DeepFake Detection Using Wavelet Packets with Vision Transformer (WPT-ViT)

Osama Rawy and AbdulRahman AlTahhan

University of Leeds, School of Computing, ODL MSc in AI, UK.
WWW home page:https://github.com/osmahus/WPT-ViT

Abstract. The latest advances in generative algorithms have raised the quality of virtually created images and videos to the point that it has become very difficult to distinguish the real from the generated ones (Deep-Fake). This stimulated hot research to build better models to detect DF. Our paper proposes a new DNN model to detect DF images using a wavelet packet transformer and a vision transformer (WPT-ViT). The study shows that attention could be found between the WPT decompositions of an image even without slicing the image into spatial patches, which is a novel modification to the original VIT model. We showed that by using smaller model sizes and lower GPU and CPU requirements, we can achieve comparable results with previous work in this research area, The model was trained and tested using two datasets, "CIFAKE" and "140k Real and Fake Faces," which are generated using StyleGAN and Stable Diffusion algorithms.

Keywords: Deep Fake Detection, Wavelet Packets, Vision Transformer

1 Introduction

In 2014, Goodfellow et al. introduced the Generative Adversarial Network (GAN), marking the beginning of the generative AI (GAI) era. Since then, researchers have shifted their focus from discriminative learning to generative learning. This wave brought various vision-generative applications to the market, such as Midjourney, Firefly, DALL-E2, and Imagen[1]. Such applications were developed using state-of-the-art architectures like GAN, Variational Autoencoders, and Diffusion to generate images and videos with high fidelity and diversity, mimicking real-world photos and videos [2]. Vision-generative technologies have shown high value in several domains. In the entertainment industry, for example, they could generate complete scenes that would otherwise be very risky for actors to perform or prohibitively expensive to produce. In Education, they could bring historical characters to life to talk with students for an immersive learning experience; similarly, the list of positive uses continues in other fields like manufacturing and marketing. However, this ability to produce synthetic content with realistic flavor was termed Deep Fake due to the unfortunate incidents in which these technologies were used to attack people through identity theft, character assassination, and faked pornography. On a larger scale, deep fakes were also used to

spread misinformation, fake news, and communal hatred. According to a report released in April 2021 by Cybernews, deep fake content over the internet doubles every six months, posing a significant threat that needs to be addressed urgently [5].
For that reason, there has been significant attention in both academic and industrial fields on finding ways to detect deep fakes with high accuracy and performance. For example, Facebook, Microsoft, and Amazon collaborated to launch the Deep Fake Detection Challenge (DFDC) on Kaggle from 2019 to 2020. A survey conducted by (Liang & Xue, 2024) showed that the number of publications on deep fake detection surpassed the number of publications on deep fake generation in 2022 and 2023 [11].
This paper introduces a new deepfake detection tool that combines the strengths of wavelet analysis to extract important image features and Vision Transformer (VIT) to create lighter models with lower GPU and CPU requirements than CNN counterparts.

2 Literature Review

When it comes to detecting deep fakes, it can be seen as a binary classification problem involving training a machine learning model on a dataset of real and fake examples. This includes extracting relevant features from the data and using these features to predict the authenticity of the content (Patel et al., 2023). Previous research has suggested different ways to categorize work in the deepfake detection field. (Patel et al., 2023) highlighted three approaches for detecting deepfakes in video and images.

The first approach involves using handcrafted algorithms to extract features of visual artifacts, such as inconsistent head poses or unusual eye blinking. The results of the feature extractor could then be passed to any classifier, such as SVM or NN, to perform the detection. An example of this approach is the work of Matern et al. (2019), which has an AUC of 0.866.

The second approach works on the pixel level to extract spatial features related to visual inconsistencies using local feature detectors (like SIFT and HOG) or steganography detectors. An example of this approach is the two-stream network proposed by (Zhou et al., 2018) with an AUC of 0.927. However, the effectiveness of the first two approaches has been reduced by the fact that the latest deepfake datasets were created using advanced image generation techniques, which decreases the likelihood of producing visual artifacts or detectable local features.

The third approach utilizes Deep Neural Networks to understand the intricate patterns and features present in the training dataset. The detection results are more accurate when a more relevant and comprehensive dataset is provided. According to Rana et al. (2022), previous deep fake detection models can be grouped into three categories: Machine learning, deep learning, and statistical. Their research indicated that 77

In a study conducted by Wang et al. (2024), it was demonstrated that while Convolutional Neural Networks (CNNs) are commonly used in deepfake detection to capture spatial relationships within images, making them effective in identifying facial manipulations and other visual irregularities at the frame level, Vision Transformers (ViTs) have distinct advantages in analyzing and comprehending the intricate details of deepfake images and videos. ViTs are especially good at understanding the overall structure of an image to identify inconsistencies or anomalies suggestive of manipulation.

However, Wang also highlighted the challenges faced by standalone ViT models in deepfake detection, such as their struggle to generalize across diverse datasets, their need for extensive training data, their difficulty in maintaining temporal consistency in video deepfakes, their limited ability to capture local spatial information, and their potential inability to fully capture the temporal and sequential dependencies present in video data. To address these limitations, hybrid models that combine ViTs with other techniques, such as CNNs or RNNs, are often utilized.

3 Fixed-Period Problems: The Sublinear Case

With this chapter, the preliminaries are over, and we begin the search for periodic solutions to Hamiltonian systems. All this will be done in the convex case; that is, we shall study the boundary-value problem

$$\begin{aligned} \dot{x} &= JH'(t,x) \\ x(0) &= x(T) \end{aligned}$$

with $H(t,\cdot)$ a convex function of x, going to $+\infty$ when $\|x\| \to \infty$.

3.1 Autonomous Systems

In this section, we will consider the case when the Hamiltonian $H(x)$ is autonomous. For the sake of simplicity, we shall also assume that it is C^1.

We shall first consider the question of nontriviality, within the general framework of (A_∞, B_∞)-subquadratic Hamiltonians. In the second subsection, we shall look into the special case when H is $(0, b_\infty)$-subquadratic, and we shall try to derive additional information.

The General Case: Nontriviality. We assume that H is (A_∞, B_∞)-subquadratic at infinity, for some constant symmetric matrices A_∞ and B_∞, with $B_\infty - A_\infty$ positive definite. Set:

$$\gamma := \text{smallest eigenvalue of } \; B_\infty - A_\infty \tag{1}$$

$$\lambda := \text{largest negative eigenvalue of } \; J\frac{d}{dt} + A_\infty \; . \tag{2}$$

Theorem 1 tells us that if $\lambda + \gamma < 0$, the boundary-value problem:

$$\begin{aligned} \dot{x} &= JH'(x) \\ x(0) &= x(T) \end{aligned} \tag{3}$$

has at least one solution $\overline{x}$, which is found by minimizing the dual action functional:

$$\psi(u) = \int_o^T \left[\frac{1}{2} \left(\Lambda_o^{-1} u, u \right) + N^*(-u) \right] dt \tag{4}$$

on the range of Λ, which is a subspace $R(\Lambda)_L^2$ with finite codimension. Here

$$N(x) := H(x) - \frac{1}{2} \left(A_\infty x, x \right) \tag{5}$$

is a convex function, and

$$N(x) \leq \frac{1}{2} \left(\left(B_\infty - A_\infty \right) x, x \right) + c \quad \forall x \ . \tag{6}$$

Proposition 1. *Assume $H'(0) = 0$ and $H(0) = 0$. Set:*

$$\delta := \liminf_{x \to 0} 2N(x) \left\| x \right\|^{-2} \ . \tag{7}$$

If $\gamma < -\lambda < \delta$, the solution $\overline{u}$ is non-zero:

$$\overline{x}(t) \neq 0 \quad \forall t \ . \tag{8}$$

Proof. Condition (7) means that, for every $\delta' > \delta$, there is some $\varepsilon > 0$ such that

$$\left\| x \right\| \leq \varepsilon \Rightarrow N(x) \leq \frac{\delta'}{2} \left\| x \right\|^2 \ . \tag{9}$$

It is an exercise in convex analysis, into which we shall not go, to show that this implies that there is an $\eta > 0$ such that

$$f \left\| x \right\| \leq \eta \Rightarrow N^*(y) \leq \frac{1}{2\delta'} \left\| y \right\|^2 \ . \tag{10}$$

Fig. 1. This is the caption of the figure displaying a white eagle and a white horse on a snow field

Since u_1 is a smooth function, we will have $\|hu_1\|_\infty \leq \eta$ for h small enough, and inequality (10) will hold, yielding thereby:

$$\psi(hu_1) \leq \frac{h^2}{2}\frac{1}{\lambda}\|u_1\|_2^2 + \frac{h^2}{2}\frac{1}{\delta'}\|u_1\|^2 \ . \tag{11}$$

If we choose δ' close enough to δ, the quantity $\left(\frac{1}{\lambda} + \frac{1}{\delta'}\right)$ will be negative, and we end up with

$$\psi(hu_1) < 0 \qquad \text{for} \quad h \neq 0 \ \text{ small} \ . \tag{12}$$

On the other hand, we check directly that $\psi(0) = 0$. This shows that 0 cannot be a minimizer of ψ, not even a local one. So $\overline{u} \neq 0$ and $\overline{u} \neq \Lambda_o^{-1}(0) = 0$. □

Corollary 1. *Assume H is C^2 and (a_∞, b_∞)-subquadratic at infinity. Let ξ_1, $\ldots, \xi_N$ be the equilibria, that is, the solutions of $H'(\xi) = 0$. Denote by ω_k the smallest eigenvalue of $H''(\xi_k)$, and set:*

$$\omega := \text{Min}\ \{\omega_1, \ldots, \omega_k\} \ . \tag{13}$$

If:

$$\frac{T}{2\pi}b_\infty < -E\left[-\frac{T}{2\pi}a_\infty\right] < \frac{T}{2\pi}\omega \tag{14}$$

then minimization of ψ yields a non-constant T-periodic solution $\overline{x}$.

We recall once more that by the integer part $E[\alpha]$ of $\alpha \in \mathbb{R}$, we mean the $a \in \mathbb{Z}$ such that $a < \alpha \leq a + 1$. For instance, if we take $a_\infty = 0$, Corollary 2 tells us that $\overline{x}$ exists and is non-constant provided that:

$$\frac{T}{2\pi}b_\infty < 1 < \frac{T}{2\pi} \tag{15}$$

or

$$T \in \left(\frac{2\pi}{\omega}, \frac{2\pi}{b_\infty}\right) \ . \tag{16}$$

Proof. The spectrum of Λ is $\frac{2\pi}{T}\mathbb{Z} + a_\infty$. The largest negative eigenvalue λ is given by $\frac{2\pi}{T}k_o + a_\infty$, where

$$\frac{2\pi}{T}k_o + a_\infty < 0 \leq \frac{2\pi}{T}(k_o + 1) + a_\infty \ . \tag{17}$$

Hence:

$$k_o = E\left[-\frac{T}{2\pi}a_\infty\right] \ . \tag{18}$$

The condition $\gamma < -\lambda < \delta$ now becomes:

$$b_\infty - a_\infty < -\frac{2\pi}{T}k_o - a_\infty < \omega - a_\infty \tag{19}$$

which is precisely condition (14). □

Lemma 1. *Assume that H is C^2 on $\mathbb{R}^{2n}\backslash\{0\}$ and that $H''(x)$ is non-degenerate for any $x \neq 0$. Then any local minimizer $\widetilde{x}$ of ψ has minimal period T.*

Proof. We know that $\widetilde{x}$, or $\widetilde{x} + \xi$ for some constant $\xi \in \mathbb{R}^{2n}$, is a T-periodic solution of the Hamiltonian system:

$$\dot{x} = JH'(x) \ . \tag{20}$$

There is no loss of generality in taking $\xi = 0$. So $\psi(x) \geq \psi(\widetilde{x})$ for all $\widetilde{x}$ in some neighbourhood of x in $W^{1,2}\left(\mathbb{R}/T\mathbb{Z}; \mathbb{R}^{2n}\right)$.

But this index is precisely the index $i_T(\widetilde{x})$ of the T-periodic solution $\widetilde{x}$ over the interval $(0, T)$, as defined in Sect. 2.6. So

$$i_T(\widetilde{x}) = 0 \ . \tag{21}$$

Now if $\widetilde{x}$ has a lower period, T/k say, we would have, by Corollary 31:

$$i_T(\widetilde{x}) = i_{kT/k}(\widetilde{x}) \geq k i_{T/k}(\widetilde{x}) + k - 1 \geq k - 1 \geq 1 \ . \tag{22}$$

This would contradict (21), and thus cannot happen. □

Notes and Comments. The results in this section are a refined version of [**?**]; the minimality result of Proposition 14 was the first of its kind.

To understand the nontriviality conditions, such as the one in formula (16), one may think of a one-parameter family x_T, $T \in \left(2\pi\omega^{-1}, 2\pi b_\infty^{-1}\right)$ of periodic solutions, $x_T(0) = x_T(T)$, with x_T going away to infinity when $T \to 2\pi\omega^{-1}$, which is the period of the linearized system at 0.

Table 1. This is the example table taken out of *The TEXbook,* p. 246

Year	World population
8000 B.C.	5,000,000
50 A.D.	200,000,000
1650 A.D.	500,000,000
1945 A.D.	2,300,000,000
1980 A.D.	4,400,000,000

Theorem 1 (Ghoussoub-Preiss). *Assume $H(t, x)$ is $(0, \varepsilon)$-subquadratic at infinity for all $\varepsilon > 0$, and T-periodic in t*

$$H(t, \cdot) \quad \text{is convex} \quad \forall t \tag{23}$$

$$H(\cdot, x) \quad \text{is} \quad T-\text{periodic} \quad \forall x \tag{24}$$

$$H(t, x) \geq n\left(\|x\|\right) \quad \text{with} \quad n(s)s^{-1} \to \infty \quad \text{as} \quad s \to \infty \tag{25}$$

$$\forall \varepsilon > 0 \ , \quad \exists c \ : \ H(t,x) \leq \frac{\varepsilon}{2} \|x\|^2 + c \ . \tag{26}$$

Assume also that H is C^2, and $H''(t,x)$ is positive definite everywhere. Then there is a sequence x_k, $k \in \mathbb{N}$, of kT-periodic solutions of the system

$$\dot{x} = JH'(t,x) \tag{27}$$

such that, for every $k \in \mathbb{N}$, there is some $p_o \in \mathbb{N}$ with:

$$p \geq p_o \Rightarrow x_{pk} \neq x_k \ . \tag{28}$$

□

*Example 1 (*External forcing*).* Consider the system:

$$\dot{x} = JH'(x) + f(t) \tag{29}$$

where the Hamiltonian H is $(0, b_\infty)$-subquadratic, and the forcing term is a distribution on the circle:

$$f = \frac{d}{dt}F + f_o \qquad \text{with} \quad F \in L^2\left(\mathbb{R}/T\mathbb{Z}; \mathbb{R}^{2n}\right) \ , \tag{30}$$

where $f_o := T^{-1} \int_o^T f(t)dt$. For instance,

$$f(t) = \sum_{k \in \mathbb{N}} \delta_k \xi \ , \tag{31}$$

where δ_k is the Dirac mass at $t = k$ and $\xi \in \mathbb{R}^{2n}$ is a constant, fits the prescription. This means that the system $\dot{x} = JH'(x)$ is being excited by a series of identical shocks at interval T.

Definition 1. *Let $A_\infty(t)$ and $B_\infty(t)$ be symmetric operators in $\mathbb{R}^{2n}$, depending continuously on $t \in [0,T]$, such that $A_\infty(t) \leq B_\infty(t)$ for all t.*

A Borelian function $H : [0,T] \times \mathbb{R}^{2n} \to \mathbb{R}$ is called (A_∞, B_∞)-subquadratic at infinity if there exists a function $N(t,x)$ such that:

$$H(t,x) = \frac{1}{2}\left(A_\infty(t)x, x\right) + N(t,x) \tag{32}$$

$$\forall t \ , \quad N(t,x) \qquad \text{is convex with respect to} \quad x \tag{33}$$

$$N(t,x) \geq n\left(\|x\|\right) \qquad \text{with} \quad n(s)s^{-1} \to +\infty \quad \text{as} \quad s \to +\infty \tag{34}$$

$$\exists c \in \mathbb{R} \ : \quad H(t,x) \leq \frac{1}{2}\left(B_\infty(t)x, x\right) + c \quad \forall x \ . \tag{35}$$

If $A_\infty(t) = a_\infty I$ and $B_\infty(t) = b_\infty I$, with $a_\infty \leq b_\infty \in \mathbb{R}$, we shall say that H is (a_∞, b_∞)-subquadratic at infinity. As an example, the function $\|x\|^\alpha$, with $1 \leq \alpha < 2$, is $(0, \varepsilon)$-subquadratic at infinity for every $\varepsilon > 0$. Similarly, the Hamiltonian

$$H(t,x) = \frac{1}{2}k\|k\|^2 + \|x\|^\alpha \tag{36}$$

is $(k, k+\varepsilon)$-subquadratic for every $\varepsilon > 0$. Note that, if $k < 0$, it is not convex.

Notes and Comments. The first results on subharmonics were obtained by Rabinowitz in [**?**], who showed the existence of infinitely many subharmonics both in the subquadratic and superquadratic case, with suitable growth conditions on H'. Again the duality approach enabled Clarke and Ekeland in [**?**] to treat the same problem in the convex-subquadratic case, with growth conditions on H only.

Recently, Michalek and Tarantello (see [**?**] and [**?**]) have obtained lower bound on the number of subharmonics of period kT, based on symmetry considerations and on pinching estimates, as in Sect. 5.2 of this article.

References

1. Bengesi, Staphord, et al. "Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers." IEEE Access (2024).
2. Raut, Gaurav, and Apoorv Singh. "Generative AI in Vision: A Survey on Models, Metrics, and Applications." arXiv preprint arXiv:2402.16369 (2024).
3. Heidari, Arash, Nima, Jafari Navimipour, Hasan, Dag, Mehmet, Unal. "Deepfake detection using deep learning methods: A systematic and comprehensive review". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 14. 2(2024): e1520.
4. Patel, Yogesh, Sudeep, Tanwar, Rajesh, Gupta, Pronaya, Bhattacharya, Innocent Ewean, Davidson, Royi, Nyameko, Srinivas, Aluvala, Vrince, Vimal. "Deepfake Generation and Detection: Case Study and Challenges". IEEE Access. (2023).
5. Pei, Gan, Jiangning, Zhang, Menghan, Hu, Guangtao, Zhai, Chengjie, Wang, Zhenyu, Zhang, Jian, Yang, Chunhua, Shen, Dacheng, Tao. "Deepfake Generation and Detection: A Benchmark and Survey". arXiv preprint arXiv:2403.17881. (2024).
6. Zhang, Tao. "Deepfake generation and detection, a survey". Multimedia Tools and Applications 81. 5(2022): 6259–6276.
7. Akhtar, Zahid. "Deepfakes generation and detection: A short survey". Journal of Imaging 9. 1(2023): 18.
8. Rana, Md Shohel, Mohammad Nur, Nobi, Beddhu, Murali, Andrew H, Sung. "Deepfake detection: A systematic literature review". IEEE access 10. (2022): 25494–25513.
9. Masood, Momina, Mariam, Nawaz, Khalid Mahmood, Malik, Ali, Javed, Aun, Irtaza, Hafiz, Malik. "Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward". Applied intelligence 53. 4(2023): 3974–4026.
10. Stroebel, Laura, Mark, Llewellyn, Tricia, Hartley, Tsui Shan, Ip, Mohiuddin, Ahmed. "A systematic literature review on the effectiveness of deepfake detection techniques". Journal of Cyber Security Technology 7. 2(2023): 83–113.
11. Gong, Liang Yu, Xue Jun, Li. "A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges". Electronics 13. 3(2024): 585.
12. Wang, Zhikan, Zhongyao, Cheng, Jiajie, Xiong, Xun, Xu, Tianrui, Li, Bharadwaj, Veeravalli, Xulei, Yang. "A Timely Survey on Vision Transformer for Deepfake Detection". arXiv preprint arXiv:2405.08463. (2024).
13. "140k Real and Fake Faces — Kaggle.https://www.kaggle.com/xhlulu/140k-real-and-fake-faces