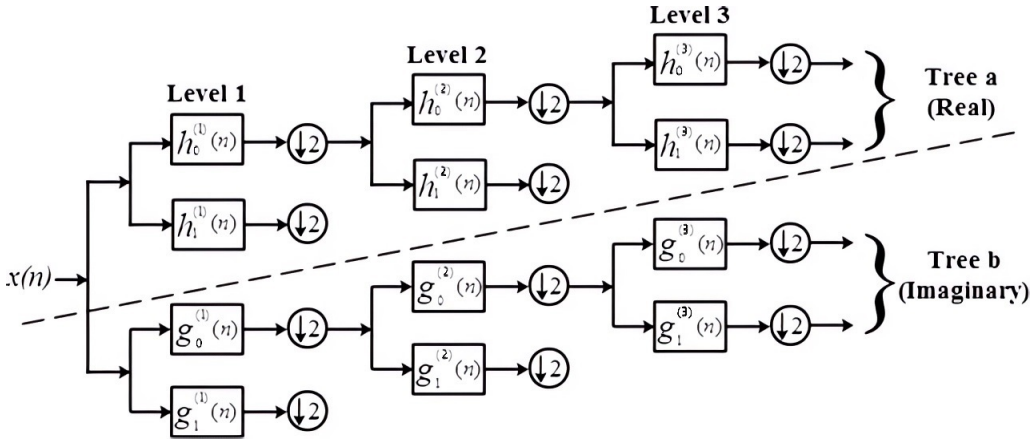

Figure 1: Representation of dual-tree complex wavelet transform over 3 levels from the original Kingsbury paper

reliable detection of inpainted regions regardless of their position in the image. The dual-tree structure of DTCWT captures both magnitude and phase information, revealing subtle inconsistencies in inpainted areas that may not be apparent through intensity analysis alone. Phase congruency, derived from phase information, can indicate disruptions in the natural flow of the image's texture or edges. Reduced Aliasing: DTCWT exhibits less aliasing in its subbands compared to traditional wavelet transforms, ensuring cleaner and more representative features, which enhances the detection of inpainted regions, especially those that are small or highly precise. DTCWT's subbands provide a clearer distinction between high-frequency noise components and inpainting artifacts, effectively isolating localized changes in frequency and directional ranges to pinpoint manipulated areas. See the below illustration - FIGURE 2 (all inpainting methods are done using [13], – and source image is from [14], – more details will be provided in the results part):

In practice only the first level of decomposition shall be applied, all 6 orientations bands are considered (both original and imaginary part) resulting in a total of 12 images that are analyzed. The DT-CWT provides complex coefficients $c_{j,k}$ representing the image $I(x)$ at scale j and position k, computed as:

$$c_{j,k} = \langle I(x), \psi_{j,k} \rangle$$

where $\psi_{j,k}$ are the wavelet basis functions. A key property of DT-CWT is approximate shift invariance, which implies that small shifts in the input image $I(x)$ result in minimal changes in the coefficients $c_{j,k}$. Mathematically, for a small shift Δx, the coefficients transform as:

$$c_{j,k}' = \langle I(x + \Delta x), \psi_{j,k} \rangle \approx c_{j,k} e^{i\theta}$$

where θ is a phase shift. This shift invariance ensures that coefficients corresponding to similar structures (e.g., uniform color or texture) remain consistent, which is critical for distinguishing authentic from tampered regions. In a truly homogeneous region (large uniform areas of the same color or texture), the wavelet coefficients should exhibit statistical uniformity. Specifically, if a segment is homogeneous, the mean μ and variance σ^2 of the coefficients should be consistent across that segment:

$$\mu = \frac{1}{N} \sum_k c_{j,k}, \quad \sigma^2 = \frac{1}{N} \sum_k \left(c_{j,k} - \mu \right)^2$$

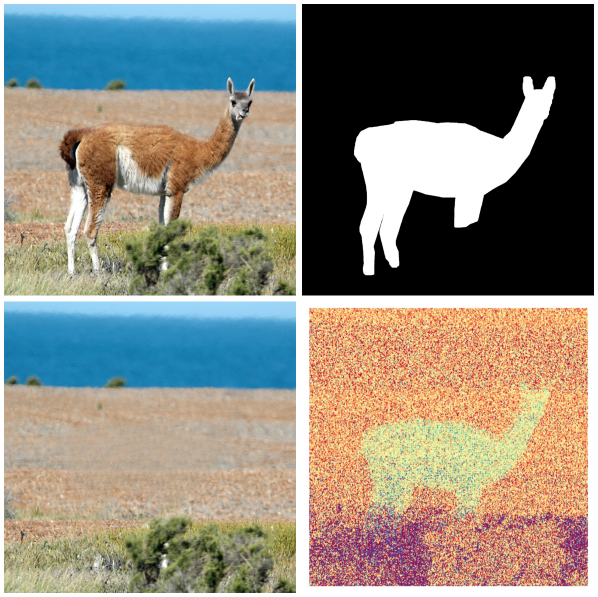

Figure 2: First row: original image and mask image of the object to be removed. Second row: image inpainted and DTCWT (one level of decomposition selected only one band – third one – real values image enhanced – used spectral coloring and values normalized

Due to the shift invariance, these statistical measures should not vary significantly within a uniform segment. Forgery introduces inconsistencies such as changes in texture, illumination, or structure, disrupting the uniformity of the coefficients. These inconsistencies can be detected by analyzing deviations in the distribution of coefficients. If a uniform segment shows unexpected variations in its coefficients, such as changes in mean, variance, or phase:

$$\text{Phase} = \arctan\left(\frac{\text{Imaginary Part}}{\text{Real Part}}\right)$$

it suggests tampering. The anomalies in the coefficients are statistically significant when compared to the expected distribution of authentic segments. Consider an image segment expected to be homogeneous. Let the coefficients $c_{j,k}$ follow a normal distribution $N\left(\mu, \sigma^2\right)$. In the presence of forgery, the coefficients deviate from this expected distribution:

$$c_{j,k}^{\text{forged}} = c_{j,k}^{\text{authentic}} + \delta$$

where δ represents the irregularity introduced by the forgery. The detection problem then reduces to hypothesis testing, where the null hypothesis H_0 assumes the segment is authentic ($\delta = 0$), and the alternative hypothesis H_1 assumes forgery ($\delta \neq 0$):

$$H_0 : c_{j,k} \sim N\left(\mu, \sigma^2\right), \quad H_1 : c_{j,k} \sim N\left(\mu + \delta_\mu, \sigma^2 + \delta_\sigma^2\right)$$

Rejecting H_0 in favor of H_1 indicates a forged segment. This is typically done using statistical measures such as Z-tests or Chi-square tests on the coefficient distributions. Thus, the DT-CWT's high-level features are instrumental in detecting forgeries because they capture the global consistency of image segments, and any deviations from this consistency serve as strong indicators of tampering.

3.2 Noise Level Analysis and Estimation

The starting point of the proposed method is the idea expressed in [15]. In this paper, the authors present a novel approach for estimating the noise level in images. The method leverages principal component analysis (PCA) on image patches to identify noise-dominant components. They assume that noise follows a Gaussian distribution. Based on this concept, they applied PCA to separate noise from signal by analyzing eigenvalues. The noisy image I can be modeled as I = I0 + N, where I0 is the clean image and N is the noise. PCA is applied to small patches of the image to decompose the data into principal components. For a set of image patches X, PCA finds the eigenvalues λ_1, λ_2, ..., λ_n of the covariance matrix of X. The smallest eigenvalue λ_{min} from the PCA of the image patches is used to estimate the noise variance σ^2. The rationale is that the smallest eigenvalue is dominated by the noise component due to its isotropic nature. To better analyze the noise inconsistencies, we proposed to alter the above method, and apply first a dual-tree complex wavelet decomposition, and then perform the noise inconsistency inside the HH bands of the dual tree complex wavelet. The above method works well on both natural and synthetic images, but in the case of highly texturized images it struggles to yield a good result. Based on the above work, we postulate that inside an image, if it were to apply color segmentation, and apply the method per segments of same color and texture, then we can extract the noise variance at segment level. Thus, defining a noise estimation at the segment level, we then check if all the patches inside color segmentation are in the same threshold as the noise estimated for the entire segment. In case there is a variance more than 30%-50%, we classify that region as suspicious, and it will enter further analysis. To demonstrate this, several images were selected. Using the above noise estimation level 3 set of noise were computed: from entire image, from the region outside the object removed and from the region where the object was removed. For the simplicity of tests, we've selected images which contain only one type of texture. FIGURE 3, we can see the original image on the left side, the mask image on the right and the altered image on the bottom. In FIGURE 4, there is another image representing the mask applied to the inpainted region – only the inside mask and only the outside mask. In the first instance the noise of the hole altered image is combined. Then the noise is computed for only the inside and outside mask separately. The method [15], is applied to for noise estimation, but instead of applying it directly to RGB image, it is applied to each band (real and imaginary) of the DT-CWTs.

Figure 3: Original image / Mask image of object to be removed / Altered (inpainted) image

For the analysis from the complex wavelet decomposition only the real and imaginary were taken. Also, all the bands (0 to 5) are analyzed. The results are highlighted in the TABLE 1.

The observed data indicates that the noise distribution is significantly impacted by the inpainted area. For example, in the image provided, an analysis of the real values across all six levels of band

Figure 4: Image altered inside mask /outside mask

Table 1: Applying noise estimations on DT-CWT bands.

Band	***Band information (real / imaginary)***	***Measurement from:***		
		Mask	**Outside Mask**	**Entire band (image)**
0	Real	205.92	278.76	276.88
0	Abs	156.84	159.88	158.88
1	Real	104.37	175.08	173.60
1	Abs	119.50	174.29	200.76
2	Real	177.33	267.01	264.88
2	Abs	86.67	191.51	193.46
3	Real	104.75	217.22	214.40
3	Abs	115.04	151.89	148.42
4	Real	120.11	237.37	231.97
4	Abs	113.70	171.00	169.79
5	Real	126.02	205.27	209.28
5	Abs	212.75	269.44	280.47

decomposition reveals a noteworthy pattern. Specifically, the noise distribution from the regions outside the masks closely resembles that of the entire image. However, there is approximately a 50% discrepancy between the noise distribution in the altered region compared to the rest of the image. This highlights the distinct alteration in noise characteristics due to inpainting.

4. PROPOSED METHOD

The proposed method (see FIGURE 5) aims to identify and enhance regions of interest within texture segments and to differentiate between regions with uniform noise distribution and those with significant variations.

The method can be described as follows:

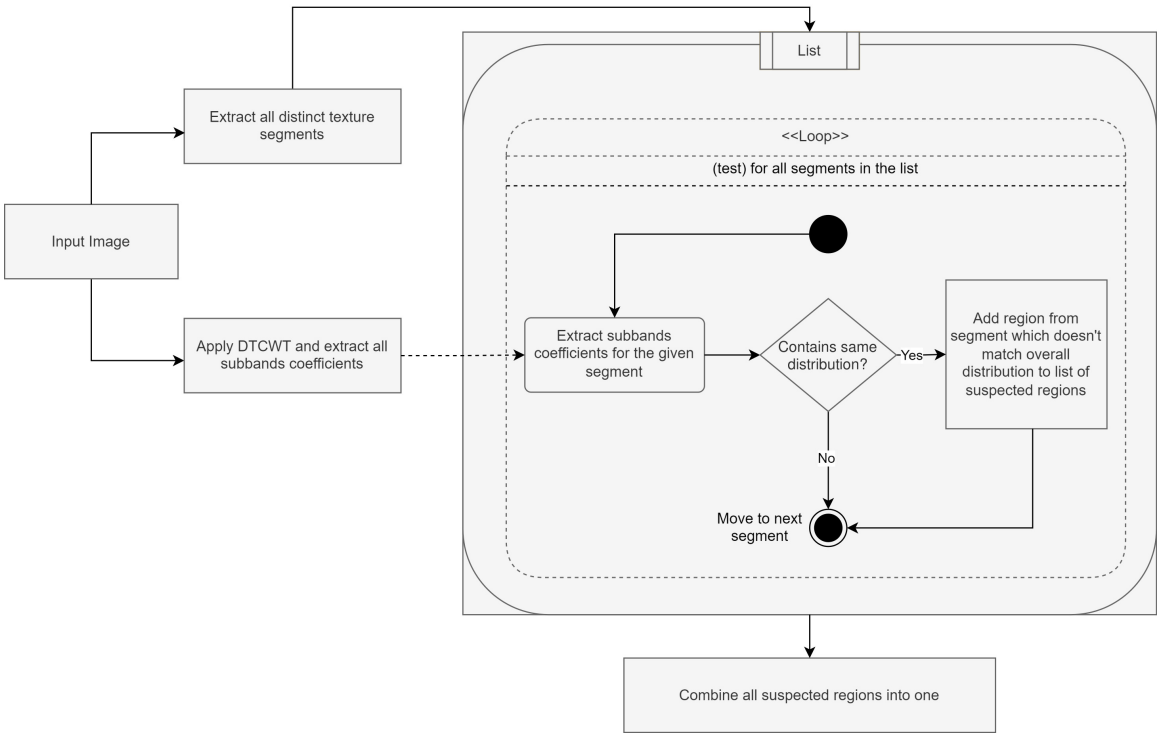

Figure 5: Proposed method

- **Texture Segmentation using HFS**: Perform texture segmentation using the Hierarchical Feature Selection (HFS) technique, resulting in a set of distinct segments (we call them also mask regions).
- **DTCWT Wavelet Decomposition**: Apply Dual-Tree Complex Wavelet Transform (DTCWT) decomposition, yielding 12 matrices corresponding to 6 real and 6 imaginary bands.
- **Operations for Each Texture Segmentation (mask)**: For each texture segmentation obtained from HFS, the following steps are performed:
  - **Extraction of Wavelet Coefficients**: For each segmentation mask, extract the relevant wavelet coefficients by applying a masking operation, where coefficients corresponding inside mask are retained, and others are set to zero.
  - **Enhancement of Matrices**: Enhance each of the masked 12 matrices using four different methods, resulting in 48 enhanced matrices.
  - **Clustering and Noise Variance Analysis**:
    * Perform clustering on each of the 48 enhanced matrices, thus obtaining for each matrix (image) a set of distinct clusters (regions).
    * Calculate the noise variance on the original band matrix and on the clustered regions. If a significant discrepancy between these variances is detected, the clustered region having the most variance is flagged as fake.
- **Generation of Proposed Mask**: All detected regions are combined, and a median value is computed across these regions to generate the final proposed mask.

In the context of image inpainting, specifically for object removal, the removed object is invariably encompassed within a larger area. Even when the object is situated at the boundary between multiple regions, each section of the removed object will be seamlessly filled with textures from the adjacent areas. This process inherently alters the noise statistical model within each modified color segment. Given these observations, we propose that the forgery detection method should be applied independently to each texture segment. This ensures that the detection is precise and accounts for the unique characteristics of each color segment. Our novel proposed method can be summarized as follows: first, the image undergoes semantic segmentation, resulting in N texture segments, labeled as S_i for i in the range [0..N). Simultaneously, a one-level DT-CWT decomposition is performed on the grayscale image. For each of the 6 bands (where k is between 0 and 5), we extract the real and imaginary parts separately, resulting in 6 W_{kReal} and $W_{kImaginary}$ matrices. Each segment's color serves as a mask for extracting information from the 12 images. We analyze wavelet coefficients, extract patches, and check for inconsistencies to identify potential forgeries. Minimal variations lead to skipping the band information, while significant variations prompt further analysis.

The extracted band data undergoes normalization, which involves a logarithmic transformation to handle negative values, followed by a z-transform to standardize the data. Several enhancement techniques are applied, including mean modified Wiener filtering, median filtering combined with bilateral filtering, Laplacian un-sharpening, and small median filter residue. These techniques were chosen to highlight the variations within the segment. For each enhanced feature, K-means or c-means clustering with two segments/groups is applied. If two distinct segments are identified within the segmented area, the segment with the largest difference in noise is added to the forged image segments. Thus, for an image, we obtain N segments * 12 matrices of wavelet coefficients * 4 different enhancement methods, resulting in 48 N distinct images. The mean value of these images is computed to obtain the inpainted detection mask. Thus, each pixel will represent an intensity. The higher the intensity, the more likely it is that the pixel was detected in several bands (enhancements).

4.1 Region Segmentation and Dual-Complex Wavelet Band Extraction

The above FIGURE 6, on the first row is presenting the original image, followed by mask image of the object to be removed. On the second row, the inpainted image is presented. As mentioned above, the first step is to apply semantic segmentations. For this Hierarchical Feature Selection [16], is employed. Hierarchical feature selection is a technique used to identify important features for machine learning tasks, but with a twist compared to traditional methods. It goes beyond just picking the "best" features and instead builds a hierarchy of feature subsets. Traditional feature selection often treats features independently. Hierarchical selection, however, considers the relationships between features. It creates a hierarchy, like a family tree, where features are grouped based on their similarity and importance. The result of applying HFS on the image is presented in the image below see FIGURE 7.

For the next operation, each individual segment must be considered. For the easiness of the exampled hereby only the largest segment will be further analyzed (in our implementations all segments having a size greater than a threshold defined at image level size are considered), see FIGURE 8.

The dual-complex tree wavelet is also applied on the entire image. From the analysis only the first decomposition level is taken. From the complex numbers the real and imaginary parts are extracted

Figure 6: First row original image / mask image. Second row inpainted image

Figure 7: Output of HFS segmentaton. Each segment has a different color

independently (FIGURE 9 and FIGURE 10), thus obtaining 12 matrices (images). In the pictures below from first row, from left to right all wavelet coefficients obtained – first image represents the real coefficients while the second image represents the imaginary part of the complex numbers.

To illustrate, if we are analyzing band 2 and focusing on its real components (assuming necessary enhancements have been applied for optimal image display), and a mask has been applied (based on

Figure 8: Segmentation masked further analyzed

FIGURE 8) to filter out only the area within the specified texture, the input image for the subsequent stages of the algorithm should resemble the following – see FIGURE 11:

As a side note, to improve perceptibility in the above images, the images are enhanced by normalizing and clamping values outside the interval (-1, 2) and applying the Matplotlib Spectral color scheme.

4.2 Wavelet Filtering

As stated above for one segment, the method will extract 12 different matrices. Some of the matrices might not contain relevant features for the given image. For e.g. in FIGURE 9 one can noticed that only bands 1,3,5 contain some visible traces, while the rest of bands (2,4,6) do not contain visible traces. To determine which bands, contain relevant information for each band, the area is split into non-overlapping patches and patches mean is computed. If there is a higher variance (FIGURE 12) among the mean patches inside the segmented area the band will be further processed, otherwise it will be skipped.

4.3 Band Enhacements

After the bands with inconsistencies have been identified, the next step consists in enhancing the altered region to be able to detect it easier. For this several filtering are employed:

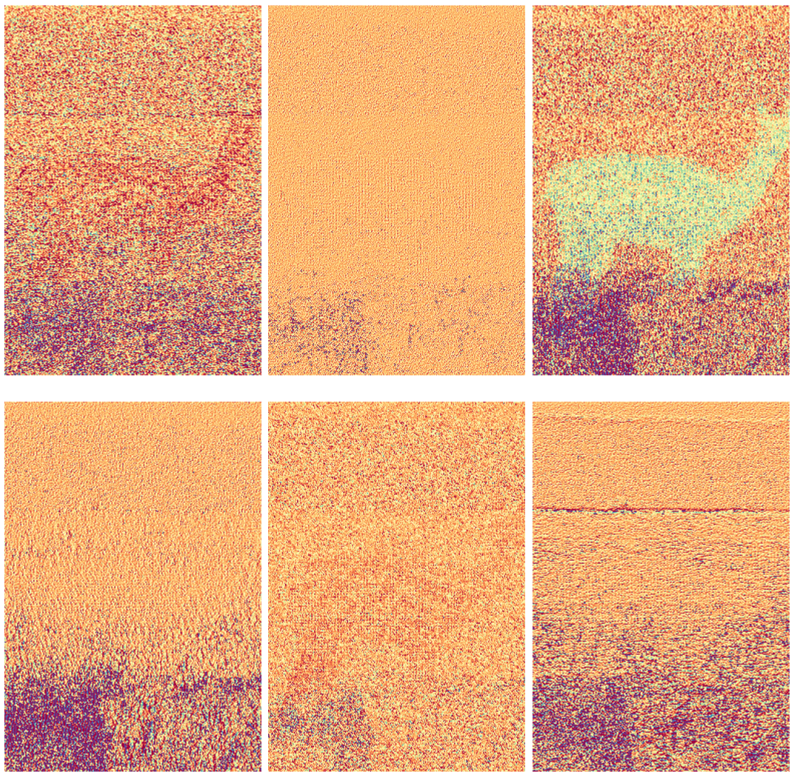

Figure 9: Real coefficients of DC-WT of the original image. On the first row are the first 3 bands and on the second row the rest of the last 3 bands

- Median Filtering: Removes impulsive noise from individual bands while preserving important structural details.
- SMFR: Isolates and enhances fine details and edges within the wavelet bands, reducing noise and highlighting features.
- Wiener Filtering: Adaptively reduces noise in each band, enhancing signal fidelity and preserving details.
- Median Modified Wiener Filtering: Combines the noise reduction and edge-preserving benefits of median and Wiener filtering, providing robust enhancement across the wavelet bands.

Median filtering is a non-linear process commonly used to reduce noise, particularly salt-and-pepper noise, while preserving edges in the image. The formula for median filtering is: Median Filtered Image (i, j) = median ({I (i+k, j+l)}), fork, l ∈ [-N/2, N/2], where N is the kernel size. The advantages of median filtering include its effectiveness in removing salt-and-pepper noise and its ability to preserve edges better than linear filters. The Small Median Filter Residue (SMFR) technique isolates fine details and noise by computing the difference between the original image and its median-filtered version, followed by another median filtering. This process can be described with the following steps: First, the filtered image is obtained using a median filter. Next, the first difference

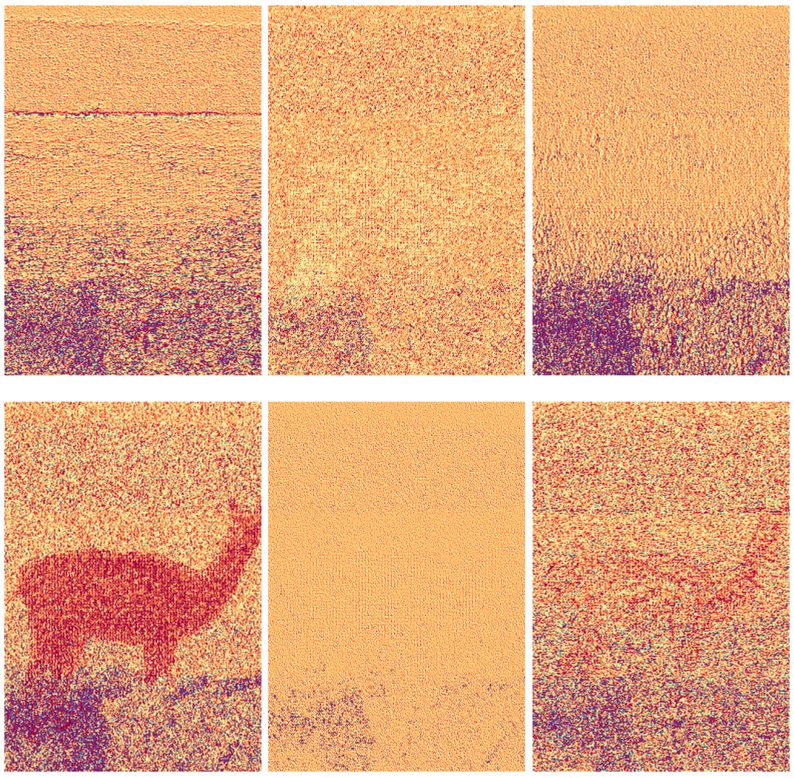

Figure 10: Imaginary coefficients of DC-WT of the original image. On the first row are the first 3 bands and on the second row the rest of the last 3 bands

is calculated as the original image minus the filtered image. Then, a second median filtering is applied to this first difference to obtain the second filtered difference. Finally, SMFR is computed as the first difference minus the second filtered difference. The advantages of SMFR include its ability to enhance fine details and edges while reducing noise and retaining important features. Wiener filtering is a technique that minimizes the mean square error between the estimated and true image by adapting to the local variance of the image. The Wiener filter can be mathematically represented as: Wiener Filtered Image (i, j) = $\mu + (\sigma^2 - \nu^2)/\sigma^2 \times$ (I (i, j) - μ), where μ and σ^2 are the local mean and variance, and ν^2 is the noise variance. The advantages of Wiener filtering include its adaptive noise reduction capabilities and its ability to preserve signal details in regions of low noise. Median modified Wiener was proposed in [17]. This filtering method combines the strengths of median filtering and Wiener filtering to effectively reduce noise and enhance the image. The process begins with the application of a median filter to remove impulsive noise, followed by Wiener filtering to adaptively reduce the remaining noise. The combined formula can be expressed as: First, apply median filter: Median Filtered Image (i, j) = median ({I (i+k, j+l)}). Then, apply Wiener filter to the median-filtered image: Wiener Filtered Median Image (i, j) = $\mu + (\sigma^2 - \nu^2)/\sigma^2 \times$ (Median Filtered Image (i, j) - μ). The advantages of this technique include its effectiveness in removing various types of noise and its ability to preserve edges and fine details.

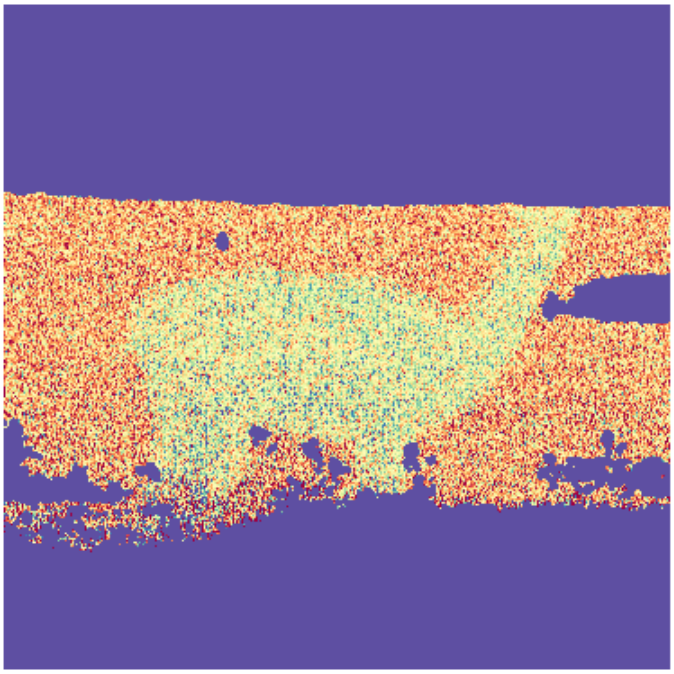

Figure 11: Band 2 Real values for the largest segment. Further on, only the values inside the mask will be used for processing.

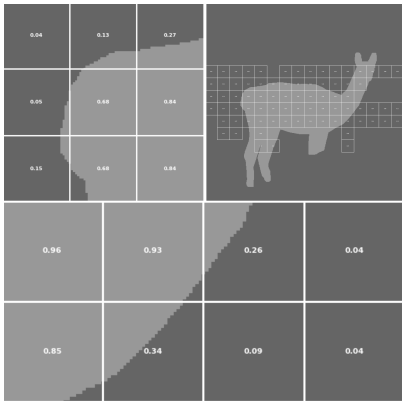

Figure 12: Noise variance (patch means) inside band 2 real coefficients

4.4 Clustering

In the proposed method, because we consistently operate on a single segment color during band decomposition, we assume that there will always be a maximum of three different clusters (2 while there is only one-color segment for entire area – e.g. image of a sky). One cluster represents the ignored area, while another represents high-level features that are evenly distributed within the segmented color. If the clustering results in two different types of texture inside our segmentation mask, it indicates that the segment contains forged areas - Fig 13.

Figure 13: Result after clustering. The dark gray area represents the outside of mask. The darker area and the lighter area represent the 2 distinct regions insisde the mask

4.5 Forged Area Selection

All the steps thus far essentially divide the color segmentation mask into two different regions. The final step is to determine which of these sub-regions is forged and which is not. To make this distinction, the noise variance method is utilized [15]. Noise variance is calculated on the original band matrix. The band matrix is divided into three distinct regions, and the noise variance is computed for each region as well as for the entire band matrix. Using these four values, the outlier, which indicates the forged area, is identified - FIGURE 14.

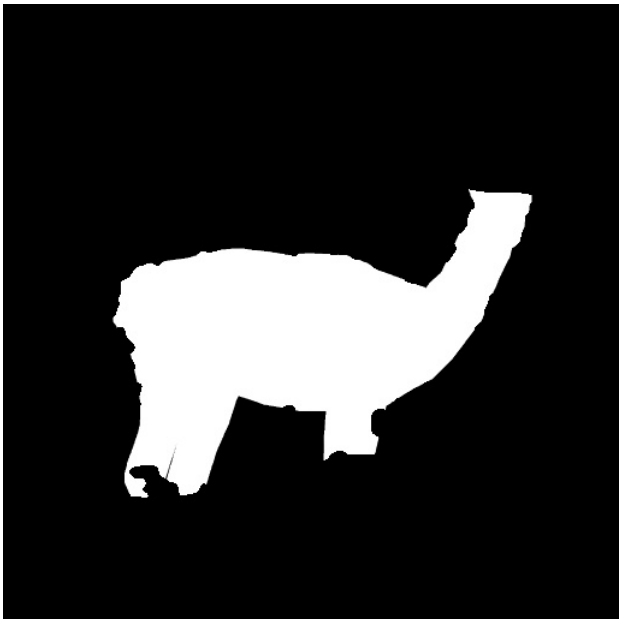

Figure 14: Final mask per band

5. EXPERIMENTAL RESULTS

5.1 Method Parameters

This chapter presents the experimental evaluation of the proposed inpainting forgery detection method employing DT-CWT and HFS. The settings used are the following: for DT-CWT the Kingsbury Q-shift filters are used (standard near-symmetric 13,19 + Q-Shift 18,18); for normalization the log normalization is employed (actually symmetric log – similar to what Matplotlib is doing); for HFS segmentations a SLIC value of 32 is used; for patch mean computation a patch size of 32 is used no padding with a tolerance of 0.5; for K-Means the cluster is either set to 3 or to 2 (when there is only on segment in the entire image – for e.g. a clear blue sky or a green field); for noise variance between areas a patch size of 8 is used with a padding of 3.

5.2 Dataset

This section examines the results of various inpainting detection techniques applied to different image inpainting processes, with a primary focus on the dataset. While other forgery datasets like Casia and MFC exist, they do not specifically address image inpainting or object removal. One of the similar datasets to what we proposed is the work done in [18], but they have some limitations. They used the [19], as the base dataset, but the difference is that they take all segmented objects and apply removal. By doing this process automatically, not all objects that are removed can be easily removed due to the highly texturized area. Also, regarding the removal method they use only one classical method [20], thus making the detection more focused on detecting the artifacts introduced by that specific inpainting method. Another similar approach were proposed in [21],

and in [22]. The problem with these 2 datasets are the following: first the image sizes are smaller and secondly the object removal mask is generated arbitrary, thus making the inpainting methods introduced a lot of visual traces (specifically blur inconsistencies), thus making the detection easier. a To address this, Google's Open Images Dataset V7 (released in October 2022) was used [23]. From this dataset, 1k images (increased from 400 from the original paper) with segmented masks were manually selected, ensuring each image used only one mask and avoided highly texturized areas, as inpainting methods struggle with such regions. The selected images were limited to a maximum size of 1024x1024 pixels. The dataset included a variety of object sizes to be removed, from small to large. Additionally, to improve inpainting results, a dilation with a 5x5 kernel was added to masks that were close to the image borders. The complete dataset, including the original images, masks, inpainting results, and forgery detection results, can be accessed at https://github.com/jmaba/ImageInpaitingDetectionAReview

5.3 Evaluation Metrics

This section evaluates the performance of the proposed method for detecting tampered regions. The algorithm's effectiveness is assessed by comparing its output binary mask (indicating tampered pixels) with ground truth masks of tampered regions in the test images. Three key metrics, commonly used in image segmentation tasks, are employed to quantify performance: accuracy, recall, and intersection over union (IoU). These metrics are calculated at the pixel level for each test image, providing a thorough evaluation of the algorithm's ability to accurately identify tampered pixels:

- **Accuracy**: Measures the proportion of correctly classified pixels, including both tampered and untampered pixels. It reflects the algorithm's overall effectiveness in distinguishing between the two classes. A high accuracy value indicates strong performance in identifying both true positives and avoiding false positives/negatives.
- **Intersection over Union (IoU)**: Combines precision and recall into a single metric. It represents the overlap area between the predicted tampered mask and the ground truth mask, divided by the total area of their union. A high IoU value signifies a good balance between precision and recall, indicating the algorithm accurately identifies most tampered pixels while minimizing false positives.
- **Recall**: Measures the proportion of actual tampered pixels that are correctly identified by the algorithm. It reflects the algorithm's ability to detect true positives among the tampered pixels. A high recall value indicates that the method is successful in capturing most of the tampered regions, although it may also capture false positives.

5.4 Results

The proposed method was evaluated using an extended version of the publicly available inpainting forgery detection dataset. The results highlight the method's effectiveness in identifying tampered regions across a broader range of image content and inpainting techniques (FIGURE 15 - FIGURE 18). We have selected several images, including the originals, masks, and inpainting results, for demonstration. For these we have applied the detection methods proposed in [21], hereafter

referred to as IID, [24], hereafter referred to as PSCCNET respectively [25], named FOCAL. Additionally, we've included also a newer method called TruFor [26], which has a similar idea to detect forgeries based on noise inconsistencies analysis by leveraging a trained neural network.

Figure 15: Original image on left, in the middle the mask of the object to be removed (also used for computing metrics), the altered image

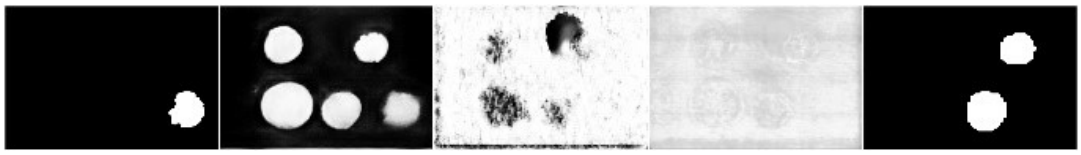

Figure 16: First image is our generated mask, second is the mask output TruFor, third is the mask generated by IID, forth is the mask output by PSCNET and last one is mask generated by FOCAL

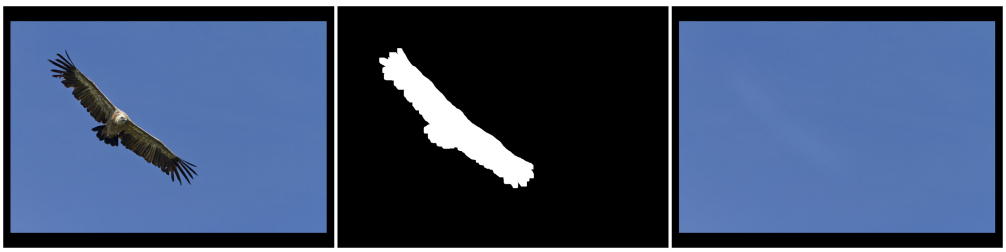

Figure 17: Original image on left, in the middle the mask of the object to be removed (also used for computing metrics), the altered image

Figure 18: First image is our generated mask, second is the mask output by TRUFOR, third is the mask generated by IID, forth is the mask output by PSCNET and last one is mask generated by FOCAL

From the above images and from TABLE 2 results, our method and IID and TRUFOR yield good results. FOCAL on the other hand yields false positive results, while PSCCNET, depending how the mask interval is analyzed, might be able to detect some portions of the mask. Analysis of the IID detection method reveals high recall values, indicating successful identification of most forged images. However, accuracy and Intersection over Union (IoU) metrics are low, suggesting that the method frequently misclassifies entire images as forged rather than precisely locating the

Table 2: Inpainting detection Results.

Detection	LAMA inpainting method		
	Accuracy	***Recall***	***iOu***
Proposed method	0.93	0.89	0.67
FOCAL	0.75	0.15	0.10
IID	0.49	0.73	0.26
FOCAL	0.58	0.60	0.21
TRUFOR	0.64	0.69	0.63

tampered regions. These conclusions can be seen on the entire dataset, not particularly to the LAMA inpainting method. While IID detection shows decent performance in some cases, it generally exhibits lower IoU scores compared to ours. IID detection tends to have high recall scores, indicating its ability to capture a large portion of the actual forged regions. The high recall values usually obtained by IID algorithms are because the method yields not a binary mask for inpainting detection, but rather for each pixel a confidence score. In some situations, the method seems to falsely attribute altered values to mostly all the pixels inside the image. IID accuracy and IoU scores vary across different inpainting methods. The PSCCNET detection method demonstrates relatively moderate performance across all inpainting methods, with accuracy, recall, and IoU scores falling between those of FOCAL and our method. The proposed method demonstrated promising results for inpainting forgery detection across various inpainting techniques. It achieved consistently high accuracy, suggesting its effectiveness in distinguishing tampered from original regions. However, the effectiveness of all methods, including ours, varied depending on the specific inpainting technique used. Regarding the TRUFOR accuracy results, given that accuracy is relatively low, this suggests that there may be a notable number of errors in classification (either false positives or false negatives). However, it is less informative on its own and should be considered alongside recall and IoU. Given that recall is higher than accuracy and IoU, TRUFOR model might be over-predicting manipulated areas (false positives). An IoU of 0.63 suggests that there is some level of mismatch between the predicted and ground truth masks. The score is moderate, indicating that while your model finds many of the manipulated areas, the segmentation might not be perfectly aligned, with some inaccuracies around the boundaries or missing parts.

5.5 Evaluation on IID Dataset

After proving the effectiveness of the proposed method on the given dataset, we have investigated also the results on several other datasets. The dataset to be analyzed is the IID dataset [21]. Several key factors to considered:

- all images have been resized to 256x256 – thus mostly of the images will appear as blurred / unclear
- although several inpainted methods have been employed – they are applied on random generated masks – thus making the inpainting method generate a lot of artifacts like blurring, staircase effect, inconsistent textures, colors, incorrect object continuation etc. [27].

Figure 19: Images from IID dataset: first image is the inpainted one, and second image is the mask

Due to the size of the images (FIGURE 19), the standard filter used (near-symmetric 13,19 tap filters + BP filter [28]) in the previous dataset must be replaced by shorter variant to better enhance the forged area (Antonini 9,7 tap filters [29]). As one can observed from the below figures - FIGURE 20 vs. FIGURE 21, the Antonini filters better enhanced the forged areas:

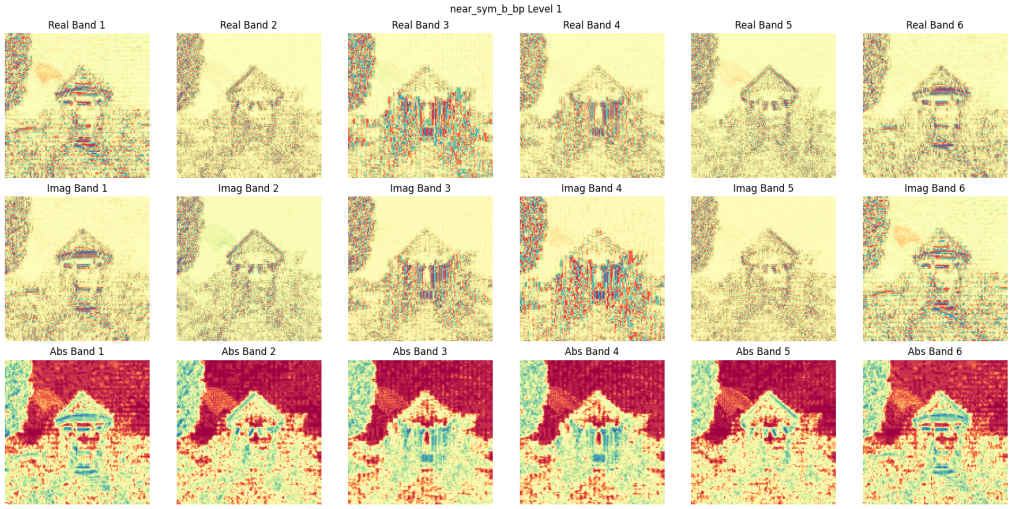

Figure 20: DCWT applied on inpainted image using near-symmetric 13,19 tap filters + BP filter. 1^{st} row represents the real part, 2^{nd} row represents the imaginary part, while 3^{rd} row represents the absolute part. Columns 1-6 represent the orrientation of the decomposition.

Another important change in the above algorithm is the segmentation SLIC parameter. Because mostly the images tend to be blurred, the SLIC parameter must be lowered to 8. Also, again due to the blurriness of the images the number of clusters must be increased (FIGURE 22). In the below images the intermediate processing results for the given test image are presented, along with the overall results FIGURE 23.

From the overall results in TABLE 3, it can be the high recall and accuracy indicate that the model is very good at identifying and classifying objects. However, the relatively low IoU points to

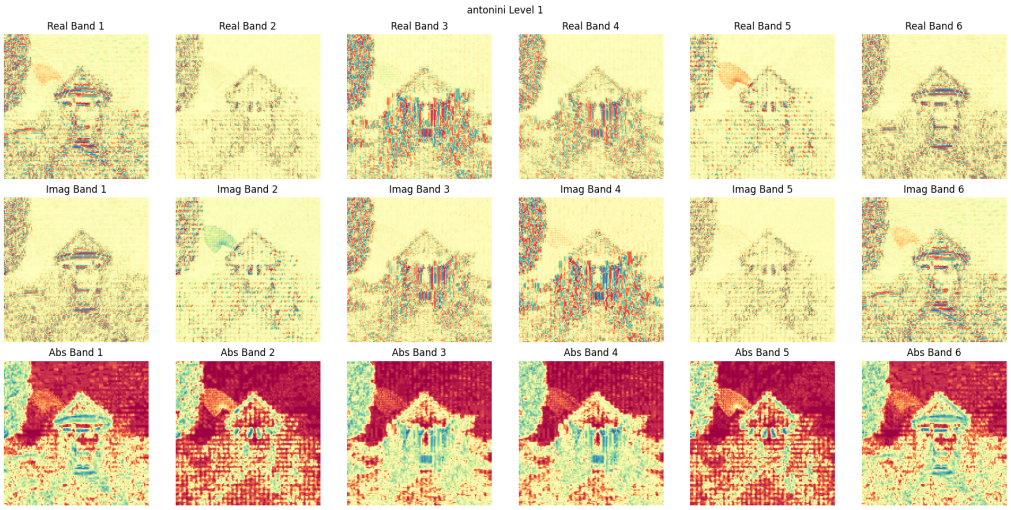

Figure 21: DCWT applied on inpainted image using Antonini filters. 1^{st} row represents the real part, 2^{nd} row represents the imaginary part, while 3^{rd} row represents the absolute part. Columns 1-6 represent the orrientation of the decomposition.

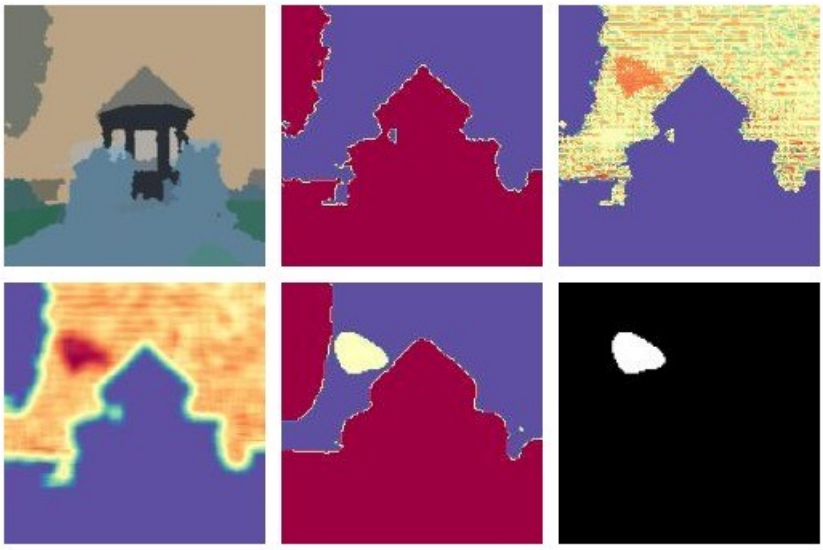

Figure 22: First row: first image represents the texture segmentation. Second image represents the area to be further analyzed. Third image represents the DCWT band 1 real with the mask from previous image. Second row: MMWF enhancemen on the previous image, 5^{th} image is the result of noise and clusterization, last image the proposed mask.

potential issues with the precision of object localization or segmentation, suggesting that the model's bounding boxes or masks do not perfectly align with the ground truth. The low IoU is since the test images are 256x256 and contain a lot of blurring artifacts, which affect the color segmentation and the noise analysis of the DTCWT coefficients. Probably by fine tuning our proposed method

Figure 23: 1^{st} image is the truth mask, second image is the IID result while 3^{rd} image is the one generated by our proposed method.

parameters, the results can be increased in terms of IoU. Comparing with IID results, high recall in our proposed method indicates that our approach is excellent at detecting all relevant positives, making it ideal in scenarios where missing positives is critical, but lower IoU (0.49 vs 0.68) suggests that while it detects positives well, it struggles with precise localization, leading to less accurate boundaries and higher overlap errors.

Table 3: proposed detection method on IID dataset.

Method	**IID Dataset detection**		
	Accuracy	***Recall***	***iOu***
Our proposed method	0.85	0.94	0.49
IID Method	0.9	0.68	0.68

6. CONCLUSION

Future studies should incorporate a broader range of inpainting methods to provide a more comprehensive assessment of how well the proposed and baseline methods perform across various scenarios. Expanding the dataset to include newer inpainting techniques would enhance the robustness of the inpainting forgery detection evaluation framework. While accuracy is an important metric for evaluating detection performance, the consistently high accuracy values achieved by the proposed method might indicate a bias towards correctly classifying untampered pixels. Future research could employ additional metrics, such as recall and IoU, which focus more on accurately identifying tampered regions.

An IoU greater than 0.5 is generally considered a good detection result, but the observed variations in recall and IoU across different inpainting techniques highlight potential limitations. The methods, including the proposed one, might not be equally effective in detecting artifacts introduced by various inpainting techniques. Additionally, processing time, particularly for the noise estimation method and clustering, is a significant limitation that needs to be addressed.

Another limitation is that some configuration values were obtained through extensive testing. These parameters may vary for other types of inpainting, suggesting the need for a machine learning-based approach to automatically adjust parameters for different scenarios. Employing adaptive methods or leveraging machine learning for parameter tuning could improve the robustness and effectiveness of forgery detection across various inpainting techniques. Furthermore, the proposed method is highly susceptible to different types of post-processing, such as resizing, blurring, and noise addition. These post-processing techniques can significantly affect the detection performance, highlighting the need for more resilient detection methods capable of handling various image manipulations.

References

[1] Criminisi A, Pérez P, Toyama K. Region Filling and Object Removal by Exemplar-Based Image Inpainting. IEEE Trans Image Process. 2004;13:1200-1212.

[2] Papafitsoros K, Schoenlieb CB, Sengul B. Combined First and Second Order Total Variation Inpainting Using Split Bregman. Image Processing On Line. 2013;3:112-136.

[3] Yu J, Lin Z, Yang J, Shen X, Lu X, et al. Generative Image Inpainting With Contextual Attention. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2018:5505-5514.

[4] Corneanu C, Gadde R, Martinez AM. Latent Paint: Image Inpainting in Latent Space With Diffusion Models. Proceedings IEEE Winter Conference on Applications of Computer Vision, WACV. 2024:4322-4331.

[5] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-Resolution Image Synthesis With Latent Diffusion Models. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2022;2022:10674-10685.

[6] Selesnick IW, Baraniuk RG, Kingsbury NC. The Dual-Tree Complex Wavelet Transform. IEEE Signal Process Mag. 2005;22:123-151.

[7] Julliand T, Nozick V, Talbot H. Image Noise and Digital Image Forensics. Lect Notes Comput Sci. 2016;9569:3-17.

[8] Zeng H, Peng A, Lin X. Exposing Image Splicing With Inconsistent Sensor Noise Levels. Multimed Tools Appl. 2020;79:26139-25154.

[9] Liu Q, Li H, Liu Z. Image Forgery Localization Based on Fully Convolutional Network With Noise Feature. Multimed Tools Appl. 2022;81:17919-17935.

[10] Zhang Z, Qian Y, Zhao Y, Zhang X, Zhu L, et al. Noise and Edge Based Dual Branch Image Manipulation Detection. In Proceedings of the 2023 4th International Conference on Computing, Networks and Internet of Things (CNIOT '23). Association for Computing Machinery. 2023:963-968.

[11] Chen J, Liao X, Wang W, Qian Z, Qin Z, et al. Snis: A Signal Noise Separation-Based Network for Post-processed Image Forgery Detection. IEEE Trans Circuits Syst For Video Technol. 2022;33:935-951.

[12] Gardella M, Muse P, Morel JM, Colom M. Noisesniffer: A Fully Automatic Image Forgery Detector Based on Noise Analysis. In 2021 IEEE International Workshop on Biometrics and Forensics (IWBF) IEEE. 2021:1-6.

[13] Suvorov R, Logacheva E, Mashikhin A, Remizova A, Ashukha A, et al. Resolution-Robust Large Mask Inpainting With Fourier Convolutions. In proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022:3172-3182.

[14] Kuznetsova A, Rom H, Alldrin N, Uijlings J, Krasin I, et al. The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale. Int J Comput Vis. 2020;128:1956-1981.

[15] Chen G, Zhu F, Ann Heng P. An Efficient Statistical Method for Image Noise Level Estimation. In Proceedings of the IEEE International Conference on Computer Vision. 2015:477-485.

[16] Cheng MM, Liu Y, Hou Q, Bian J, Torr P, et al. HFS: Hierarchical Feature Selection for Efficient Image Segmentation. In: Leibe B, Matas J, Sebe N, Welling M, (eds). Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science. Springer 2016;9907:867-882.

[17] Cannistraci CV, Montevecchi FM, Alessio M. Median-Modified Wiener Filter Provides Efficient Denoising, Preserving Spot Edge and Morphology in 2-DE Image Processing. Proteomics. 2009;9:4908-4919.

[18] Mahfoudi G, Tajini B, Retraint F, Morain-Nicolier F, Dugelay JL, et al. Defacto: Image and Face Manipulation Dataset. 27th European Signal Processing Conference.2019:1-5.

[19] Lin TY, Maire M, Belongie S, Hays J, Perona P, et al. Microsoft COCO: Common Objects in Context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds). Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science. Springer. 2014;8693:740-755.

[20] Daisy M, Buyssens P, Tschumperle D, Lezoray O. A Smarter Exemplar-Based Inpainting Algorithm Using Local and Global Heuristics for More Geometric Coherence. IEEE International Conference on Image Processing, ICIP. 2014:4622-4626.

[21] Wu H, Zhou J. IID-Net: Image Inpainting Detection Network via Neural Architecture Search and Attention. IEEE Trans Circuits Syst For Video Technol. 2021;32:1172-1185.

[22] Ding X, Deng Y, Zhao Y, Zhu W. Aftlnet: An Efficient Adaptive Forgery Traces Learning Network for Deep Image Inpainting Localization. J Inf Sec Appl. 2024;84:103825.

[23] Benenson R, Popov S, Ferrari V. Large-Scale Interactive Object Segmentation With Human Annotators. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019:11700-11709.

[24] Liu X, Liu Y, Chen J, Liu X. PSCC-Net: Progressive Spatio-Channel Correlation Network for Image Manipulation Detection and Localization. IEEE Trans Circuits Syst For Video Technol. 2022;32:7505-7517.

[25] Wu H, Chen Y, Zhou J. Rethinking Image Forgery Detection via Contrastive Learning and Unsupervised Clustering. 2023. ArXiv preprint: https://arxiv.org/pdf/2308.09307

[26] Guillaro F, Cozzolino D, Sud A, Dufour N, Verdoliva L. Trufor: Leveraging All-Round Clues for Trustworthy Image Forgery Detection and Localization. In proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:20606-20615.

[27] Barglazan AA, Brad R, Constantinescu C. Image Inpainting Forgery Detection: A Review. J Imaging. 2024;10:42.

[28] Rotation-Invariant Local Feature Matching With Complex Wavelets. In 2006 14th European Signal Processing Conference. IEEE. 2006:1-5.

[29] Antonini M, Barlaud M, Mathieu P, Daubechies I. Image Coding Using Wavelet Transform. IEEE Trans Image Process. 1992;1:205-220.