

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 Mid-journey, 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 immediately[3].

2 Literature Review

[?] has established a new TD(0) method that replays all past experiences. On the hand, [?] has taken this further to include a target that incorporates all past updates via TD(λ). [?] applied conjugate gradient update on TD.

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 $\|x\| \rightarrow \infty$.

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 \bar{x} , which is found by minimizing the dual action functional:

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

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

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

is a convex function, and

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

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

$$\delta := \liminf_{x \rightarrow 0} 2N(x) \|x\|^{-2}. \quad (7)$$

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

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

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

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

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

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

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:

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

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

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

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 $\bar{u} \neq 0$ and $\bar{u} \neq \Lambda_o^{-1}(0) = 0$. \square

Corollary 1. *Assume H is C^2 and (a_∞, b_∞) -subquadratic at infinity. Let ξ_1, \dots, ξ_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:*

$$\omega := \text{Min } \{\omega_1, \dots, \omega_k\} . \quad (13)$$

If:

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

then minimization of ψ yields a non-constant T -periodic solution \bar{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 \bar{x} exists and is non-constant provided that:

$$\frac{T}{2\pi} b_\infty < 1 < \frac{T}{2\pi} \quad (15)$$

or

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

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

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

Hence:

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

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

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

which is precisely condition (14). \square

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

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

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

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

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

$$i_T(\tilde{x}) = 0 . \quad (21)$$

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

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

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

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 , $T \in (2\pi\omega^{-1}, 2\pi b_\infty^{-1})$ of periodic solutions, $x_T(0) = x_T(T)$, with x_T going away to infinity when $T \rightarrow 2\pi\omega^{-1}$, which is the period of the linearized system at 0.

Table 1. This is the example table taken out of *The T_EXbook*, 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*

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

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

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

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

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

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

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

$$p \geq p_o \Rightarrow x_{pk} \neq x_k . \quad (28)$$

□

Example 1 (External forcing). Consider the system:

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

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

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

where $f_o := T^{-1} \int_0^T f(t)dt$. For instance,

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

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

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

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

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

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

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

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

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. Patel, Yogesh, et al. "Deepfake Generation and Detection: Case Study and Challenges." IEEE Access (2023).
4. Oussidi, Achraf, Azeddine, Elhassouny. "Deep generative models: Survey." 2018 International conference on intelligent systems and computer vision (ISCV). 2018.
5. Mehmood, R, R, Bashir, KJ, Giri. "Deep generative models: a review". Indian Journal of Science and Technology 16. 7(2023): 460–467.
6. Bartos, Gaye Ediboglu, Serel, Akyol. "Deep Learning for Image Authentication: A Comparative Study on Real and AI-Generated Image Classification".
7. Wodajo, Deressa, Solomon, Atnafu. "Deepfake video detection using convolutional vision transformer". arXiv preprint arXiv:2102.11126. (2021).
8. Hossain, Md Zahid, Farhad Uz, Zaman, Md Rakibul, Islam. "Advancing AI-Generated Image Detection: Enhanced Accuracy through CNN and Vision Transformer Models with Explainable AI Insights." 2023 26th International Conference on Computer and Information Technology (ICCIT). 2023.
9. Zhao, Hanqing, Wenbo, Zhou, Dongdong, Chen, Tianyi, Wei, Weiming, Zhang, Nenghai, Yu. "Multi-attentional deepfake detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
10. Sharma, Jatin, Sahil, Sharma, Vijay, Kumar, Hany S, Hussein, Hammam, Alshazly. "Deepfakes Classification of Faces Using Convolutional Neural Networks.". Traitement du Signal 39. 3(2022).
11. Sharma, Jatin, Sahil, Sharma. "Challenges and solutions in deepfakes". arXiv preprint arXiv:2109.05397. (2021).
12. Atwan, Jaffar, Mohammad, Wedyan, Dheeb, Albashish, Elaf, Aljaafrah, Ryan, Al-turki, Bandar, Alshawi. "Using Deep Learning to Recognize Fake Faces.". International Journal of Advanced Computer Science and Applications 15. 1(2024).
13. Eissa, Khaled, Friedhelm, Schwenker. "Fake vs. Real Face Discrimination Using Convolutional Neural Networks." Pan African Conference on Artificial Intelligence. 2023.
14. Bagherzadeh, Fahimeh, Razieh, Rastgoo. "Deepfake image detection using a deep hybrid convolutional neural network". Journal of Modeling in Engineering 21. 75(2023): 19–28.

15. Lanzino, Romeo, Federico, Fontana, Anxhelo, Diko, Marco Raoul, Marini, Luigi, Cinque. "Faster Than Lies: Real-time Deepfake Detection using Binary Neural Networks." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.
16. Gupta, Achal Shankar, Krishan Pratap, Shreneter, Smriti, Sehgal. "Visual Veracity: Advancing AI-Generated Image Detection with Convolutional Neural Networks." *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. 2024.
17. Dolhansky, Brian, Joanna, Bitton, Ben, Pflaum, Jikuo, Lu, Russ, Howes, Menglin, Wang, Cristian Canton, Ferrer. "The deepfake detection challenge (dfdc) dataset". *arXiv preprint arXiv:2006.07397*. (2020).
18. 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.
19. Matern, Falko, Christian, Riess, Marc, Stamminger. "Exploiting visual artifacts to expose deepfakes and face manipulations." *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*. 2019.
20. Agarwal, Shruti, Hany, Farid, Yuming, Gu, Mingming, He, Koki, Nagano, Hao, Li. "Protecting World Leaders Against Deep Fakes.." *CVPR workshops*. 2019.
21. Ciftci, Umur Aybars, Ilke, Demir, Lijun, Yin. "Fakecatcher: Detection of synthetic portrait videos using biological signals". *IEEE transactions on pattern analysis and machine intelligence*. (2020).
22. Yang, Xin, Yuezun, Li, Siwei, Lyu. "Exposing deep fakes using inconsistent head poses." *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019.
23. McCloskey, Scott, Michael, Albright. "Detecting GAN-generated imagery using saturation cues." *2019 IEEE international conference on image processing (ICIP)*. 2019.
24. Guarnera, Luca, Oliver, Giudice, Sebastiano, Battiato. "Deepfake detection by analyzing convolutional traces." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020.
25. Nguyen, Huy H, Fuming, Fang, Junichi, Yamagishi, Isao, Echizen. "Multi-task learning for detecting and segmenting manipulated facial images and videos." *2019 IEEE 10th international conference on biometrics theory, applications and systems (BTAS)*. 2019.
26. Khalil, Samar Samir, Sherin M, Youssef, Sherine Nagy, Saleh. "iCaps-Dfake: An integrated capsule-based model for deepfake image and video detection". *Future Internet* 13. 4(2021): 93.
27. Afchar, Darius, Vincent, Nozick, Junichi, Yamagishi, Isao, Echizen. "Mesonet: a compact facial video forgery detection network." *2018 IEEE international workshop on information forensics and security (WIFS)*. 2018.
28. Wang, Yaohui, Antitza, Dantcheva. "A video is worth more than 1000 lies. Comparing 3DCNN approaches for detecting deepfakes." *2020 15Th IEEE international conference on automatic face and gesture recognition (FG 2020)*. 2020.
29. Rafique, Rimsha, Rahma, Gantassi, Rashid, Amin, Jaroslav, Frnda, Aida, Mustapha, Asma Hassan, Alshehri. "Deep fake detection and classification using error-level analysis and deep learning". *Scientific Reports* 13. 1(2023): 7422.
30. Dosovitskiy, Alexey, Lucas, Beyer, Alexander, Kolesnikov, Dirk, Weissenborn, Xi-aohua, Zhai, Thomas, Unterthiner, Mostafa, Dehghani, Matthias, Minderer, Georg,

- Heigold, Sylvain, Gelly, others. "An image is worth 16x16 words: Transformers for image recognition at scale". arXiv preprint arXiv:2010.11929. (2020).
31. Percival, Donald B, Andrew T, Walden. Wavelet methods for time series analysis. Vol. 4. Cambridge university press, 2000.
 32. Ryan, Øyvind, Geir Dahl, and Knut Mørken. "Nonlinear optimization Lecture notes for the course MAT-INF2360." University of Oslo (2014).
 33. Marcellin, Michael W, Michael J, Gormish, Ali, Bilgin, Martin P, Boliek. "An overview of JPEG-2000." Proceedings DCC 2000. Data compression conference. 2000.
 34. Nadler, Ethan O, Elise, Darragh-Ford, Bhargav Srinivasa, Desikan, Christian, Conaway, Mark, Chu, Tasker, Hull, Douglas, Guilbeault. "Divergences in color perception between deep neural networks and humans". Cognition 241. (2023): 105621.
 35. Li, Qiufu, Linlin, Shen, Sheng, Guo, Zhihui, Lai. "WaveCNet: Wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification". IEEE Transactions on Image Processing 30. (2021): 7074–7089.
 36. Cody, Mac A. "The wavelet packet transform: Extending the wavelet transform". Dr. Dobb's Journal 19. (1994): 44–46.
 37. Luo, Xiangyang, Fenlin, Liu, Jianming, Chen, Yining, Zhang. "Image universal steganalysis based on wavelet packet transform." 2008 IEEE 10th Workshop on Multimedia Signal Processing. 2008.
 38. Davidson, Jennifer, Clifford, Bergman, Eric, Bartlett. "An artificial neural network for wavelet steganalysis." Mathematical Methods in Pattern and Image Analysis. 2005.
 39. Wolter, Moritz, Felix, Blanke, Raoul, Heese, Jochen, Garcke. "Wavelet-packets for deepfake image analysis and detection". Machine Learning 111. 11(2022): 4295–4327.
 40. Rossler, Andreas, Davide, Cozzolino, Luisa, Verdoliva, Christian, Riess, Justus, Thies, Matthias, Nießner. "Faceforensics++: Learning to detect manipulated facial images." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
 41. Bird, Jordan J, Ahmad, Lotfi. "Cifake: Image classification and explainable identification of ai-generated synthetic images". IEEE Access. (2024).
 42. Chen, Zhongxi, Ke, Sun, Ziyin, Zhou, Xianming, Lin, Xiaoshuai, Sun, Liujuan, Cao, Rongrong, Ji. "DiffusionFace: Towards a Comprehensive Dataset for Diffusion-Based Face Forgery Analysis". arXiv preprint arXiv:2403.18471. (2024).
 43. "140k Real and Fake Faces — Kaggle.<https://www.kaggle.com/xhlulu/140k-real-and-fake-faces>