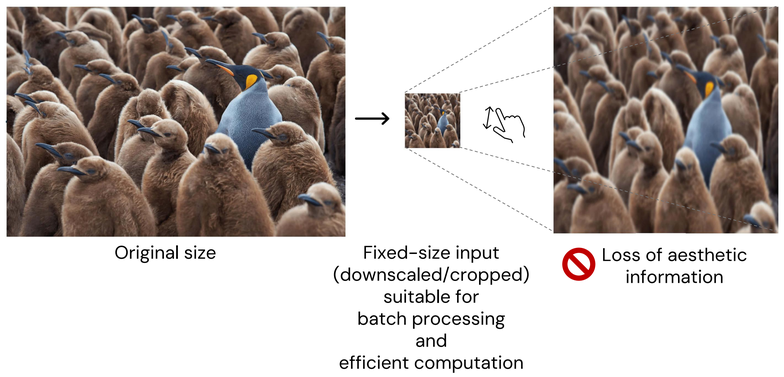
CHARM is essential for human perception and aesthetics, and can be implemented in computer vision, too.

Fatemeh Behrad, Tinne Tuytelaars, Johan Wagemans
KU Leuven University, Belgium

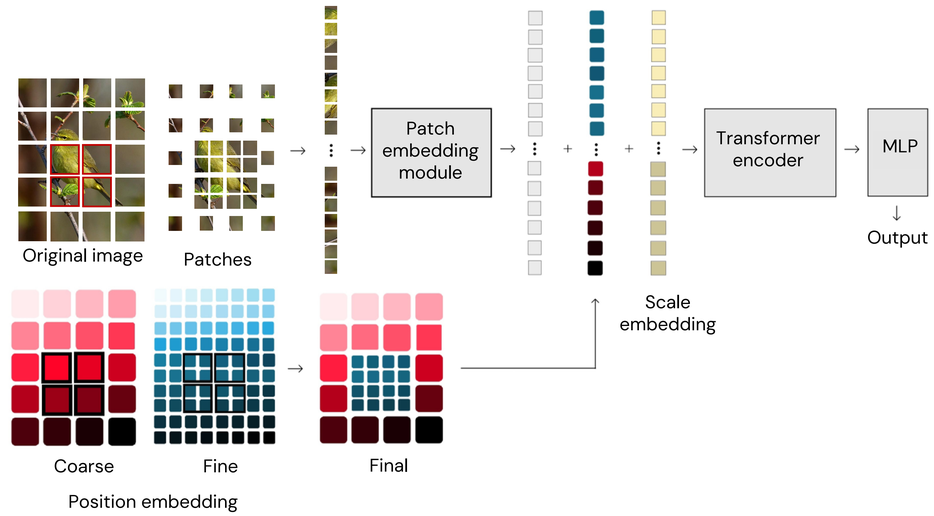
Code

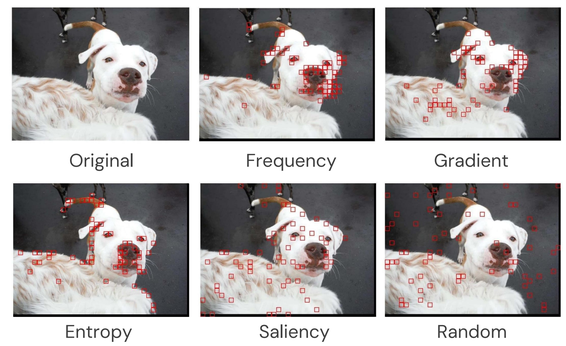
1. Problem: Current AI models fail to capture aesthetics



- Downscaling and cropping can **distort aesthetic properties** by altering composition, changing aspect ratios, and removing high-resolution details, resulting in **inaccurate** image aesthetic assessment (IAA).
- Existing solutions often **slow convergence**, **degrade IAA model performance**, and **increase computational costs considerably**.

2. Solution: Filling the gap between human and machine vision using CHARM



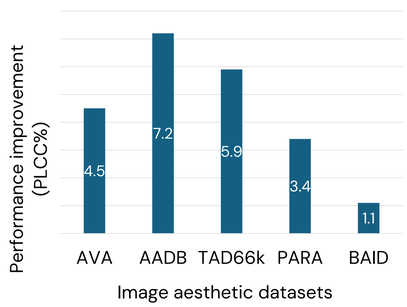


Patch selection

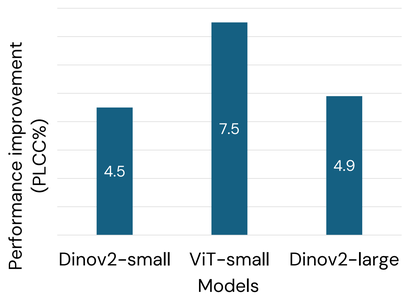
Best approach:
- Training: Random
- Inference: Frequency

- Inspired by the human brain, CHARM processes visual information by preserving **multi-resolution details**, maintaining **spatial relationships**, and applying **selective attention**.

3. Evaluation: When CHARM works like a charm



Performance improvement of Dinov2-small with CHARM

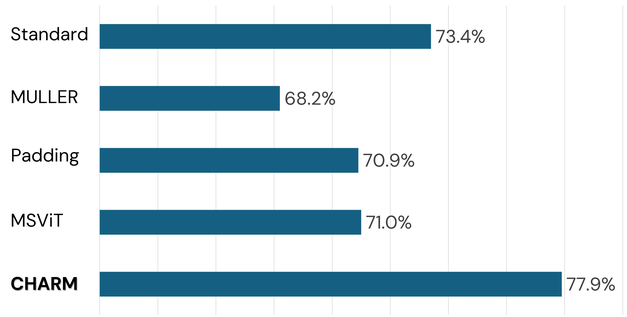


Performance improvement of different models with CHARM on the AVA dataset

- CHARM **significantly improves the performance** of various ViTs on different datasets.
- CHARM can be integrated with other methods for further performance gain.

4. CHARM vs Others

- Unlike existing methods, CHARM tokenization improves ViT performance on IAA by preserving high-resolution details, aspect ratio, and multi-scale information **simultaneously**.



Dinov2-small performance (PLCC) using different image preprocessing approaches on the AVA dataset

5. Computational costs of CHARM

Model	Input size	CHARM	#tokens	ms	GMACs	MB
Dinov2 -small	224 x 224	-	256	**5.7**	**6.11**	**202.9**
	640 x 640	-	2070	32.8	84.01	2091.8
		✓	2-scale:512	7.3 (↓ 77.7%)	13.46 (↓ 84%)	346.0 (↓ 83.5%)
		✓	3-scale:700	9.3 (↓ 71.6%)	19.60 (↓ 76.7%)	494.3 (↓ 76.4%)

Dinov2-small inference cost breakdown for processing a single image.

- CHARM **significantly decreases** the costs of processing an image in its original size.

6. The resolution dilemma: How much detail is enough?

Approach	PLCC	SRCC	ACC
Standard	0.904	0.855	0.863
Maximum edge = 1024	0.938	0.905	0.892
Maximum edge = 1500	**0.940**	**0.908**	**0.900**

Performance of Dinov2-small on the PARA dataset with different image resolutions.

- There is a limit to performance improvement from preserving high-resolution details; beyond a certain point, the gains become minimal.

7. Conclusion

- Processing images at their **original** resolution is crucial for accurate image aesthetic assessment.
- **CHARM** boosts IAA model performance by efficiently preserving **C**omposition, **H**igh resolution, **A**spect **R**atio, and **M**ulti-scale information without overwhelming computational resources.