Novelty Assessment Report

Paper: AnyUp: Universal Feature Upsampling
PDF URL: https://openreview.net/pdf?id=Y9UAgPehqo
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-01

Abstract

We introduce AnyUp, a method for feature upsampling that can be applied to any vision feature at any resolution, without encoder-specific training. Existing learning-based upsamplers for features like DINO or CLIP need to be re-trained for every feature extractor and thus do not generalize to different feature types at inference time. In this work, we propose an inference-time feature-agnostic upsampling architecture to alleviate this limitation and improve upsampling quality. In our experiments, AnyUp sets a new state of the art for upsampled features, generalizes to different feature types, and preserves feature semantics while being efficient and easy to apply to a wide range of downstream tasks.

Core Task Landscape

This paper addresses: **universal feature upsampling across vision encoders and resolutions**

A total of **24 papers** were analyzed and organized into a taxonomy with **12 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Universal Feature Upsampling Methods**
- **Encoder-Specific Feature Upsampling**
- **Task-Specific Resolution Enhancement**

Complete Taxonomy Tree

- universal feature upsampling across vision encoders and resolutions Survey Taxonomy
- Universal Feature Upsampling Methods
  - Inference-Time Universal Upsamplers ★ (2 papers)
  - [0] AnyUp: Universal Feature Upsampling (Anon et al., 2026) View paper
  - [20] Upsample Anything: A Simple and Hard to Beat Baseline for Feature Upsampling (Minseok Seo, 2025) View paper
  - Trainable Universal Upsamplers (3 papers)
  - [2] JAFAR: Jack up Any Feature at Any Resolution (Couairon, 2025) View paper
  - [10] NAF: Zero-Shot Feature Upsampling via Neighborhood Attention Filtering (Loick Chambon, 2025) View paper
  - [18] FeatUp: A Model-Agnostic Framework for Features at Any Resolution (Fu, 2024) View paper
- Encoder-Specific Feature Upsampling
  - Vision Foundation Model Feature Enhancement (4 papers)
  - [3] FeatSharp: Your Vision Model Features, Sharper (Heinrich, 2025) View paper
  - [8] LoftUp: Learning a Coordinate-Based Feature Upsampler for Vision Foundation Models (Huang Hai-wen, 2025) View paper
  - [12] Benchmarking Feature Upsampling Methods for Vision Foundation Models using Interactive Segmentation (Huang Hai-wen, 2025) View paper
  - [23] Visibility Improvement in Grad-CAM via High-Resolution Feature Maps with FeatUp (KAMEYA, 2025) View paper
  - Transformer-Based Upsampling Architectures (2 papers)
  - [5] MGD-SAM2: Multi-view Guided Detail-enhanced Segment Anything Model 2 for High-Resolution Class-agnostic Segmentation (Shen Haoran, 2025) View paper
  - [9] VST++: Efficient and Stronger Visual Saliency Transformer (Nian Liu, 2023) View paper
- Task-Specific Resolution Enhancement
  - Image Super-Resolution Methods
  - Arbitrary-Scale Super-Resolution (3 papers)
    - [6] Image neural field diffusion models (Yinbo Chen, 2024) View paper
    - [7] Continuous remote sensing image super-resolution based on context interaction in implicit function space (Keyan Chen, 2023) View paper
    - [22] OPE-SR: Orthogonal Position Encoding for Designing a Parameter-free Upsampling Module in Arbitrary-scale Image Super-Resolution (Gaochao Song, 2023) View paper
  - Frequency-Guided Super-Resolution (2 papers)
    - [13] Fourier-Guided Attention Upsampling for Image Super-Resolution (Daejune Choi, 2025) View paper
    - [21] SR-AFU: super-resolution network using adaptive frequency component upsampling and multi-resolution features (Kejia Chen, 2022) View paper
  - Multi-Frame Super-Resolution (1 papers)
    - [11] Adaptive Feature Consolidation Network for Burst Super-Resolution (Nancy Mehta, 2022) View paper
  - Dense Prediction Task Upsampling

- Cross-Resolution Detection and Segmentation (2 papers)
  - [15] Cross Resolution Encoding-Decoding For Detection Transformers (Kumar, 2024) View paper
  - [16] Domain-specific augmentations with resolution agnostic self-attention mechanism improves choroid segmentation in optical coherence tomography images (Burke, 2024) View paper
- Depth and Height Estimation (2 papers)
  - [4] Depth2Elevation: Scale Modulation with Depth Anything Model for Single-view Remote Sensing Image Height Estimation (Zhongcheng Hong, 2025) View paper
  - [17] Multi-scale Feature Distribution Prediction Decoder for Lightweight Monocular Depth Estimation (Xiasheng Ma, 2024) View paper
- Multi-Scale Optical Flow Estimation (1 papers)
  - [24] High Resolution Multi-Scale RAFT (Robust Vision Challenge 2022) (Jahedi, 2022) View paper
- Domain-Specific Resolution Enhancement
- Video Action Recognition (1 papers)
  - [1] Leveraging cross-resolution attention for effective extreme low-resolution video action recognition (OÄuzhan OÄuz, 2024) View paper
- Medical and Scientific Imaging (2 papers)
  - [14] Resolution- and Stimulus-Agnostic Super-Resolution of Ultra-High-Field Functional MRI: Application to Visual Studies (Hongwei Bran Li, 2024) View paper
  - [19] Enhancing Object Detection in Layout Analysis: Leveraging Vision Transformer and Fourier Neural Operator (Yahui Yang, 2024) View paper

Narrative

Core task: universal feature upsampling across vision encoders and resolutions. The field addresses the challenge of recovering high-resolution spatial detail from coarse feature maps produced by diverse vision encoders, which is essential for dense prediction tasks such as segmentation and depth estimation. The taxonomy organizes approaches into three main branches. Universal Feature Upsampling Methods aim to develop encoder-agnostic techniques that generalize across different backbone architectures and input resolutions, often leveraging learned upsampling modules or implicit representations. Encoder-Specific Feature Upsampling tailors solutions to particular network families, exploiting architectural priors or training-time adaptations to achieve tighter integration with specific encoders. Task-Specific Resolution Enhancement focuses on domain-driven strategies, where upsampling is optimized for particular applications such as medical imaging, remote sensing, or video analysis, often incorporating task-relevant inductive biases.

Recent work has explored trade-offs between generality and performance. Universal methods like FeatUp[18] and Upsample Anything[20] pursue broad applicability by training upsampling networks that can handle features from multiple encoders without retraining, while AnyUp[0] extends this paradigm by proposing an inference-time universal upsampler that adapts on-the-fly to unseen encoders and resolutions. This contrasts with encoder-specific approaches such as Cross Resolution Attention[1] or task-driven techniques like MGD-SAM2[5], which sacrifice some generality for tighter coupling to particular architectures or domains. A key open question is whether universal upsamplers can match the fidelity of specialized methods while maintaining their flexibility. AnyUp[0] sits within the inference-time universal branch alongside Upsample Anything[20], emphasizing zero-shot adaptability, whereas FeatUp[18] represents an earlier training-based universal approach that requires pre-training on a fixed set of encoders.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Upsample Anything: A Simple and Hard to Beat Baseline for Feature Upsampling

Authors: Minseok Seo, Mark Hamilton, Changick Kim | **Year/Venue**: 2025 | **URL**: View paper

Abstract

We present \textbf{Upsample Anything}, a lightweight test-time optimization (TTO) framework that restores low-resolution features to high-resolution, pixel-wise outputs without any training. Although Vision Foundation Models demonstrate strong generalization across diverse downstream tasks, their representations are typically downsampled by 14x/16x (e.g., ViT), which limits their direct use in pixel-level applications. Existing feature upsampling approaches depend on dataset-specific retraining ...

Relationship Analysis

Both papers belong to the Inference-Time Universal Upsamplers category, focusing on feature upsampling without encoder-specific training. While AnyUp proposes a learned feature-agnostic architecture with window attention and crop-based training to generalize across encoders and resolutions, Upsample Anything takes a fundamentally different approach using test-time optimization (TTO) with anisotropic Gaussian kernels, requiring no training at all but optimizing per-image at inference. The key distinction is that AnyUp is a trainable model designed for universal applicability, whereas Upsample Anything is a training-free optimization method that learns kernels on-the-fly for each input.

Contributions Analysis

Overall novelty summary. The paper proposes AnyUp, an inference-time feature upsampling method designed to work with any vision encoder at any resolution without encoder-specific training. Within the taxonomy, it resides in the 'Inference-Time Universal Upsamplers' leaf, which contains only two papers including this one. This represents a sparse research direction within the broader 'Universal Feature Upsampling Methods' branch, suggesting the work addresses a relatively underexplored problem space compared to encoder-specific or task-specific upsampling approaches that dominate other branches.

The taxonomy reveals that most upsampling research concentrates on encoder-specific methods (six papers across vision foundation models and transformer architectures) or task-specific approaches (fifteen papers spanning super-resolution, dense prediction, and domain-specific applications). The 'Trainable Universal Upsamplers' sibling branch contains three papers that require training on diverse features, whereas AnyUp's inference-time approach diverges by eliminating training requirements entirely. This positioning suggests the work bridges a gap between the flexibility of universal methods and the practicality of zero-shot deployment.

Among twenty candidates examined across three contributions, none were identified as clearly refuting the core claims. The main contribution 'AnyUp: feature-agnostic upsampling model' examined ten candidates with zero refutable matches, as did the 'Feature-agnostic layer' contribution. The 'Window attention architecture with crop-based training' was not evaluated against any candidates. Given this limited search scope of twenty papers from semantic search and citation expansion, the analysis suggests no immediate prior work overlap within the examined set, though the small candidate pool means substantial related work may exist beyond this sample.

Based on the limited literature search covering twenty candidates, the work appears to occupy a novel position within a sparse research direction. The taxonomy structure indicates that while universal upsampling is an established goal, inference-time approaches without training remain rare. However, the small search scope and the presence of only one sibling paper limit confidence in assessing broader field coverage or potential overlaps with work outside the examined candidates.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: AnyUp: feature-agnostic upsampling model

Description: AnyUp is a universal feature upsampling method that can be trained once and then applied to features from any vision encoder at any resolution without requiring encoder-specific retraining, unlike existing methods that must be retrained for each feature extractor.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Recent advances in 2d image upscaling: a comprehensive review

URL: View paper

Brief Assessment

Image Upscaling Review[39] focuses on 2D image upscaling techniques for visual quality enhancement, not feature upsampling for vision encoders. The candidate discusses image resizing and visual quality, while the original paper addresses upsampling semantic features from vision transformers for downstream tasks.

---

2. Swin transformer v2: Scaling up capacity and resolution

URL: View paper

Brief Assessment

Swin Transformer V2[35] focuses on scaling transformer architectures for vision tasks and transferring models across image resolutions, not on universal feature upsampling across different vision encoders. The candidate addresses position bias transfer for a single architecture, while the original proposes encoder-agnostic feature upsampling applicable to any vision model.

---

3. Arbitrary-Scale Image Generation and Upsampling Using Latent Diffusion Model and Implicit Neural Decoder

URL: View paper

Brief Assessment

Arbitrary Scale Generation[36] focuses on image generation and super-resolution in pixel/latent space using diffusion models and implicit neural decoders, not on universal feature upsampling across different vision encoders. The candidate operates on RGB images rather than arbitrary vision encoder features.

---

4. SFA-Net: Semantic Feature Adjustment Network for Remote Sensing Image Segmentation

URL: View paper

Brief Assessment

SFA-Net[41] focuses on semantic segmentation of remote sensing images using a hybrid CNN-transformer architecture with feature adjustment modules, not on universal feature upsampling across different vision encoders.

---

5. FeatSharp: Your Vision Model Features, Sharper

URL: View paper

Brief Assessment

FeatSharp[3] focuses on upsampling features from specific pre-trained vision models using tiling and denoising techniques, but requires model-specific training. It does not address the universal, encoder-agnostic upsampling capability that is central to the original paper's contribution.

---

6. Upsample guidance: Scale up diffusion models without training

URL: View paper

Brief Assessment

Upsample Guidance[37] focuses on upsampling diffusion model outputs (images/videos) without training, while AnyUp addresses upsampling vision encoder features for downstream tasks. These are fundamentally different domains with different technical approaches.

---

7. Learned image downscaling for upscaling using content adaptive resampler

URL: View paper

Brief Assessment

Content Adaptive Resampler[38] focuses on learned image downscaling for subsequent upscaling in the image domain (RGB to RGB), not on upsampling features from vision encoders across different architectures and resolutions.

---

8. U-repa: Aligning diffusion u-nets to vits

URL: View paper

Brief Assessment

U-repa[40] focuses on aligning diffusion U-Net hidden states with ViT encoders for generative modeling, not on universal feature upsampling across different vision encoders for downstream tasks. The technical domains and objectives are fundamentally different.

---

9. Nas-fpn: Learning scalable feature pyramid architecture for object detection

URL: View paper

Brief Assessment

NAS-FPN[42] focuses on learning feature pyramid architectures for object detection through neural architecture search, specifically designing cross-scale connections for multi-scale object detection. This is fundamentally different from AnyUp's universal feature upsampling approach that works across any vision encoder without retraining.

---

10. Annotation-free open-vocabulary segmentation for remote-sensing images

URL: View paper

Brief Assessment

Open Vocabulary Segmentation[43] focuses on semantic segmentation in remote sensing using SimFeatUp, which is an upsampler trained on remote sensing data for a specific domain application. AnyUp addresses universal feature upsampling across any vision encoder without encoder-specific retraining, representing a fundamentally different scope and generalization capability.

Contribution 2: Feature-agnostic layer

Description: A convolutional layer design that processes input channels independently using a learned kernel basis and aggregates contributions across channels, enabling the model to handle features of arbitrary dimensionality while capturing structural information.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Learning One Convolutional Layer with Overlapping Patches

URL: View paper

Brief Assessment

Learning Overlapping Patches[32] focuses on learning convolutional filters for overlapping patches in a supervised learning setting with specific patch structures. The original paper's feature-agnostic layer is designed for upsampling arbitrary-dimensional features from vision encoders in a self-supervised manner, which is a fundamentally different problem domain and architecture.

---

2. Splinecnn: Fast geometric deep learning with continuous b-spline kernels

URL: View paper

Brief Assessment

SplineCNN[26] processes spatial geometric relationships in graphs/meshes using B-spline kernels, not arbitrary-dimensional features from vision encoders. The methods address fundamentally different domains and input types.

---

3. gWaveNet: Classification of gravity waves from noisy satellite data using custom kernel integrated deep learning method

URL: View paper

Brief Assessment

gWaveNet[27] proposes a checkerboard kernel for gravity wave detection in satellite images, which is domain-specific and operates on single-channel grayscale images. The original paper's feature-agnostic layer processes multi-channel features of arbitrary dimensionality from vision encoders, representing a fundamentally different architectural component and application domain.

---

4. K3DN: Disparity-Aware Kernel Estimation for Dual-Pixel Defocus Deblurring

URL: View paper

Brief Assessment

K3DN[34] focuses on dual-pixel defocus deblurring using disparity-aware kernel estimation, not on feature upsampling with arbitrary dimensionality processing. The technical domains and objectives are fundamentally different.

---

5. Expert Kernel Generation Network Driven by Contextual Mapping for Hyperspectral Image Classification

URL: View paper

Brief Assessment

Expert Kernel Generation[31] focuses on hyperspectral image classification using context-aware mapping for dynamic convolution kernels in 3D-CNNs, not on general feature upsampling across arbitrary vision encoders. The technical approach and application domain differ fundamentally from the original paper's feature-agnostic upsampling method.

---

6. Fully convolutional mesh autoencoder using efficient spatially varying kernels

URL: View paper

Brief Assessment

Mesh Autoencoder[25] focuses on mesh convolution with spatially varying kernels for 3D geometry processing, not on processing arbitrary-dimensional features from vision encoders. The technical domains and applications are fundamentally different.

---

7. Convolutional kernel-based classification of industrial alarm floods

URL: View paper

Brief Assessment

Alarm Flood Classification[29] focuses on alarm time series classification using convolutional kernel-based transformations (multirocket) for industrial process control, not on processing arbitrary dimensionality features with learned kernel basis for vision tasks.

---

8. Deep Network With Irregular Convolutional Kernels and Self-Expressive Property for Classification of Hyperspectral Images

URL: View paper

Brief Assessment

Irregular Convolutional Kernels[33] focuses on hyperspectral image classification using PCA-based irregular patches as kernels, not a learned kernel basis for arbitrary dimensionality features. The technical approaches are fundamentally different.

---

9. Towards a General Purpose CNN for Long Range Dependencies in D

URL: View paper

Brief Assessment

General Purpose CNN[28] focuses on continuous convolutional kernels for handling arbitrary resolutions and dimensionalities in spatial data (1D, 2D, 3D), not on processing feature channels independently with learned kernel basis for arbitrary-dimensional features as in the original paper's feature-agnostic layer.

---

10. Omni-dimensional dynamic convolution feature coordinate attention network for pneumonia classification

URL: View paper

Brief Assessment

Omni-dimensional Dynamic Convolution[30] focuses on pneumonia classification using dynamic convolution across four dimensions (spatial, input/output channels, kernel count) for medical imaging, not on creating a general feature-agnostic upsampling layer for arbitrary vision features.

Contribution 3: Window attention architecture with crop-based training

Description: An upsampling architecture that restricts attention computation to local windows and employs a training strategy using randomly sampled image crops as supervision, combined with consistency regularization to preserve the original feature space.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] AnyUp: Universal Feature Upsampling View paper
- [1] Leveraging cross-resolution attention for effective extreme low-resolution video action recognition View paper
- [2] JAFAR: Jack up Any Feature at Any Resolution View paper
- [3] FeatSharp: Your Vision Model Features, Sharper View paper
- [4] Depth2Elevation: Scale Modulation with Depth Anything Model for Single-view Remote Sensing Image Height Estimation View paper
- [5] MGD-SAM2: Multi-view Guided Detail-enhanced Segment Anything Model 2 for High-Resolution Class-agnostic Segmentation View paper
- [6] Image neural field diffusion models View paper
- [7] Continuous remote sensing image super-resolution based on context interaction in implicit function space View paper
- [8] LoftUp: Learning a Coordinate-Based Feature Upsampler for Vision Foundation Models View paper
- [9] VST++: Efficient and Stronger Visual Saliency Transformer View paper
- [10] NAF: Zero-Shot Feature Upsampling via Neighborhood Attention Filtering View paper
- [11] Adaptive Feature Consolidation Network for Burst Super-Resolution View paper
- [12] Benchmarking Feature Upsampling Methods for Vision Foundation Models using Interactive Segmentation View paper
- [13] Fourier-Guided Attention Upsampling for Image Super-Resolution View paper
- [14] Resolution- and Stimulus-Agnostic Super-Resolution of Ultra-High-Field Functional MRI: Application to Visual Studies View paper
- [15] Cross Resolution Encoding-Decoding For Detection Transformers View paper
- [16] Domain-specific augmentations with resolution agnostic self-attention mechanism improves choroid segmentation in optical coherence tomography images View paper
- [17] Multi-scale Feature Distribution Prediction Decoder for Lightweight Monocular Depth Estimation View paper
- [18] FeatUp: A Model-Agnostic Framework for Features at Any Resolution View paper
- [19] Enhancing Object Detection in Layout Analysis: Leveraging Vision Transformer and Fourier Neural Operator View paper
- [20] Upsample Anything: A Simple and Hard to Beat Baseline for Feature Upsampling View paper
- [21] SR-AFU: super-resolution network using adaptive frequency component upsampling and multi-resolution features View paper
- [22] OPE-SR: Orthogonal Position Encoding for Designing a Parameter-free Upsampling Module in Arbitrary-scale Image Super-Resolution View paper
- [23] Visibility Improvement in Grad-CAM via High-Resolution Feature Maps with FeatUp View paper
- [24] High Resolution Multi-Scale RAFT (Robust Vision Challenge 2022) View paper
- [25] Fully convolutional mesh autoencoder using efficient spatially varying kernels View paper
- [26] Splinecnn: Fast geometric deep learning with continuous b-spline kernels View paper
- [27] gWaveNet: Classification of gravity waves from noisy satellite data using custom kernel integrated deep learning method View paper
- [28] Towards a General Purpose CNN for Long Range Dependencies in D View paper
- [29] Convolutional kernel-based classification of industrial alarm floods View paper
- [30] Omni-dimensional dynamic convolution feature coordinate attention network for pneumonia classification View paper
- [31] Expert Kernel Generation Network Driven by Contextual Mapping for Hyperspectral Image Classification View paper
- [32] Learning One Convolutional Layer with Overlapping Patches View paper
- [33] Deep Network With Irregular Convolutional Kernels and Self-Expressive Property for Classification of Hyperspectral Images View paper
- [34] K3DN: Disparity-Aware Kernel Estimation for Dual-Pixel Defocus Deblurring View paper
- [35] Swin transformer v2: Scaling up capacity and resolution View paper
- [36] Arbitrary-Scale Image Generation and Upsampling Using Latent Diffusion Model and Implicit Neural Decoder View paper
- [37] Upsample guidance: Scale up diffusion models without training View paper
- [38] Learned image downscaling for upscaling using content adaptive resampler View paper
- [39] Recent advances in 2d image upscaling: a comprehensive review View paper
- [40] U-repa: Aligning diffusion u-nets to vits View paper
- [41] SFA-Net: Semantic Feature Adjustment Network for Remote Sensing Image Segmentation View paper
- [42] Nas-fpn: Learning scalable feature pyramid architecture for object detection View paper
- [43] Annotation-free open-vocabulary segmentation for remote-sensing images View paper