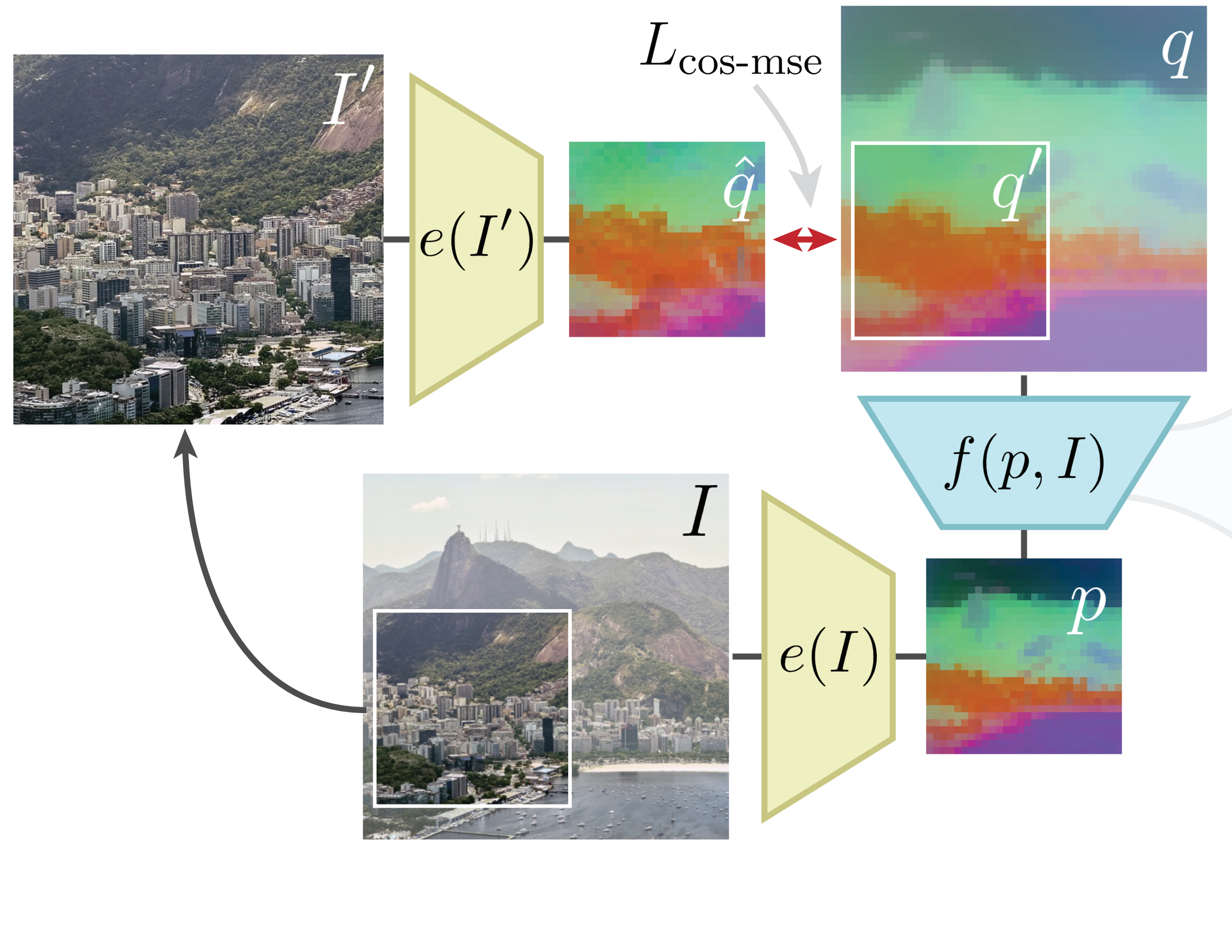
AnyUp: Universal Feature Upsampling

Thomas Wimmer[1,2], Prune Truong[3], Marie-Julie Rakotosaona[3], Michael Oechsle[3], Federico Tombari[3,4], Bernt Schiele[1], Jan Eric Lenssen[1]
[1]Max Planck Institute for Informatics, [2]ETH Zurich, [3]Google, [4]Technical University of Munich

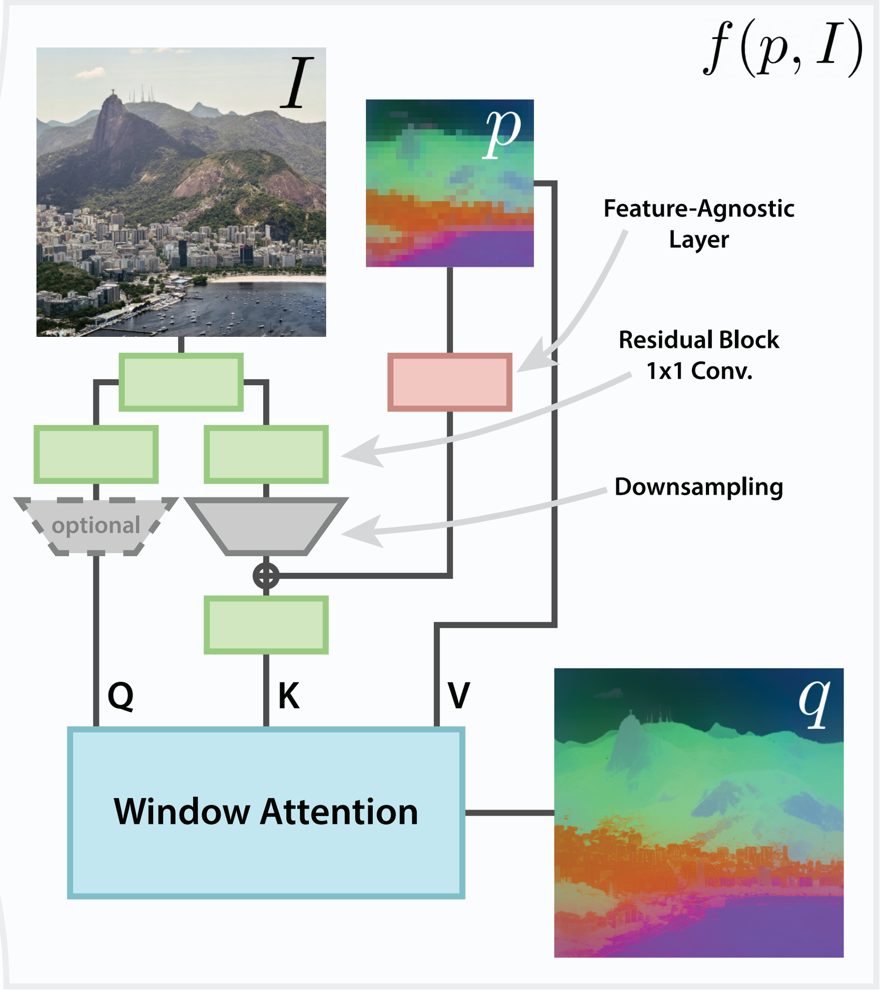
https://wimmerth.github.io/anyup

Upsample Any Feature at Any Resolution for Any Task without Retraining!

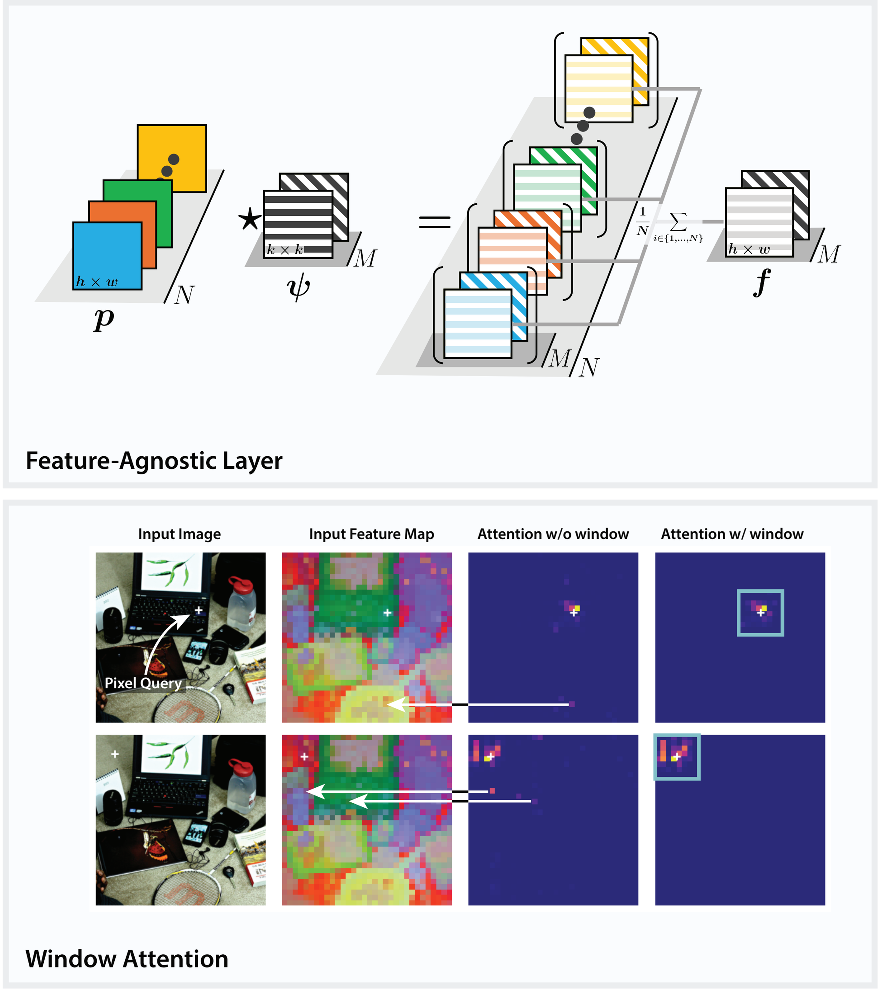
Method



Training Strategy



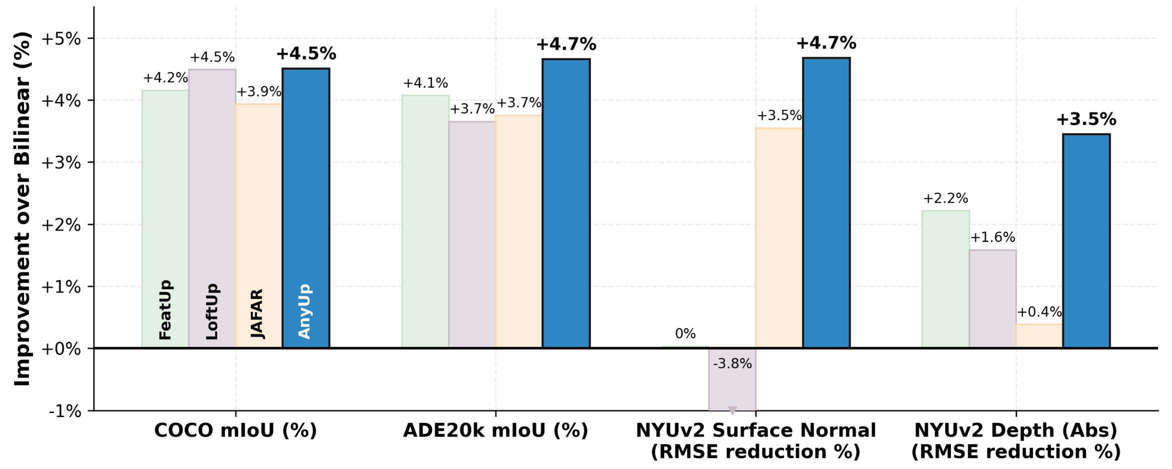
Architecture



Details

Results

AnyUp beats feature-specific upsamplers in downstream linear probing.

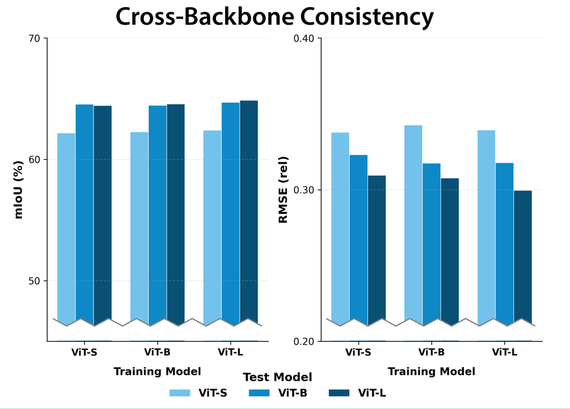


AnyUp generalizes well to unseen feature extractors at test time.

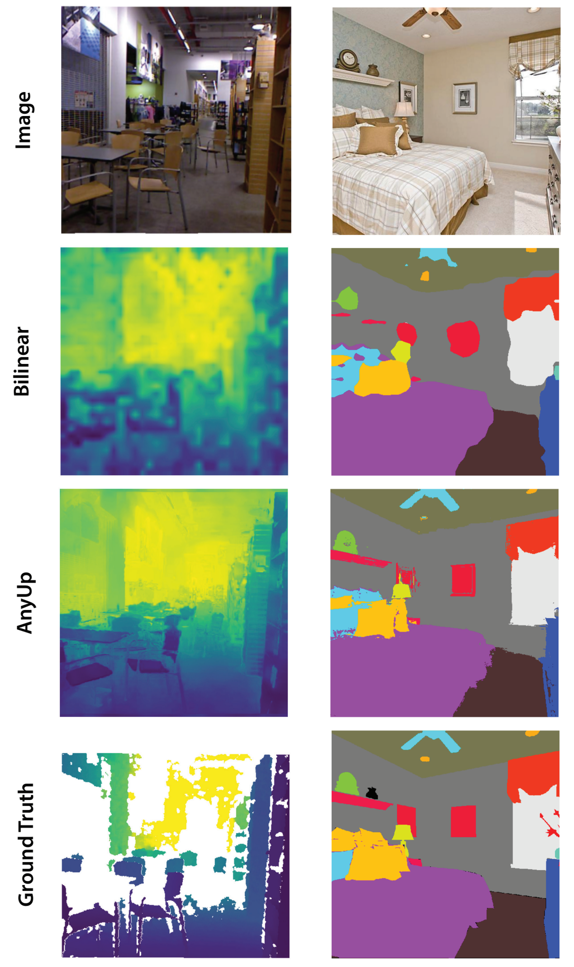
Generalization to other models

	Train Model	Test Model	mIoU (Acc)	RMSE (abs / rel)
AnyUp	DINOv2 (ViT-S)	SigLIP 2 LoftUp	51.68 (73.35)	0.59 / 0.48
AnyUp	SigLIP 2 LoftUp	SigLIP 2 LoftUp	**54.45 (75.49)**	**0.57 / 0.46**
LoftUp	SigLIP 2 LoftUp	SigLIP 2 LoftUp	40.73 (64.87)	0.72 / 0.60
AnyUp	DINOv2 (ViT-S)	SigLIP 2 JAFAR	58.51 (78.36)	0.91 / 0.58
AnyUp	SigLIP 2 JAFAR	SigLIP 2 JAFAR	**60.32 (79.57)**	**0.90 / 0.57**
JAFAR	SigLIP 2 JAFAR	SigLIP 2 JAFAR	60.10 (79.40)	0.93 / 0.60
AnyUp	DINOv2 (ViT-S)	DINOv3 (ViT-S+)	62.96 (81.82)	0.51 / 0.37
AnyUp	DINOv3 (ViT-S+)	DINOv3 (ViT-S+)	62.99 (81.84)	0.51 / 0.37

Cross-Backbone Consistency

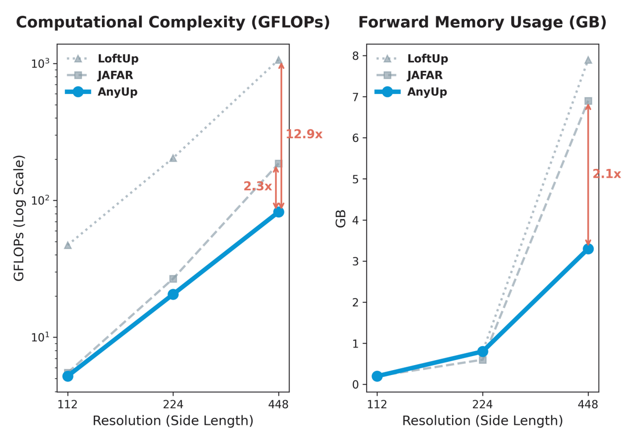


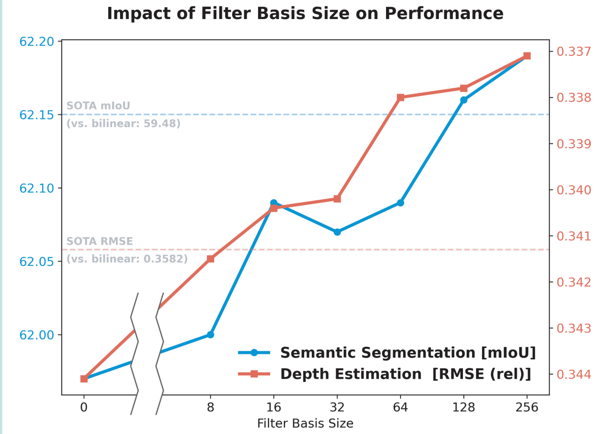
Qualitative Probing Results



Analysis

AnyUp is more efficient than previous attention-based feature upsamplers through its usage of window attention.



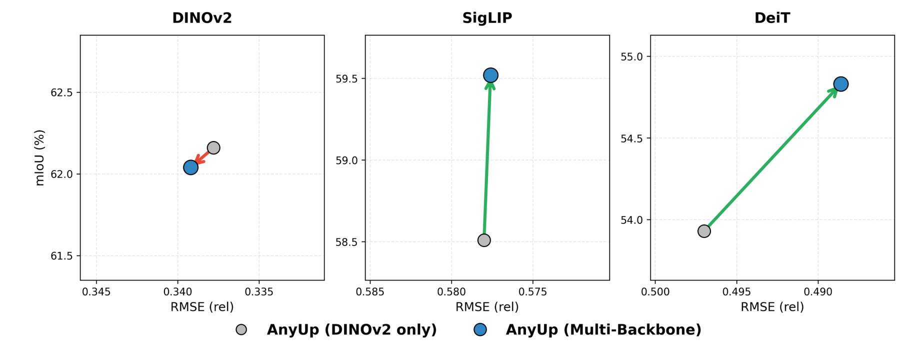


"Only looking at the RGB image" by matching colors and positions between a high-resolution image and its downsampled version is a surprisingly strong baseline.

AnyUp retains the input feature space while improving upsampling quality.

	Semantic Segmentation mIoU (↑)	Semantic Segmentation Acc. (↑)	Depth Estimation RMSE (↓)	Depth Estimation δ_1 (↑)
Bilinear	39.73	73.32	0.506	0.816
Guided Filter	37.54	72.25	0.518	0.813
FeatUp	40.19	74.05	0.504	0.818
LoftUp	4.27	46.58	0.765	0.802
JAFAR	39.06	73.75	0.503	0.815
AnyUp	**40.83**	**74.94**	**0.498**	**0.822**

Training on multiple feature extractors improves generalization while achieving comparable performance on seen backbones.



Limitations:

- Upsampled features are a linear combination of input features
  → Sub-patch-level information in the features is not leveraged
- Feature upsampling is post-hoc and bound to the quality of extracted low-resolution features.

Usage:

```python
# load the AnyUp upsampler model
upsampler   = torch.hub.load('wimmerth/anyup', 'anyup')

# upsampled high-resolution features (B, C, H, W)
hr_features = upsampler(hr_image, lr_features)
```