TaxaDiffusion: Progressively Trained Diffusion Model for Fine-Grained Species Generation

Amin Karimi Monsefi, Mridul Khurana, Rajiv Ramnath, Anuj Karpatne, Wei-Lun Chao, Cheng Zhang

ICCV 2025

THE OHIO STATE UNIVERSITY

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Motivation

- Gemini 2.5 Flash: Golden Retriever
  - Prompt: generate an image of Golden Retriever.

Real Image

Generated Image

Motivation & Challenges

- Gemini 2.5 Flash: Abudefduf abdominalis
  - Prompt: generate an image of Abudefduf abdominalis.

Real Image

Generated Image

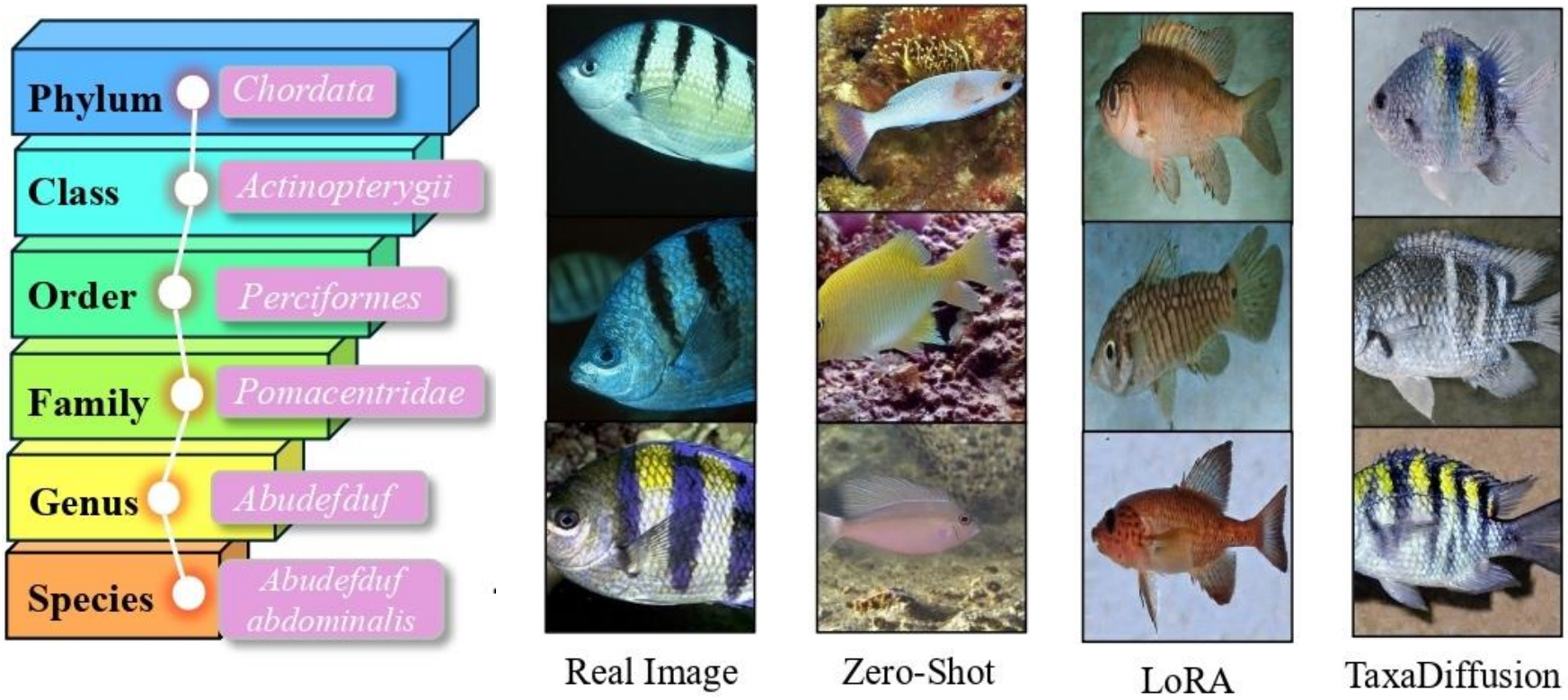
Motivation & Challenges

- Why generative models can’t generate some species?
  - Limited training data per species.
  - High morphological diversity within species.

- Question:
  - How can we learn high-quality generative models for animals without a large amount of data per species?

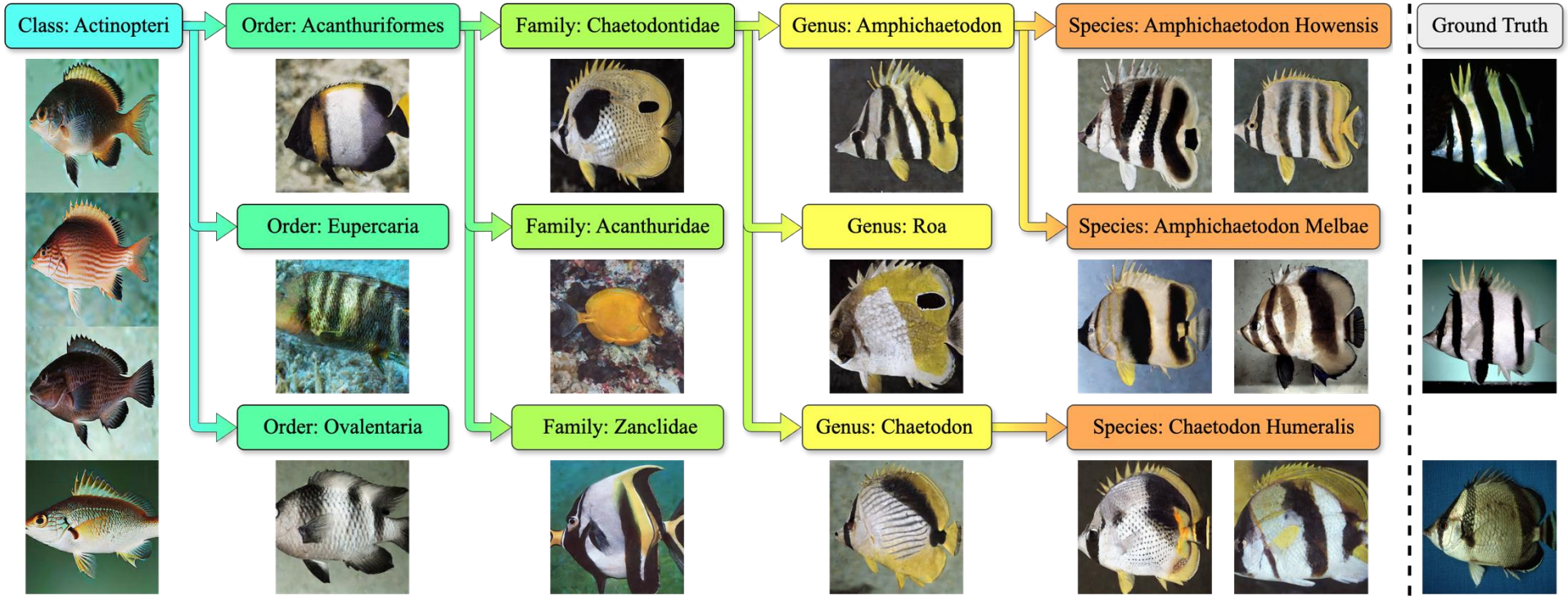
Goal

- Generate accurate, fine-grained species images.
  - We introduced: TaxaDiffusion.



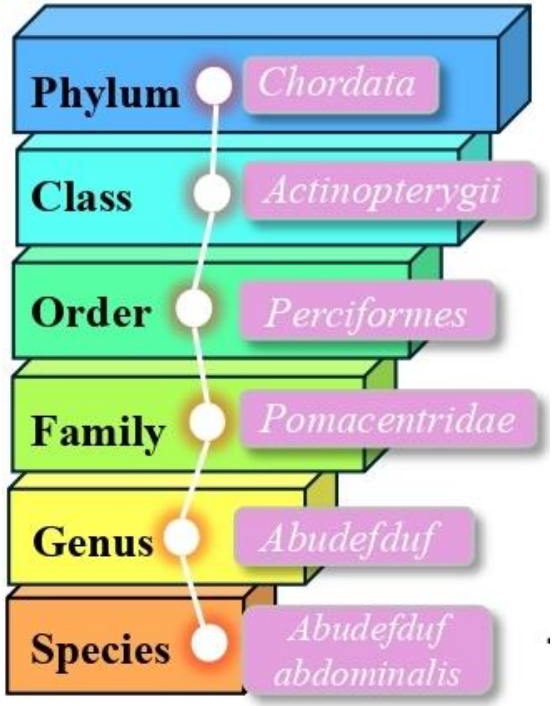
Key Insight - Leveraging Taxonomy

- Taxonomy:
  - Shared ancestry leads to similar characteristics in species.
  - Closely related species (same Genus/Family) differ mainly in subtle shape, color, or pattern variations.



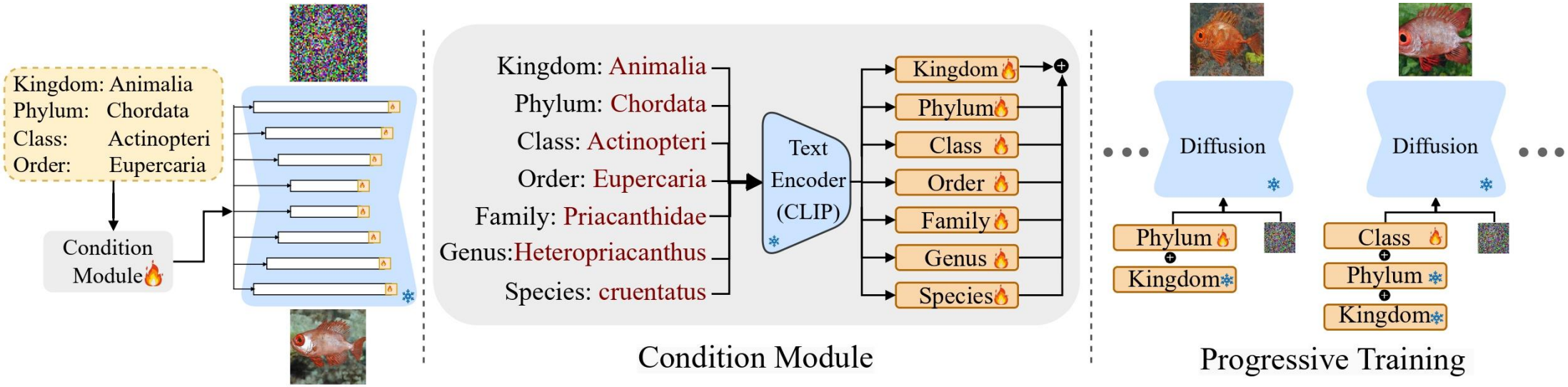
Key Insight - Leveraging Taxonomy

- Question: How can we use taxonomy in training diffusion models?



TaxaDiffusion

- Efficient domain adaptation using LoRA.
- Use of CLIP embeddings refined through transformer-based module.
- Progressive Training (Kingdom → Species).



TaxaDiffusion: Fine-grained Generation

- Classifier-Free Guidance (CFG):
  - Generation quality and diversity by refining predictions without a separate classifier.
  - Combines conditional and unconditional score estimates.

$$\tilde{\boldsymbol{\epsilon}}_\theta(\mathbf{x}_t, t, \mathbf{c}) = (1+w) \times \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{c}) - w \times \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t),$$

  - TaxaDiffusion CFG:

$$(1+w) \times \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{c}^{(i)}) - w \times \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{c}^{(0)}),$$

Experimental Setup

Evaluation on 3 Datasets:

- FishNet: 17,357 distinct fish
- BIOSCAN-1M: 8,355 insect species
- iNaturalist: 10,000 species

Metrics:

- FID, LPIPS (image quality)
- BioCLIP (semantic alignment)

Results & Performance

	Ground Truth	TaxaDiffusion (ours)	Stable Diffusion	Stable Diffusion + LoRA	Stable Diffusion + Finetuning
Abalistes filamentosus					

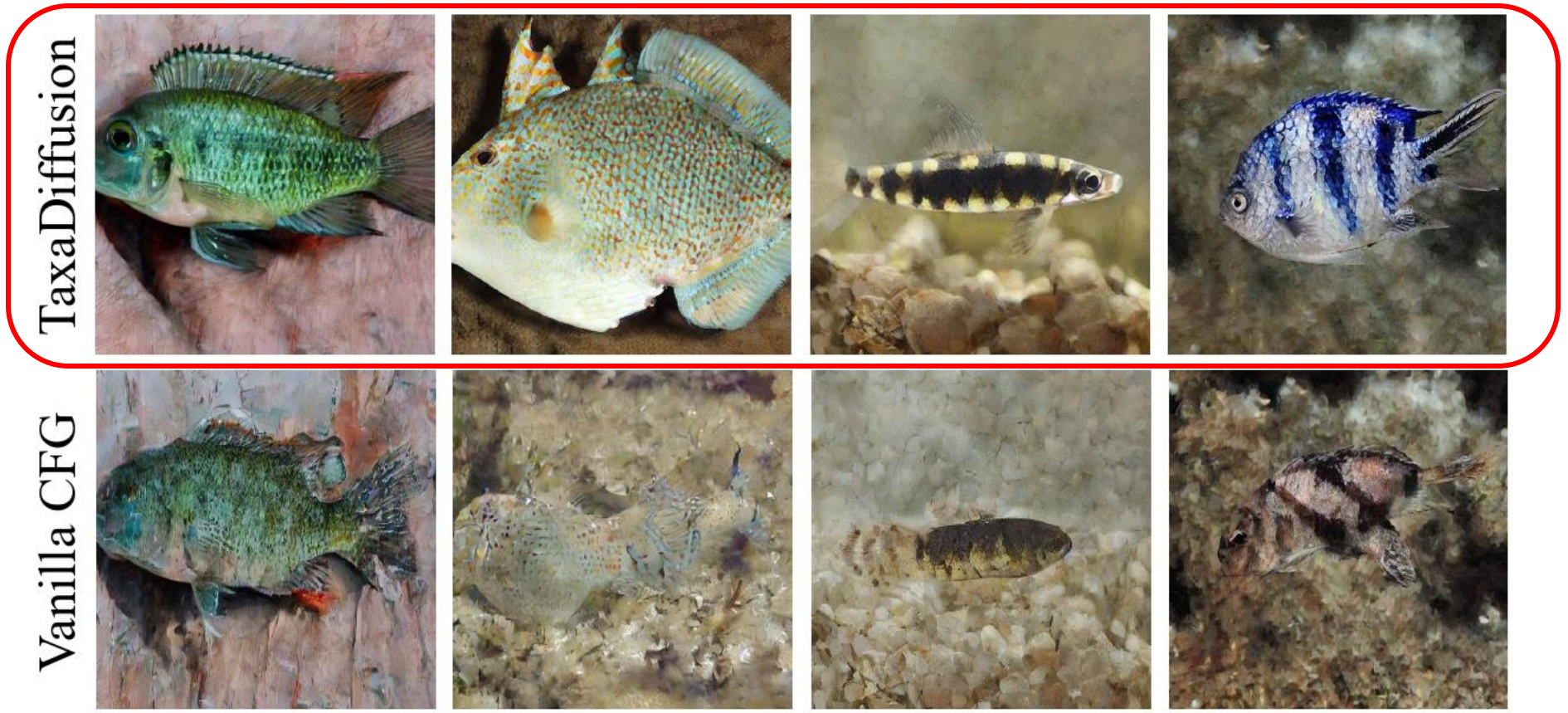
Results & Performance

Table 1. **Quantitative results on FishNet** [17]. We report FID and LPIPS to assess image generation quality and BioCLIP for text-to-image alignment, where the text corresponds to the species "taxonomic name." We generate 10 images per category at each level for evaluation, i.e., `Order`, `Family`, `Genus`, and `Species`. TaxaDiffusion outperforms the baselines for generating species-specific images.

Method	FID ↓				LPIPS ↓				BioCLIP - Score ↑			
	Order	Family	Genus	Species	Order	Family	Genus	Species	Order	Family	Genus	Species
SD [36]	74.98	67.46	65.12	61.93	0.7854	0.7859	0.7821	0.7737	7.48	8.12	8.15	3.35
SD + LoRA [14]	56.12	52.81	45.41	43.91	0.7698	0.7705	0.7617	0.7574	10.11	10.77	12.20	7.61
SD + Full [36]	50.72	47.35	41.72	39.41	0.7536	0.7616	0.7582	0.7574	11.83	11.98	13.74	8.31
TaxaDiffusion (ours)	**41.92**	**40.16**	**27.35**	**31.87**	**0.7303**	**0.7324**	**0.7349**	**0.7319**	**15.06**	**16.67**	**18.42**	**10.43**

Results & Performance

- Ablation of guidance:



Results & Performance

- Importance of progressive training:

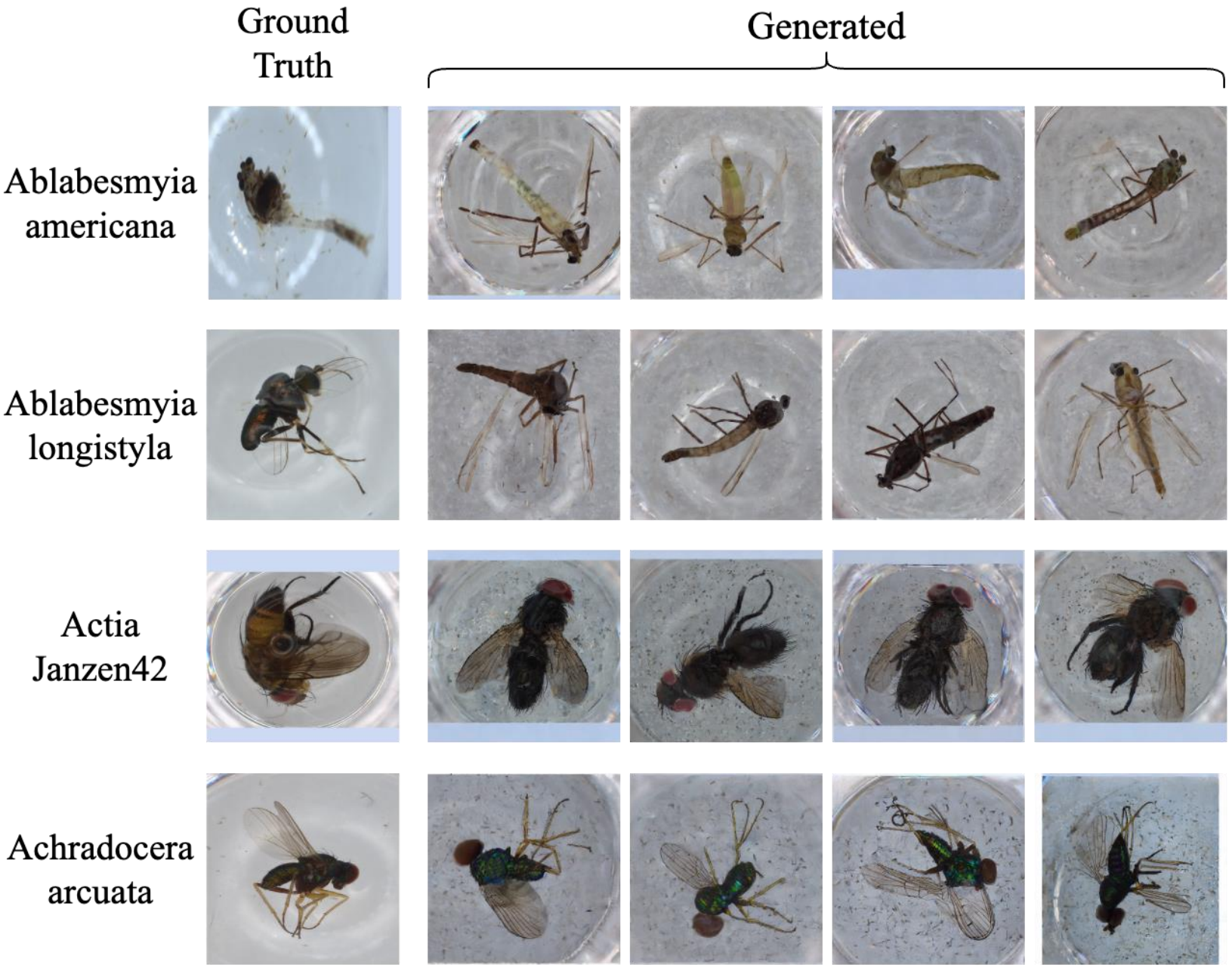
Strategy	FID ↓	LPIPS ↓	BioCLIP ↑
All	76.96	0.7273	5.66
Random	**72.31**	0.7178	6.68
Progressive (ours)	82.75	**0.7097**	**7.87**

Results

- iNaturalist:

Results

- BIOSCAN-1M:



Conclusion

- **Taxonomy-Guided Training**: Learns coarse-to-fine traits by progressing through Kingdom→…→Species.

- **Rare-Species Transfer**: Generates high-fidelity images with minimal samples.

Questions