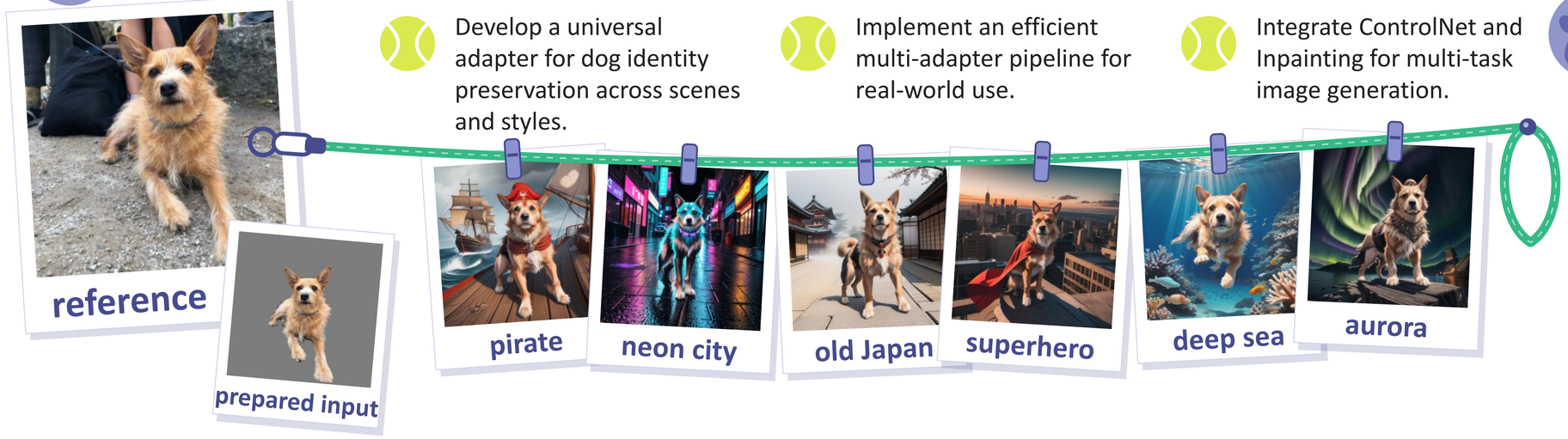
SINGLE-REFERENCE IDENTITY PRESERVATION FOR DOGS THROUGH LARGE-SCALE ADAPTER TRAINING

VYTAUTAS MAGNUS UNIVERSITY
Faculty of Informatics

AUTHORS: Hande Safiye Erdal · Eimantas Zaranka · Monika Katilienė · Tadas Žiemys · Justina Mandravickaitė · Tomas Krilavičius

1 GOAL



- Develop a universal adapter for dog identity preservation across scenes and styles.
- Implement an efficient multi-adapter pipeline for real-world use.
- Integrate ControlNet and Inpainting for multi-task image generation.

2 DATASET

- **1,255,805** images available in application
- **601,360** training images after filtering
- **273** unique breeds
- **92.1%** of eligible images pass through the pipeline

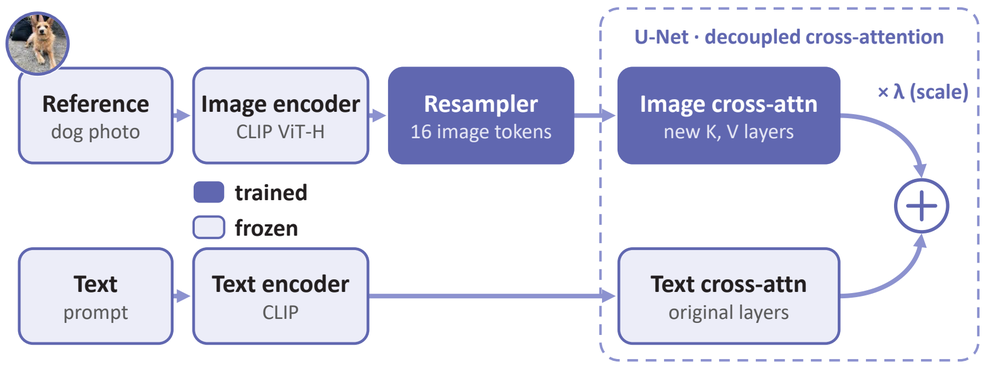
Data preparation

1. **Detection** — YOLOv8x + 11x
2. **Crop** — square 1024x1024
3. **Segmentation** — BiRefNet
4. **Mask cleaning** — blob filter
5. **Re-detection** — on the cutout
6. **De-duplication** — dHash
7. **Captioning** — Moondream2

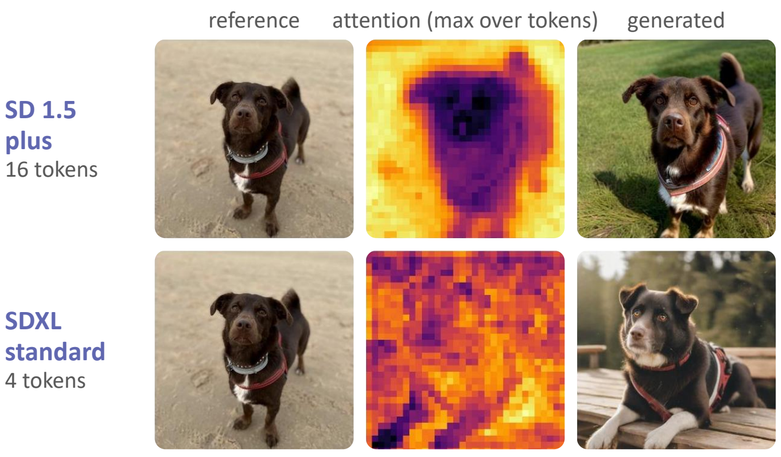


3 METHOD

Final run — SD 1.5 · 512×512 · no background · 5 epochs · LR 1e-4

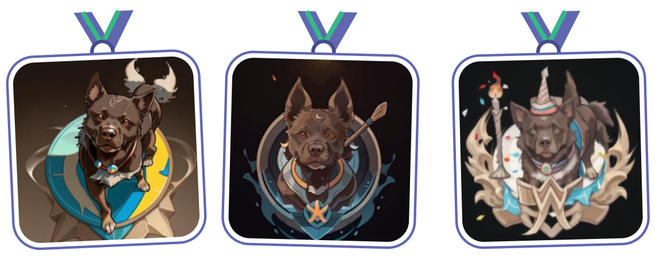


4 EXPLAINABILITY



5 CONTROLLED GENERATION

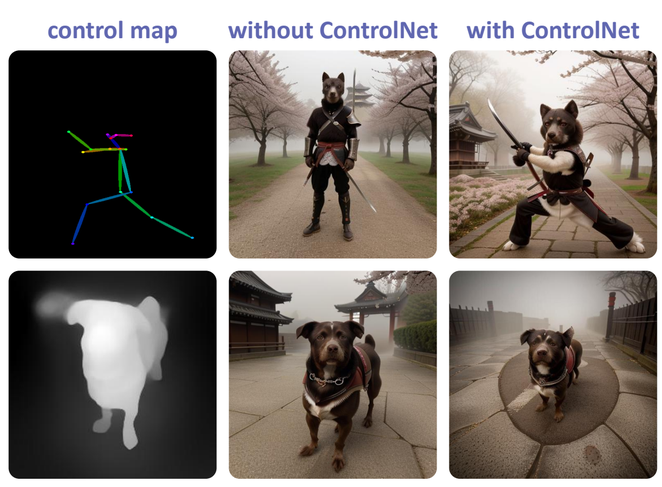
LoRA badges



Inpainting badges

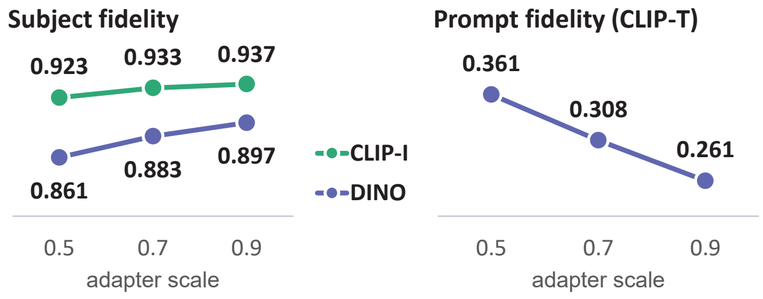


ControlNet



6 QUANTITATIVE EVALUATION

Method	DINO	CLIP-I	CLIP-T
Real images (DreamBooth)*	0.774	0.885	–
DreamBooth, Imagen	0.696	0.812	0.306
DreamBooth, Stable Diffusion	0.668	0.803	0.305
Textual Inversion, SD	0.569	0.780	0.255
This work, SD 1.5 plus, scale 0.5	**0.861**	**0.923**	**0.361**



*DreamBooth rows are from the original paper [1], not a direct comparison. [1] Ruiz, N. et al. (2023). DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. arXiv:2208.12242.

7 COMMERCIAL VS OURS



Prompt: "the dog in a (halloween costume)1.2, standing at a table, surrounded by decorations, flickering lights, dark room, eerie atmosphere"



8 CONCLUSIONS

- Trained a universal adapter on 601,360 segmented dog photos to preserve unseen dog identities from a single reference photo.
- Expert evaluation showed that segmented images with clean backgrounds and clear dog features produced the best results.
- At scale 0.5: DINO 0.861, CLIP-I 0.923, CLIP-T 0.361, outperforming published DreamBooth figures.
- Identity preservation remains possible when changing base models or integrating LoRA, ControlNet and Textual Inversion.

ACKNOWLEDGEMENTS

These results were obtained as part of the project "Development of Innovative Products at Dogo App UAB" (project No. 02-048-K-0006), co-funded by the European Union through the European Regional Development Fund.

Co-funded by the European Union