

Progressive Fourier Neural Representation for Sequential Video Compilation

Haeyong Kang, Jaehong Yoon, DaHyun Kim, Sungju Hwang, and Chang D. Yoo

Progressive Fourier Neural Representation for Sequential Video Compliation



Progressive Fourier Neural Representation for Sequential Video Compilation 🦥 ICLR

Motivation:

(1) It has not been explored sequential video so far in Continual Learning.

(2) There is also no robust continual baseline in Video Continual Learning.



➔ We propose a new method to show the effectiveness of reused winning tickets (WSN, ICML2022) in Video Continual Representations.

Progressive Fourier Neural Representation for Sequential Video Compilation



- Backgrounds of Neural Implicit Rep.



- The world around us is not discrete,
- Yet, we choose to represent real-world signals such as images or sound in a discrete manner.

(-) Discrete only contain a discrete amount of information regarding the signal.

(-) Given a 256x256 pixel grid for an image, we are not able to scale it up to a 512x512 image.

 \leftarrow Not enough information

- Neural Implicit Representations (NIR) are <u>neural</u> networks (e.g. MLPs) that estimate the function **f** that represents a signal continuously, by training on discretely represented samples of the same signal.
- **NIP** learns how to estimate the underlying (continuous) function *f* (denoted *F* below):

 $F(x, \Phi, \nabla_x \Phi, \nabla_x^2 \Phi, \cdots) = 0, \ \Phi : x \mapsto \Phi(x)$

- The network parameterizes Φ . After training on the discretely represented samples, the estimated **f** would be *implicitly* encoded in the network, hence the name "**Neural Implicit Representation**".



Discrete representations of various signals (SIREN: Sinusoidal Representation Networks, NeuralPS2020)

Progressive Fourier Neural Representation for Sequential Video Compilation

v

v



Coordinate-based Methods



SIREN: Sinusoidal Representation Networks, NeuralPS2020

(+) map coordinates to images.



(+) to learn representation with high frequencies

512 256 128 131 128 →□ point coordinates Concatenate Copy and Concatenate FC, Leaky ReLU FC, Sigmoid AAAAAA

1024

2048

A value indicating the inside/outside



(+) Feature + coordinate led to cleaner interpolation results.



Index-based Methods (1)







NeRV, NeuralPS2021

(+) NeRV represents videos or image datasets as neural neworks, <u>taking an image index as input</u>, and <u>outputting the whole image</u>.

(+) Faster than coordinate-based methods.

E-NeRV, ECCV2022

(+) The size of the parameters were reduced **<u>by introducing</u> <u>disentangled spatial-temporal representations</u> with a light network, while maintaining the majority of performance.**

(+) They distributed the saved parameters to increase channel dimensions in convolution blocks, resulting in an E-NeRV model with similar or fewer parameters but much better performance.

5

Progressive Fourier Neural Representation for Sequential Video Compilation

Backgrounds of Neural Implicit Representation (NIR)



Index-based Methods (2)



D-NeRV, CVPR2023

(+) NeRV optimizes representation to every video independently while **D-NeRV encodes all videos by a shared model**.



NIRVANA, CVPR2023

6

(+) <u>NIRVANA performs spatio-temporal patch-wise prediction</u> and <u>fits</u> <u>individual neural networks to groups of frames (clips)</u> which are initialized using networks trained on the previous group.

(+) Such an autoregressive patch-wise approach exploits both spatial and temporal redundancies present in videos while promoting scalability and adaptability to varying video content, resolution or duration.



Progressive Fourier Neural Representation for Sequential Video Compilation 🦥 ICLR

Index-based sequential Neural Implicit Representation (sequential NIR)

WSN, ICML2022



(+) Providing forget-free continual learning solutions.
(+) Reused Subnetworks Adaptably tailer to tasks.
(+) Fast Convergence .



(+) Providing forget-free video continual learning.(+) Fast Convergence.

(-) WSN does not have enough parameters for complex video representations.

→ Needs more parameters to deal with the issue.

Progressive Fourier Neural Representation for Sequential Video Compilation



Index-based sequential Neural Implicit Representation (sequential NIR)

(+) Fourier Subneural Operator (FSO): Enough Reusable parameters in Frequency domain



Progressive Fourier Neural Representation for Sequential Video Compilation

Index-based sequential Neural Implicit Representation (sequential NIR)

(+) Fourier Subneural Operator (FSO): Enough Reusable parameters in Frequency domain



Progressive Fourier Neural Representation for Sequential Video Compilation

KAIST

PFNR's Statistics of Variance and Frequency



Progressive Fourier Neural Representation for Sequential Video Compilation

KAIST

10

ICLR

PFNR's Statistics of Variance and Frequency





Progressive Fourier Neural Representation for Sequential Video Compilation





Algorithm 1 Progressive Fourier Neural Representation (PFNR) for VCL input: $\{\mathcal{D}_s\}_{s=1}^N$, model weights of FSO $\theta_* = \{\theta, \phi_{FSO}\}$, score weights of FSO $\rho_* = \{\rho, \rho_{FSO}\}$, binary mask $\mathbf{M}_0 = \{\mathbf{0}^{|\boldsymbol{\theta}|}, \mathbf{0}^{|\boldsymbol{\theta}_{FSO}|}\}$, and layer-wise capacity c%. 1: randomly initialize θ_* and ρ_* . 2: for session $s = 1, \dots, |\mathcal{S}|$ do 3: if s > 1 then randomly re-initialize ρ_* . 4: 5: end if for batch $\mathbf{b}_t \sim \mathcal{D}_s$ do 6: obtain mask \mathbf{m}_s of the top-c% scores $\boldsymbol{\rho}_*$ at each layer 7: compute $\mathcal{L}(f(\boldsymbol{e}_{s,t}; \boldsymbol{\theta}_* \odot \mathbf{m}_s), \mathbf{b}_t)$, where input embedding, $\boldsymbol{e}_{s,t} = [\boldsymbol{e}_s; \boldsymbol{e}_t]$. 8: $\boldsymbol{\theta}_{*} \leftarrow \boldsymbol{\theta}_{*} - \eta \left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_{*}} \odot \left(\mathbf{1} - \mathbf{M}_{s-1} \right) \right)$ 9: \triangleright trainable weight update $\boldsymbol{\rho}_* \leftarrow \boldsymbol{\rho}_* - \eta(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\rho}_*})$ 10: \triangleright weight score update 11: end for $\hat{\boldsymbol{ heta}}_s = \boldsymbol{ heta}_* \odot \boldsymbol{m}_s$ 12: \triangleright accumulate binary mask 13: $\mathbf{M}_{s} \leftarrow \mathbf{M}_{s-1} \lor \mathbf{m}_{s}$ 14: end for output: $\{\hat{\boldsymbol{\theta}}_s\}_{s=1}^N$







(+) Quantization of PFNR & Bits per pixel (BPP)



(+) PFNR's Performances and Progressive Model Capacity

Progressive Fourier Neural Representation for Sequential Video Compilation





S1	31.50	7.63	8.81	10.02	8.70	7.29	9.83	9.54	9.76	8.96	8.50	10.21	10.04	9.32	7.36	7.49	7.89	S1	33.63	7.57	8.75	9.97	8.70	7.28	9.77	9.49	9.72	8.88	8.48	10.13	9.98	9.29	7.36	7.50	7.83
	31.50	34.3	7 12.44	9.45	5.90	7.02	9.59	9.63	8.91	11.42	7.29	9.59	9.84	7.30	5.27	4.48	10.41		33.63	39.24	12.44	9.45	5.90	7.02	9.60	9.63	8.91	11.42	7.29	9.59	9.84	7.30	5.27	4.48	10.41
	31.50	34.3	7 31.00	11.09	5.92	6.91	10.50	11.14	9.92	14.93	7.36	11.30	11.56	7.48	5.28	4.73	12.22		33.63	39.24	4 34.21	11.08	5.92	6.91	10.49	11.13	9.91	14.91	7.36	11.29	11.54	7.48	5.28	4.73	12.21
	31.50	34.3	7 31.00	32.38	7.97	8.61	10.55	11.59	10.72	11.23	9.24	10.52	11.62	8.79	7.29	6.60	9.50		33.63	39.24	4 34.21	37.79	7.96	8.59	10.53	11.55	10.69	11.19	9.23	10.49	11.58	8.78	7.29	6.60	9.48
S5	31.50	34.3	7 31.00	32.38	29.26	8.42	8.36	6.72	8.58	5.60	9.32	7.11	7.05	10.79	12.35	10.87	5.32	S5	33.63	39.24	4 34.21	37.79	34.05	8.41	8.33	6.70	8.56	5.58	9.30	7.08	7.02	10.75	12.31	10.85	5.30
	31.50	34.3	7 31.00	32.38	29.26	23.08	8.18	7.79	9.29	7.02	9.62	7.06	7.45	8.84	7.51	6.03	7.13		33.63	39.24	4 34.21	37.79	34.05	27.17	8.08	7.69	9.18	6.90	9.56	6.97	7.35	8.83	7.53	6.07	6.99
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	10.04	11.24	11.03	8.75	11.55	11.58	8.88	7.37	5.94	10.56		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	10.02	11.21	11.00	8.74	11.52	11.55	8.86	7.36	5.93	10.54
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	10.49	11.44	8.11	11.02	11.82	7.77	6.03	5.33	10.31		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	10.43	11.30	8.10	10.95	11.73	7.78	6.05	5.37	10.19
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	10.71	9.70	11.41	11.80	9.60	7.67	6.19	10.41		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	10.55	9.56	11.18	11.58	9.48	7.61	6.16	10.24
S10	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	33.48	7.44	11.62	12.08	7.37	4.97	4.27	14.75	S10	33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	36.18	7.44	11.61	12.07	7.37	4.97	4.27	14.71
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	33.48	18.34	7.61	7.73	9.22	8.84	7.23	7.30		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	36.18	22.97	7.50	7.63	9.08	8.74	7.21	7.10
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	33.48	18.34	20.45	14.64	7.91	6.32	5.28	11.90		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	36.18	22.97	24.36	14.14	7.83	6.31	5.30	11.44
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	33.48	18.34	20.45	27.21	8.11	6.26	5.33	11.51		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	36.18	22.97	24.36	32.50	8.09	6.26	5.33	11.42
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	33.48	18.34	20.45		24.33	8.41	8.40	6.95		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	36.18	22.97	24.36	32.50	30.22	8.39	8.36	6.93
S15	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	33.48	18.34	20.45		24.33	23.09	9.37	4.75	S15	33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	36.18	22.97	24.36	32.50	30.22	27.62	9.32	4.75
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	33.48	18.34	20.45		24.33	23.09	21.23	3.68		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	36.18	22.97	24.36	32.50		27.62	29.15	3.64
	31.50	34.3	7 31.00	32.38	29.26	23.08	31.96	22.64	22.07	33.48	18.34	20.45		24.33	23.09	21.23	29.13		33.63	39.24	4 34.21	37.79	34.05	27.17	38.17	29.79	26.56	36.18	22.97	24.36	32.50		27.62	29.15	35.68
	S1				S5					S10					S15				S1				S5					S10					S15		
					(;	a) V	WS	Ν,	<i>c</i> =	= 30	0.0	%								(b) PFNR, $c = 30.0\%$, <i>f</i> -NeRV3															

(+) PFNR's Forget-free Video Continual Learning

Progressive Fourier Neural Representation for Sequential Video Compilation

PFNR's Performances on the UVG17 datasets





WSN (29.26, PSNR)



PFNR, *f*-NeRV2 (31.24, PSNR)



PFNR, *f*-NeRV3 (34.05, PSNR)

(+) PFNR's Video Generation

Progressive Fourier Neural Representation for Sequential Video Compilation



PFNR's Performances on the UVG17 datasets





(+) 8-bit PFNR's Video Generation

Progressive Fourier Neural Representation for Sequential Video Compilation



- Neural Implicit Representation (NIR) has recently gained significant attention due to its remarkable ability to encode complex and high-dimensional data into representation space and easily reconstruct it through a trainable mapping function.
- However, NIR methods assume a one-to-one mapping between the target data and representation models regardless of data relevancy or similarity. This results in poor generalization over multiple complex data and limits their efficiency and scalability.
- Motivated by continual learning, this work investigates how to accumulate and transfer neural implicit representations for multiple complex video data over sequential encoding sessions. To overcome the limitation of NIR, we propose a novel method, <u>Progressive Fourier Neural Representation (PFNR), that aims to find an</u> <u>adaptive and compact sub-module in Fourier space</u> to encode videos in each training session.
- This sparsified neural encoding allows the neural network to hold free weights, enabling an improved adaptation for future videos. In addition, when learning a representation for a new video, PFNR transfers the representation of previous videos with frozen weights.
- This design allows the model to continuously accumulate high-quality neural representations for multiple videos while ensuring lossless decoding that perfectly preserves the learned representations for previous videos. We <u>validate our PFNR method on the UVG8/17 and DAVIS50 video sequence benchmarks</u> and achieve impressive performance gains over strong continual learning baselines.

