

Compass RoPE: Isotropic Rotary Position Embeddings for Vision Transformers

Compass-RoPE solves Mix-RoPE's directional collapse via polar frequency decoupling and intra-head Fourier mixing, stabilizing resolution extrapolation and boosting ImageNet-1K multi-resolution AVG Top-1 by +2.2.

Authors

Chengxi Min, Wei Wang, Yao Zhao
{25120327, Wei Wang}@bjtu.edu.cn



北京交通大学

Beijing Jiaotong University

Paper



Code

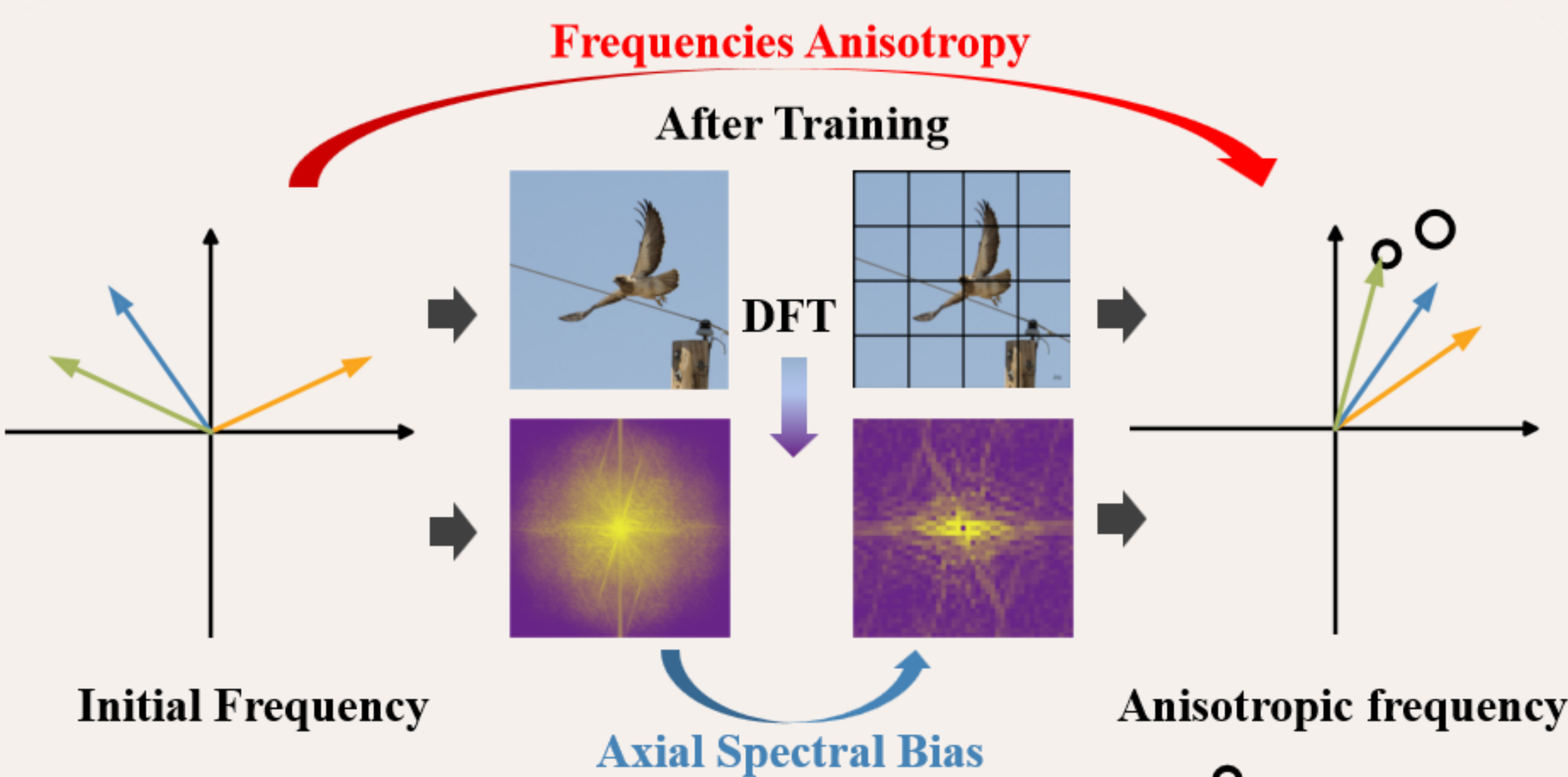


ICML
International Conference
On Machine Learning

Introduction

Motivation

Patch tokenization induces **axial spectral bias** in visual tokens, which enters attention computation through q-k interactions and drives RoPE frequencies toward **anisotropic** directions. This weakens spatial dependency modeling under resolution shifts, motivating an **isotropic RoPE** design with balanced frequency coverage.



$$2D\text{ RoPE} \begin{cases} q'(p) = q(p)e^{i\alpha(p)} \\ \alpha(p) = w_x p_x + w_y p_y \\ \varphi = \arctan(\frac{w_y}{w_x}) \end{cases}$$

Anisotropic frequencies(φ) ignore some directions under extrapolation. ❌

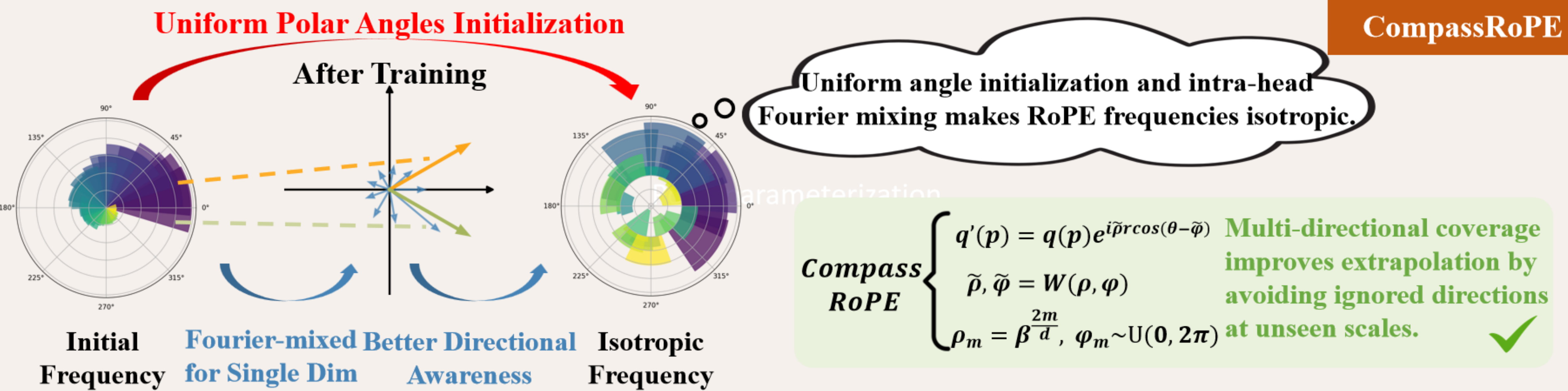
Contribution

- Method** - We propose Compass-RoPE to achieve isotropic frequency coverage via **uniform initialization** for frequency and **DFT mixing**.
- Effect** - Compass-RoPE improves **resolution extrapolation** and generalizes well to classification, detection, segmentation, and generation tasks.

Methods

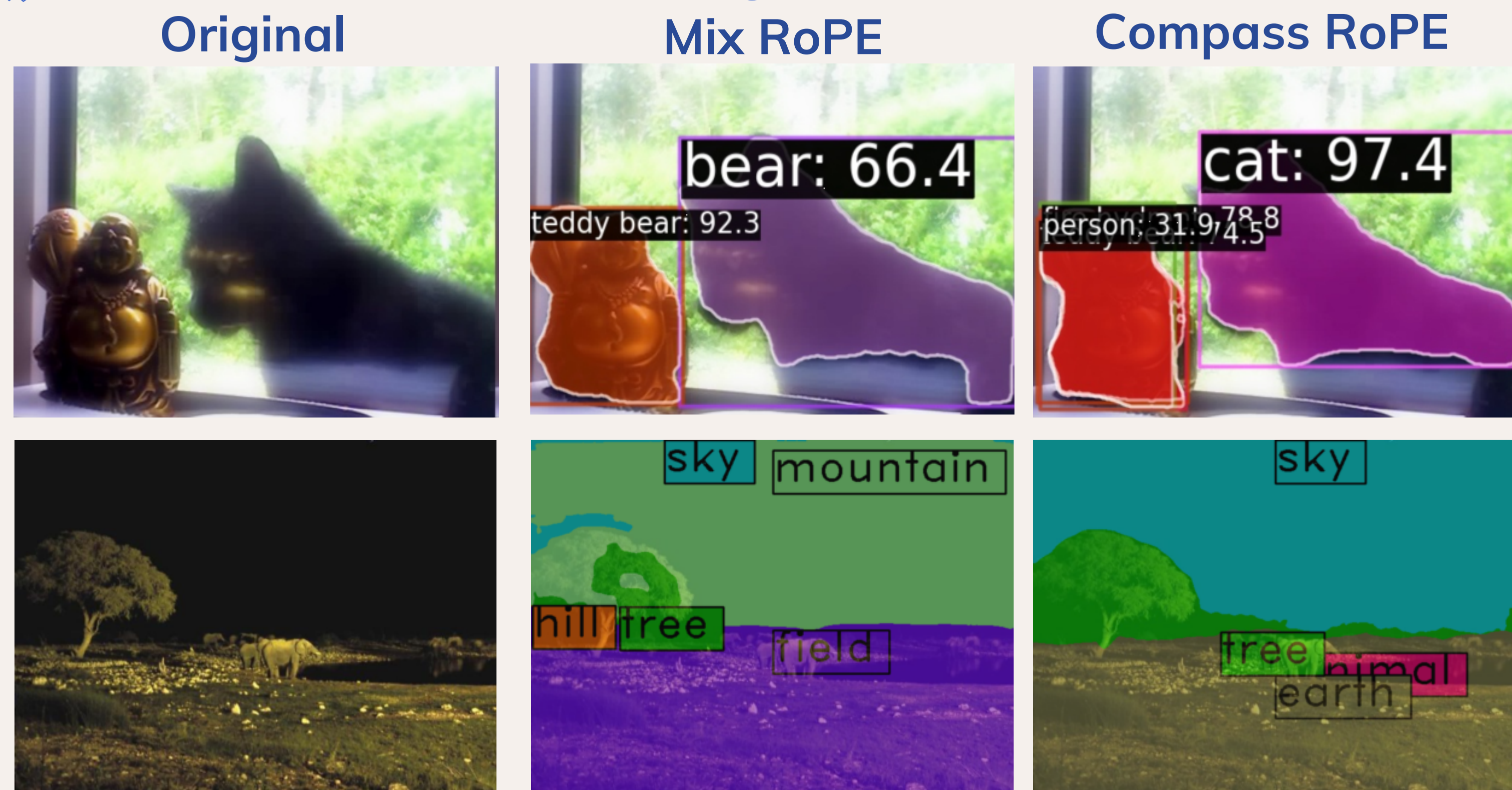
Compass RoPE

- Polar RoPE** - Parameterize 2D frequencies by scale and angle, with **uniformly initialized** angles for isotropic directional coverage.
- DFT Mixing** - Apply intra-head Fourier mixing to enrich **angular expressiveness**



Downstream Tasks Results

Detection and Segmentation

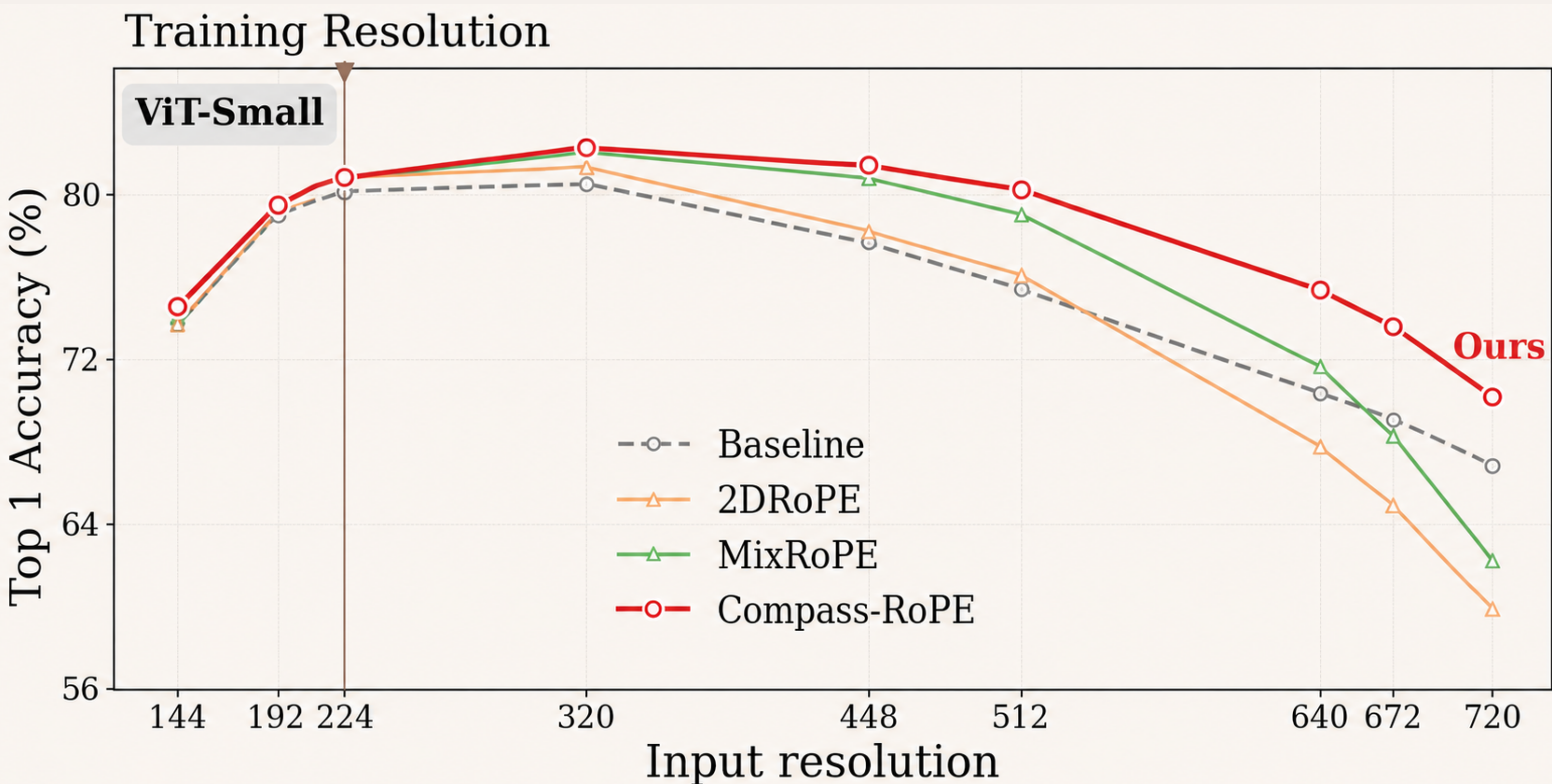


Compass-RoPE improves detection and segmentation with stronger spatial understanding.

Experiment Results

Classification on ImageNet-1K

Compass-RoPE achieves the best average accuracy across **all resolutions** and becomes much better at **high resolutions**.



Other Vision Tasks

Dense Prediction Tasks

Task	ViT (APE)	Mix-RoPE	Compass-RoPE
Pascal VOC Detection mAP (%)	78.3	79.1	79.5 ↑
ADE20K Segmentation mIoU (%)	41.1	43.2	43.4 ↑

Image Generation Task

Model	FID ↓	sFID ↓	IS ↑	Precision ↑
DiT	67.3	12.5	20.3	0.365
DiT + Compass-RoPE	61.2 ↓	11.5 ↓	22.6 ↑	0.393 ↑

Comparison With Other Vision RoPEs on ImageNet-100

