

LR0.FM: LOW-RESOLUTION ZERO-SHOT CLASSIFICATION BENCHMARK FOR FOUNDATION MODELS

<https://ucf-crcv.github.io/lr0.fm>



Priyank Pathak¹



Shyam Marjit²



Shruti Vyas¹



Yogesh S Rawat¹

¹UNIVERSITY OF CENTRAL FLORIDA

² IIIT GUWAHATI

{PRIYANK, SHRUTI, YOGESH}@UCF.EDU,

SHYAM.MARJIT@IIITG.AC.IN

Real World is low resolution

- Surveillance footage (CCTV)
- Satellite
- Pixelated

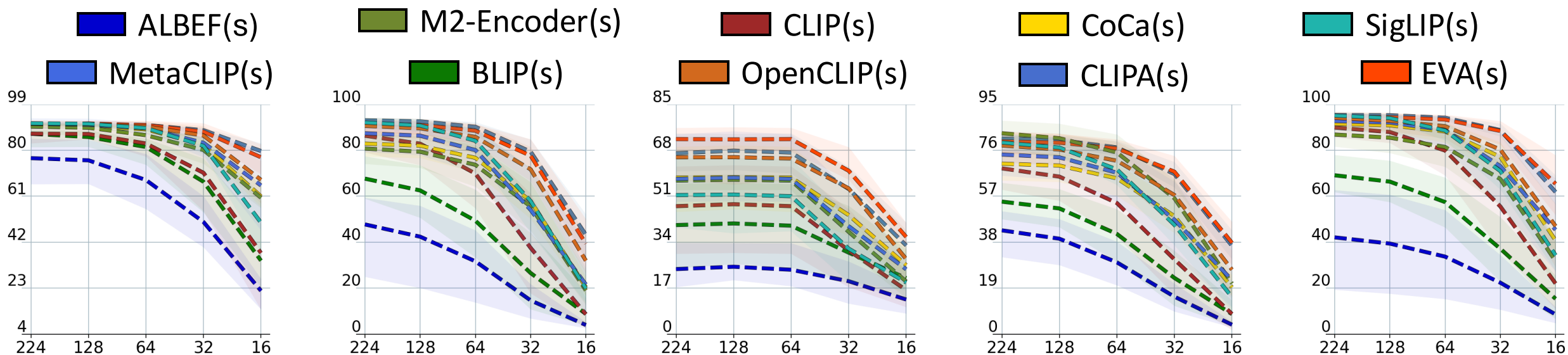


Moon
Maraca (toy)



Deer
Coyote

Performance Goes down with resolution



10 Foundation Models | 66 Backbones | 15 datasets

Benchmarking : What makes a model robust?

- Robustness is measured by **Relative Robustness**.

Models	Dataset	Random Prediction	Acc 224 x 224	Acc 16 x 16	Relative Robustness
ALBE36.9F (4M)	Aircraft	1%	2.7%	1.0%	37.1%
BLIP-ViT-B/16			4.8%	1.8%	36.9%

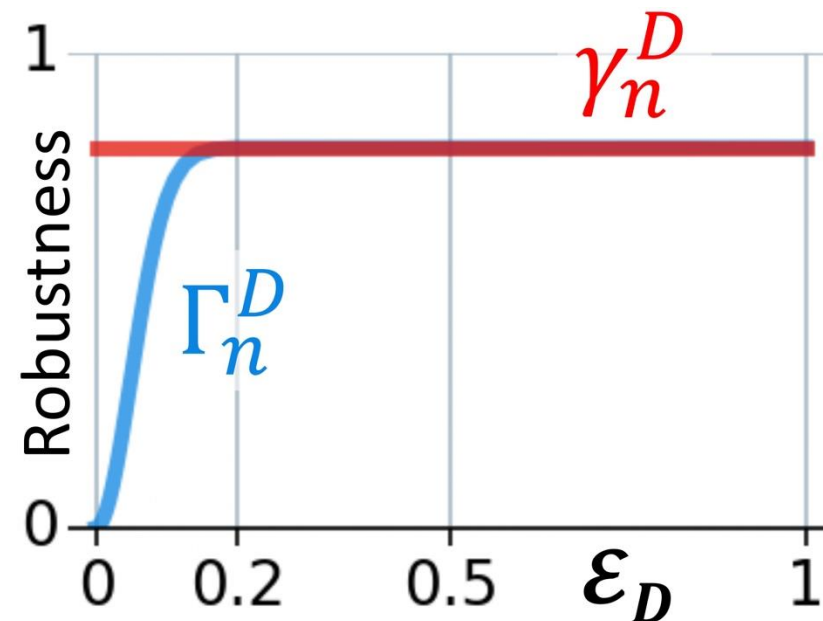
Improved relative robustness

Proposed improved robustness

Zero near random prediction

$$\Gamma_n^D = \gamma_b^d \times (1 - e^{\{-\alpha \epsilon^2\}})$$

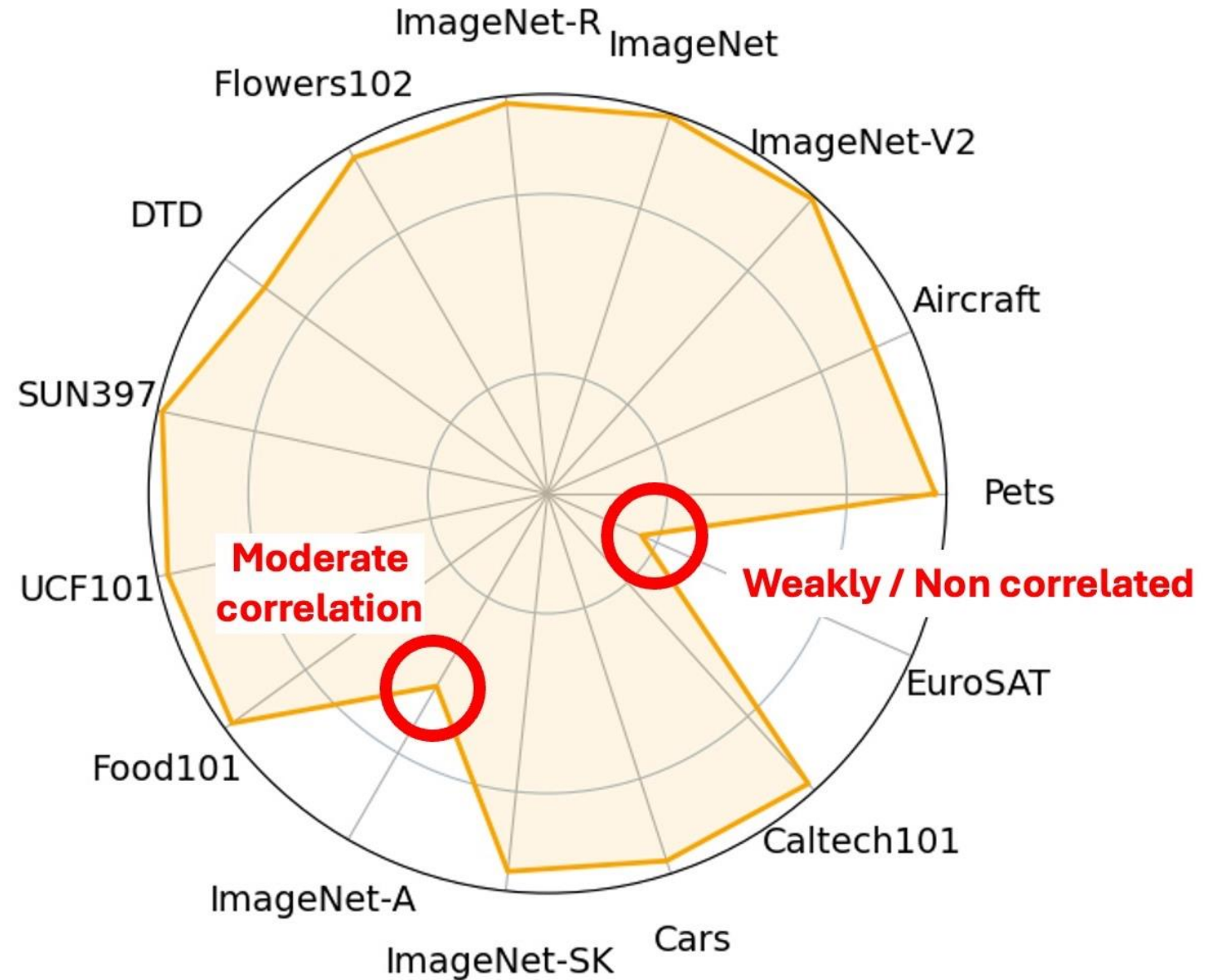
Traditional robustness



Models	Random Prediction	Acc 224 x 224	Acc 16 x 16	Relative Robustness	Improved Robustnes
ALBE36.9F (4M)	1%	2.7%	1.0%	37.1%	2.1%
BLIP-ViT-B/16		4.8%	1.8%	36.9%	9.3%

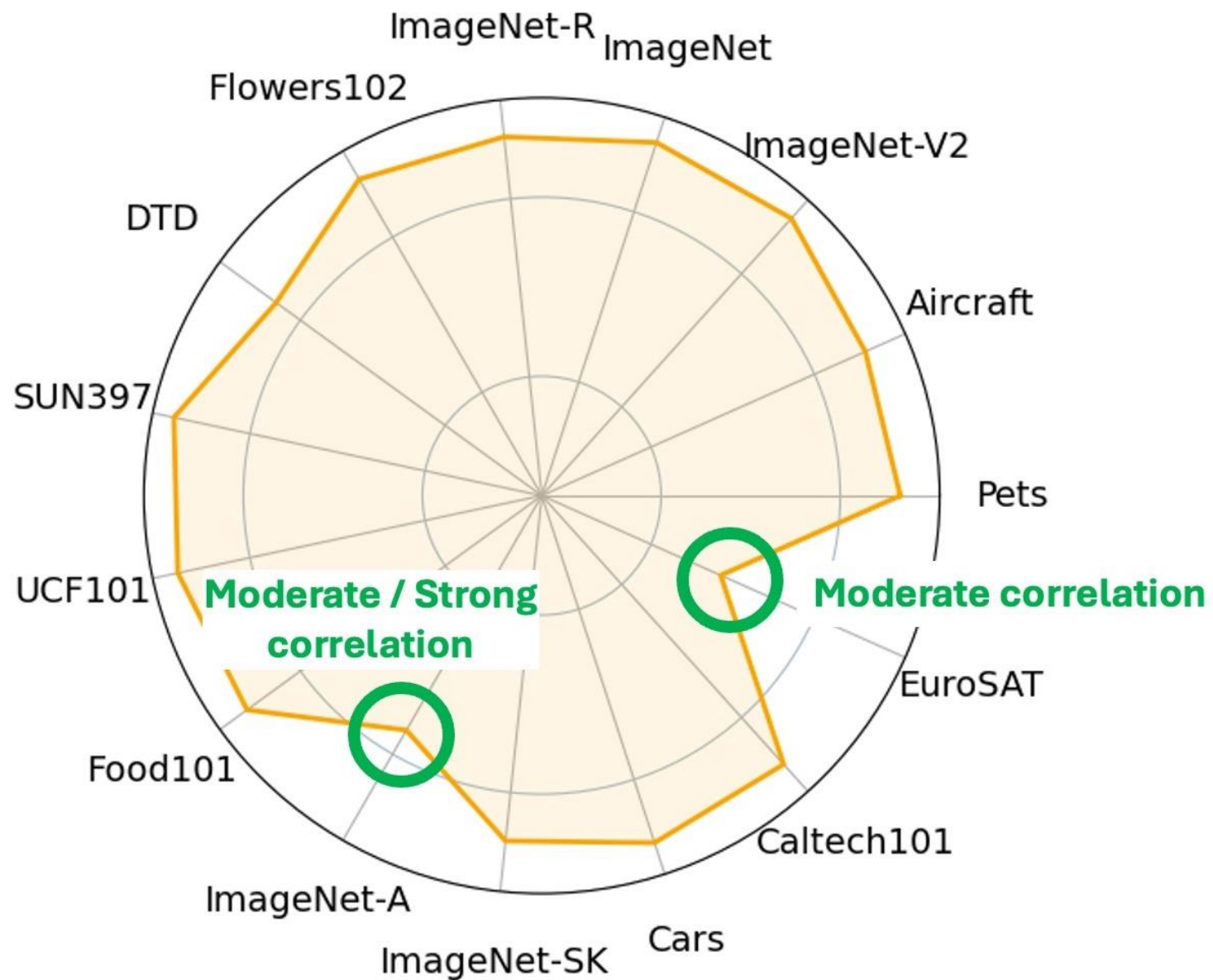
Uniform Weight for each dataset (traditional)

Average Robustness
across dataset
ignores dataset



Weight Aggregate Metric (WAR, Proposed)

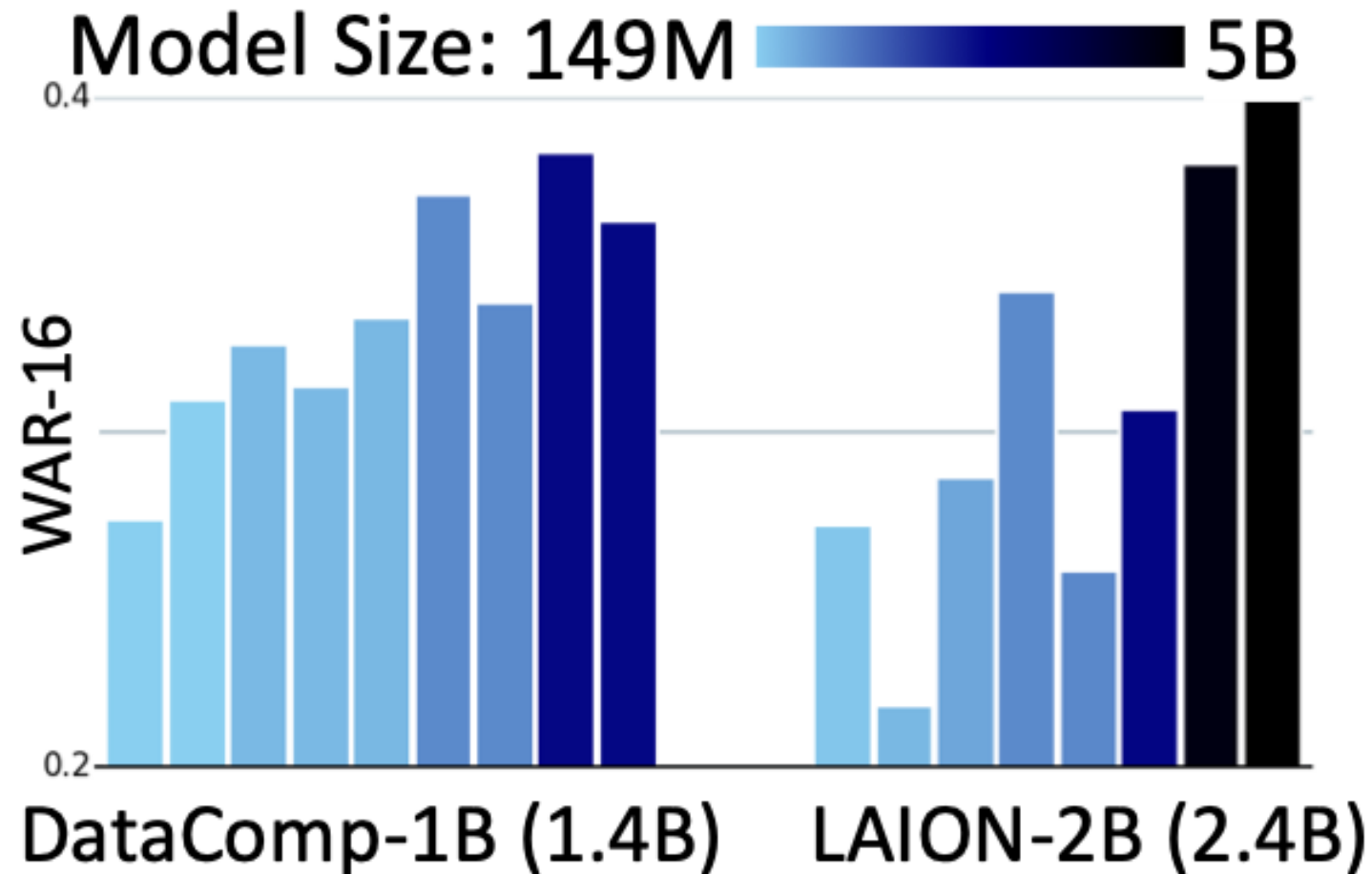
Weighted Average
represents each
dataset



Benchmark Analysis

Smaller Pretraining Dataset is more robust

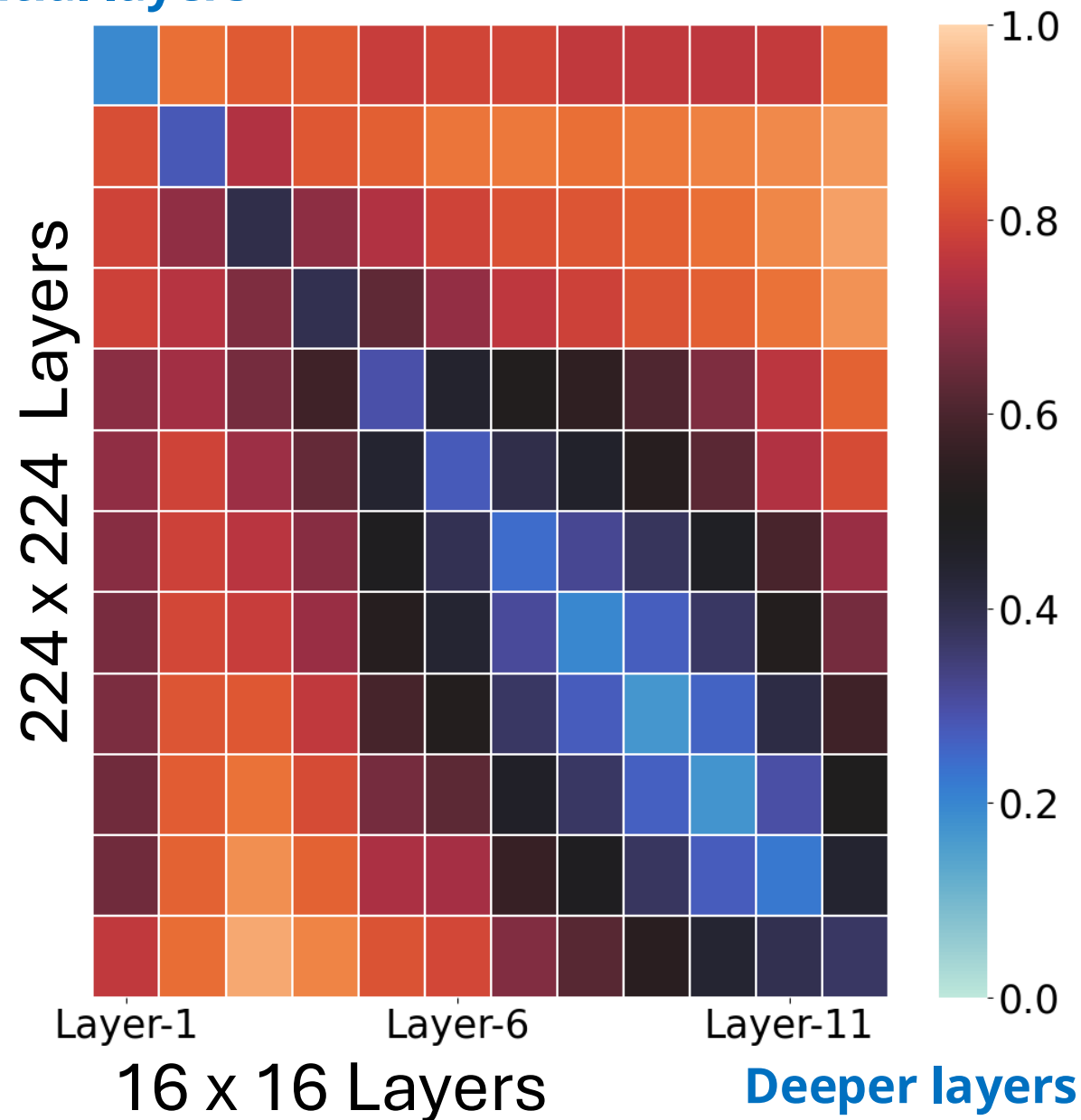
Bigger models are more robust



Benchmark Analysis

More analysis in the
paper

Initial layers



Model makes
correct
predictions in
Low
resolution

224 x 224



Baseball

128 x 128



Snail

224 x 224



Goldfinch

32 x 32



Pole

Preserve
Pretrained
weights

224 x 224



Volcano

64 x 64



Mountain

224 x 224



Hen

16 x 16



Rooster

LR-TK0

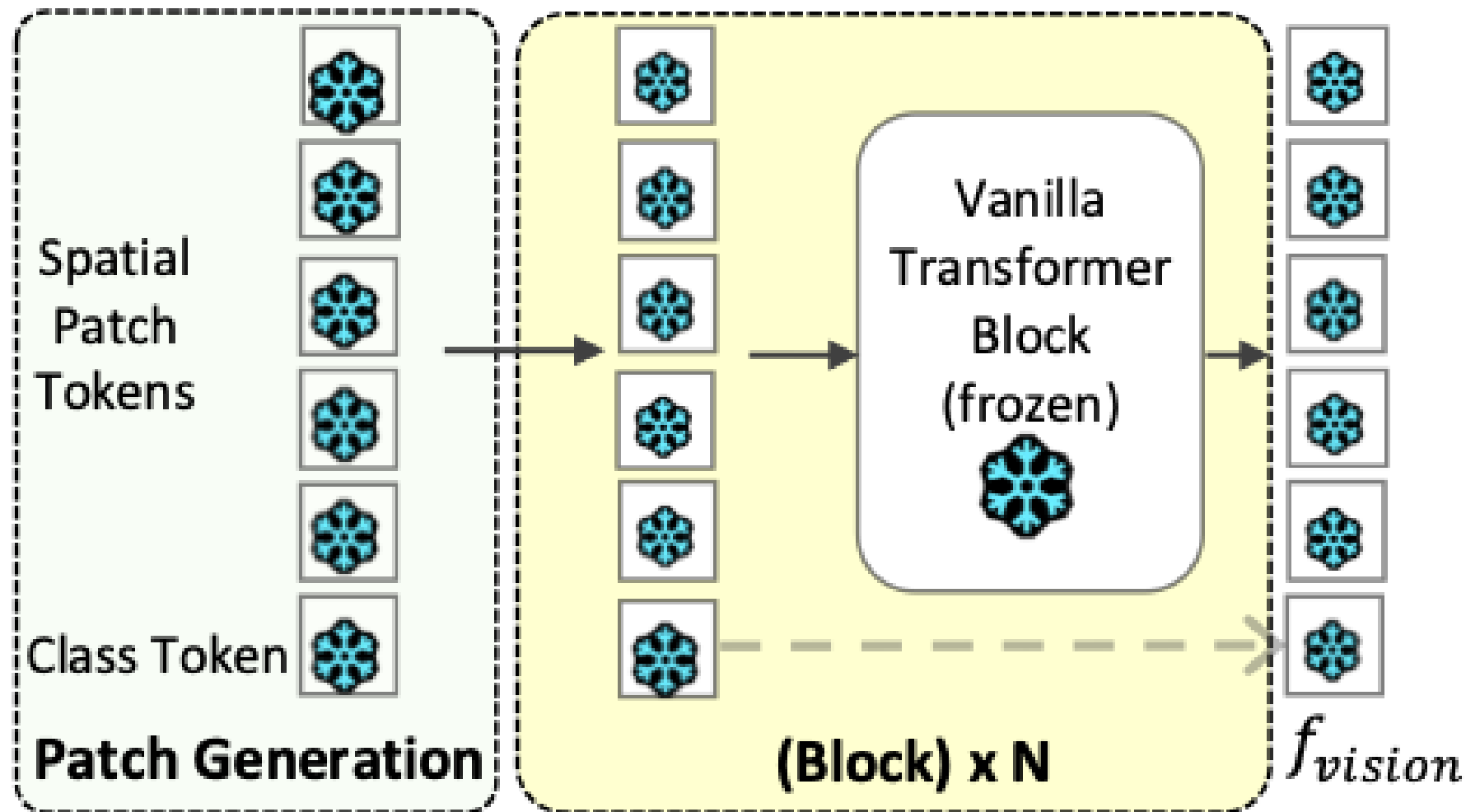
Frozen Parameters



Trainable LR Tokens



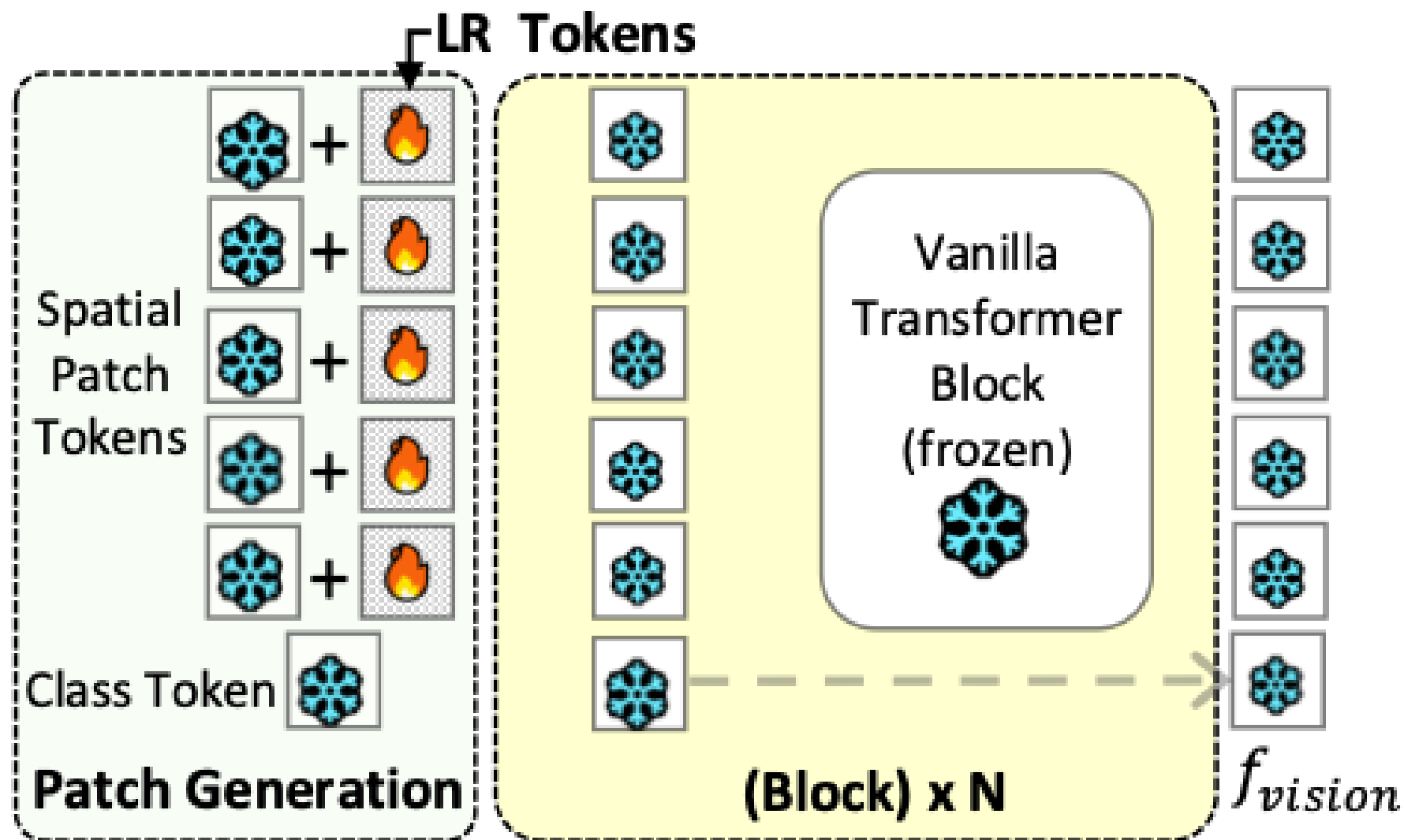
Add trainable
tokens at
every layer



LR-TK0

Add trainable tokens at every layer

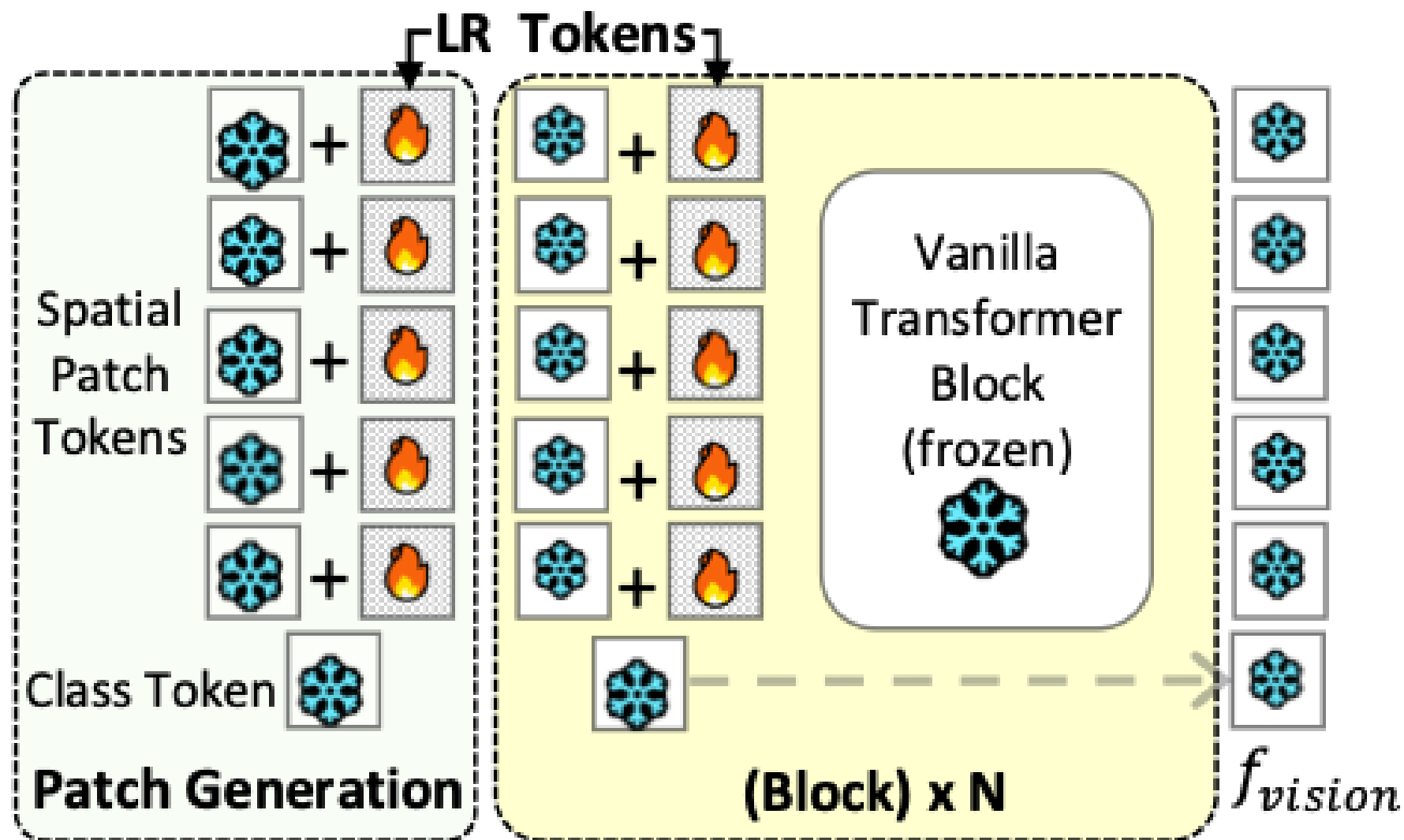
Frozen Parameters  Trainable LR Tokens 



LR-TK0

Add trainable tokens at every layer

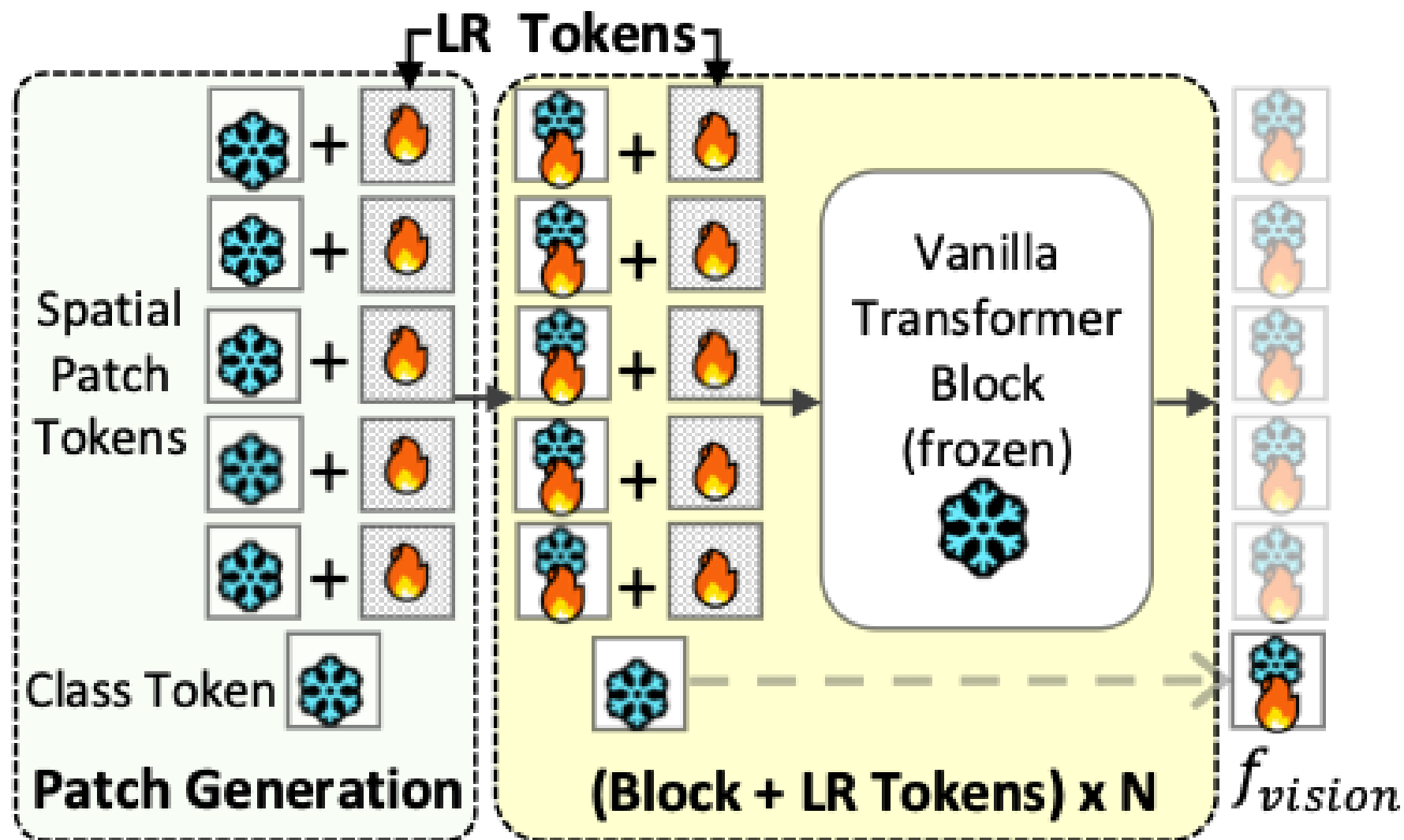
Frozen Parameters  Trainable LR Tokens 



LR-TK0

Add trainable tokens at every layer

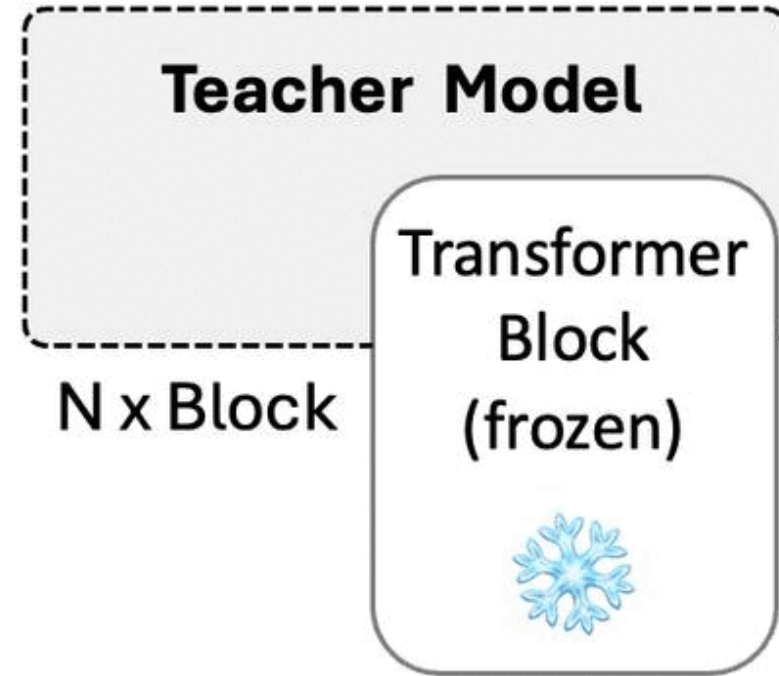
Frozen Parameters  Trainable LR Tokens 



Train via self-supervision

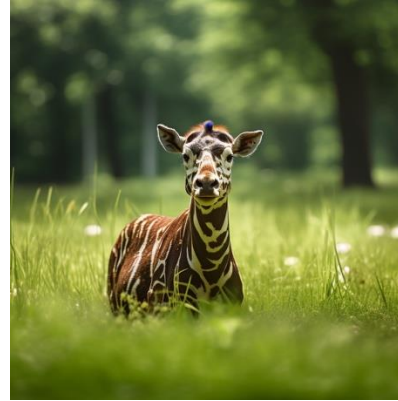
**No labels
required**

Synthetic Dataset



Train on Synthetic Dataset

- Diffusion generated image on random captions
- **Maintain zero-shot** nature



*animal in the grass
of the park*

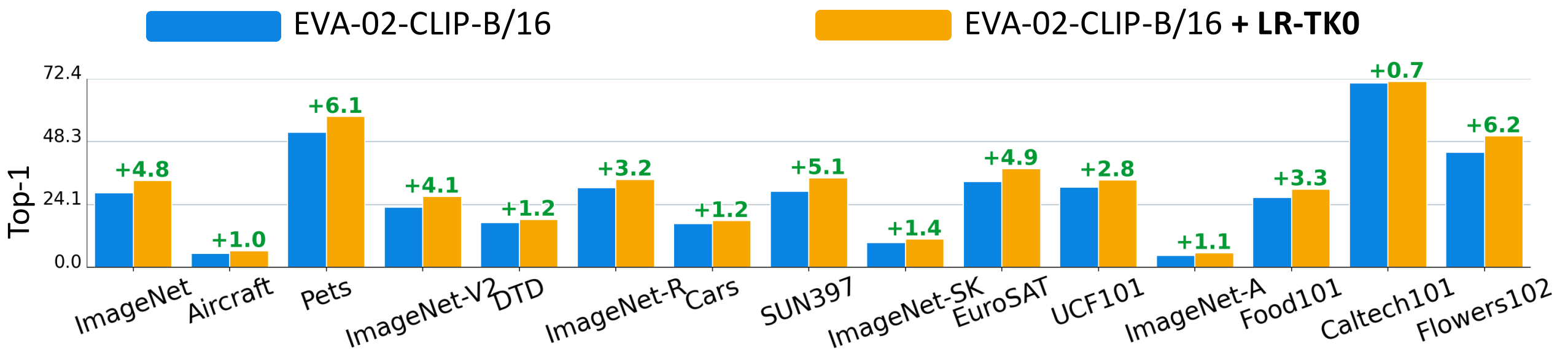


*logo with head of
a fox*



“wedding under the oak trees”

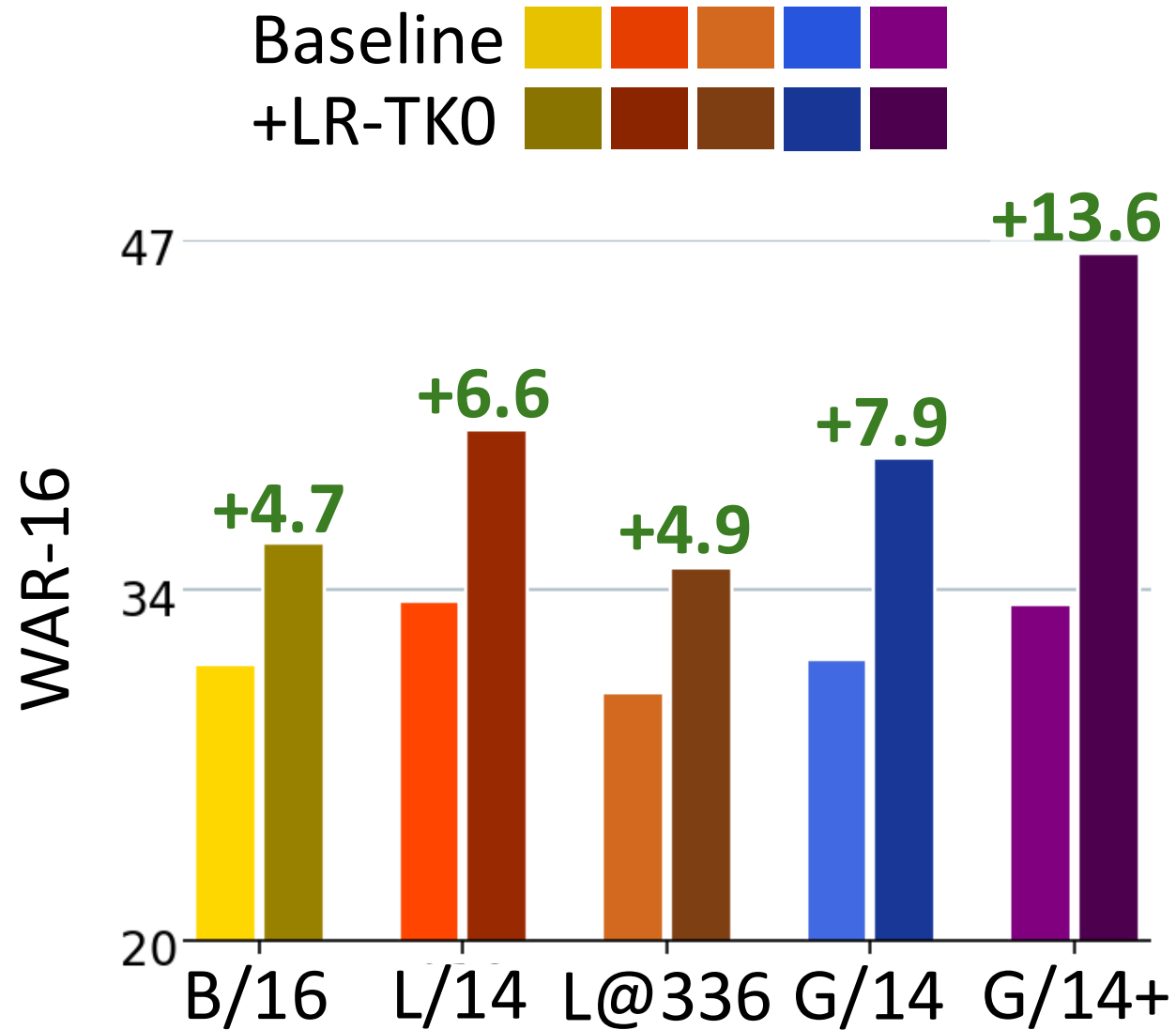
Results



Generalization to other models

Results

Bigger backbones sees larger improvement





Thank you

**Session 3 Poster Session
25th April 10 AM – 12:30 PM**