Training-Free Synthetic Data Generation with Dual IP-Adapter Guidance

Luc Boudier*[1], Loris Manganelli*[1], Eleftherios Tsonis*[1], Nicolas Dufour[1 2], Vicky Kalogeiton[1]
[1] LIX, École Polytechnique, IP Paris, CNRS [2] LIGM, École des Ponts, IP Paris, CNRS, UGE

BMVC 2025

SCAN ME

Introduction and Related Works

- **Task**: Few-Shot Classification (classify images with <16 examples).
- **Importance**: Critical for data-scarce fields (e.g., medical, rare species).
- **Challenges**: Existing synthetic data methods are often:
  - **Complex**: Require heavy fine-tuning
  - **Brittle**: Rely on external captioning/filtering

Related Works

Method	Training-Free	Captioning/Filtering-Free
DataDream [Kim 2024 ECCV]	✗	✓
DISEF [da Costa 2023 arXiv]	✓	✗
DIPSY (Ours)	✓	✓

We introduce **DIPSY**, a training and captioning-free pipeline producing faithful and specific images using **double image guidance**.

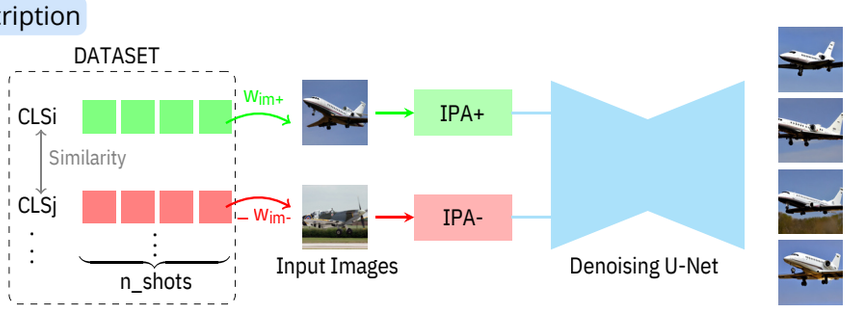
Method

Double Image Guidance: Steers towards the class description

$$\hat{\varepsilon}_\theta(x_t, c_{\text{text}}, c_{\text{im+}}, c_{\text{im-}}) = \varepsilon_\theta(x_t) + w_{\text{text}}\left(\varepsilon_\theta(x_t|c_{\text{text}}) - \varepsilon_\theta(x_t)\right) + w_{\text{im+}}\left(\varepsilon_\theta(x_t|c_{\text{text}}, c_{\text{im+}}) - \varepsilon_\theta(x_t|c_{\text{text}})\right) - w_{\text{im-}}\left(\varepsilon_\theta(x_t|c_{\text{text}}, c_{\text{im+}}, c_{\text{im-}}) - \varepsilon_\theta(x_t|c_{\text{text}}, c_{\text{im+}})\right).$$

Enforces target class features

Discourages features from other classes given by **similarity sampling**.



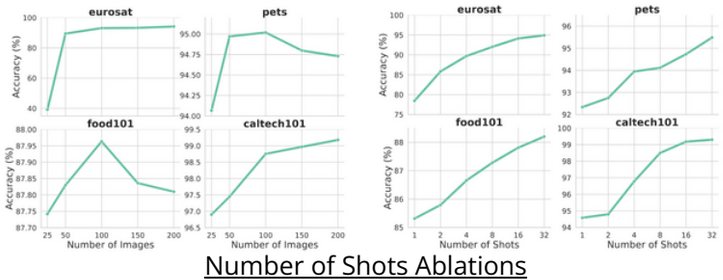
DIPSY Generation Pipeline

Quantitative Results

Method	Venue	R	S	Train Free	Captioning Filtering-free	DTD	Food	Pets	SUN	EuSAT	AirC	IN	FLO	Cars	CAL	Avg. Acc.
Visual Prompt Tuning [32]	ECCV22	✓				66.06	86.99	91.84	70.47	92.33	36.21	69.57	90.95	69.01	95.40	66.31
Text Prompt Tuning [65]	IJCV'22	✓				70.73	83.65	89.89	72.95	87.05	45.50	67.97	97.62	81.40	95.23	79.20
VPT + TPT	arXiv'23	✓				72.02	83.70	88.87	72.58	90.42	48.03	67.34	98.12	81.99	95.75	79.88
Classifier Tuning [26]	ICLR23	✓				73.64	87.28	92.81	76.16	87.13	46.73	73.41	86.52	82.55	96.01	80.22
DISEF [13]	arXiv'23	✓	✓	✓		75.36	87.11	94.33	77.43	94.31	63.89	73.94	98.85	88.63	96.94	85.08
DataDream [36]	ECCV24	✓	✓		✓	73.76	87.51	93.79	76.78	92.61	63.52	72.74	98.52	90.28	98.14	84.76
DIPSY (ours)		✓	✓	✓	✓	75.55	87.81	94.73	77.78	94.13	62.35	73.24	98.56	88.95	99.18	85.23

Number of Classes	DIPSY (ours)	DataDream
10	1s	36m
100	57s	6h
1000	8m	61h

Ablations



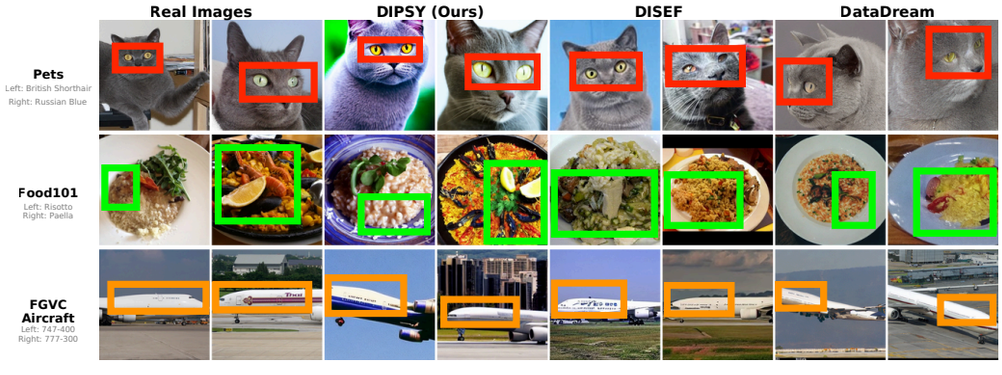
Number of Shots Ablations

Negative Image Prompt	Class Similarity	Data Augmentation	DTD	Pets	EuSAT	AirC
✗	✗	✗	75.02	94.72	93.44	59.70
✗	✗	✓	74.70	94.57	92.74	60.13
✗	✓	✗	Class Sim. only defined for negative.			
✗	✓	✓				
✓	✗	✗	75.22	94.79	93.36	61.54
✓	✗	✓	75.36	94.70	93.23	60.49
✓	✓	✗	74.93	94.77	93.49	61.55
✓	✓	✓	75.55	94.73	94.13	62.35

Components Ablation

Ablation Finding: The negative image prompt drives the largest performance gain. While more real data always helps, generating more than 200 synthetic images for 16 shots offers no further benefit.

Qualitative Results



DIPSY excels at capturing, **fine-grained details** (like pet eye color and food toppings) that other methods miss.

Conclusion: DIPSY is a simple image gen pipeline for few-shot learning and achieves SOTA results without fine-tuning, filtering or captioning.