Appendix

A. Impact of the VAE

VAE	Acc_{pixel}	$Acc_{highpass}$	$Acc_{lowpass}$
-	73.0	51.6	64.5
SD1.5	70.5	37.6	63.1
SDXL	69.7	37.1	62.8
Flux	70.6	37.7	63.4

Table 1. We quantify the impact of the VAE by performing the reconstruction of our ImageNet subset and then training on it. An evaluation on unaltered real data reveals that the VAE accounts for a significant drop in accuracy in the high-frequency regime.

B. Additional Vision Tasks

In addition to classification, we test object detection and segmentation to determine whether the observed trend holds across other vision tasks. Tested models do not all support additional spatial ques, so we label generated images using the SAM3 model to obtain bounding boxes and segmentation masks. Due to computational constraints, we reuse the pixel-space images from the class-name case of the main experiment. To provide a fair comparison with real data, the same labelling method is applied to the ImageNet images.

For object detection, we follow torchvision [1] recipe and train object detection Faster-RCNN with Imagenet1k Resnet-50 backbone for 26 epochs with Adam optimizer and batch size of 16. For segmentation, we also follow the torchvision recipe for the DeepLabv3 model with Imagenet1k Resnet-50 backbone. We train the model using the Adam optimizer for 30 epochs with a batch size of 20.

B.1. Object Detection

	Test Scores		
Data Source	AP	AP_{50}	AP_{75}
ImageNet	20.2	47.6	14.1
SD V1.5 (Oct 2022)	11.9	32.1	6.2
SD V3.0 (Feb 2024)	11.0	28.7	6.4
Flux-dev (Aug 2024)	6.5	17.2	3.2
Lumina2 (Jan 2025)	6.1	17.3	3.2
Qwen-Image (Aug 2025)	3.1	8.2	1.7

Table 2. Performance of the Faster-RCNN model evaluated on the ImageNet validation subset labeled with SAM3 using the COCO evaluation protocol. We observe that the downward trend also applies to the object detection task, with the SD15 model outperforming recent state-of-the-art models.

[1] https://github.com/pytorch/vision/tree/6f131f1f56f1b78c6301eb4

B.2. Semantic Segmentation

	Test scores			
Data Source	Pixel Acc	mIoU	FWIoU	Dice Score
ImageNet	94.3	75.2	89.6	84.6
SD V1.5 (Oct 2022)	87.9	49.3	79.1	62.6
SD V3.0 (Feb 2024)	85.2	40.2	75.0	54.1
Flux-dev (Aug 2024)	78.2	27.9	66.5	40.6
Lumina2 (Jan 2025)	80.9	26.0	67.6	38.2
Qwen-Image (Aug 2025)	75.7	12.6	59.1	19.8

Table 3. Performance of the DeepLabv3 model evaluated on the ImageNet validation subset labeled with SAM3. We observe that the downward trend also applies to the semantic segmentation task, with the SD15 model outperforming recent state-of-the-art models.

C. Performance on Synthetic Test Set

	Real Test Set			Synth Test Set		
Data Source	Acc	AP	mIoU	Acc	AP	mIoU
Imagenet	73.0	20.2	47.6	-	-	-
SD1.5	45.5	11.9	49.3	79.6	25.5	71.2
SD2.1	30.5	-	-	51.6	-	-
SDXL	30.5	-	-	83.2	-	-
SDXL turbo	10.3	-	-	99.2	-	-
SD3.0	39.4	11.0	40.2	92.6	47.7	81.9
SD3.5 medium	38.9	-	-	94.6	-	-
SD3.5 large	28.2	-	-	77.6	-	-
SD3.5 turbo	15.0	-	-	97.6	-	-
Sana	22.1	-	-	97.3	-	-
Flux-dev	19.3	6.5	27.9	95.5	50.1	88.1
Flux-schnell	16.1	-	-	94.4	-	-
Lumina2	24.7	6.1	26.0	93.3	41.9	81.6
Qwen	9.9	3.1	12.6	97.9	69.5	93.7

Table 4. We evaluate the trained model on the same-sized synthetic validation set obtained from the same data source. We find that synthetic models generally achieve a much higher performance than on the real test set. This excludes bad fit as the source of the observed trend and instead strongly hints at a more easily class-separable data manifold. Accuracy (Acc) reported for the Resnet-50 network.

D. Scaling Behavior of Synthetic Data

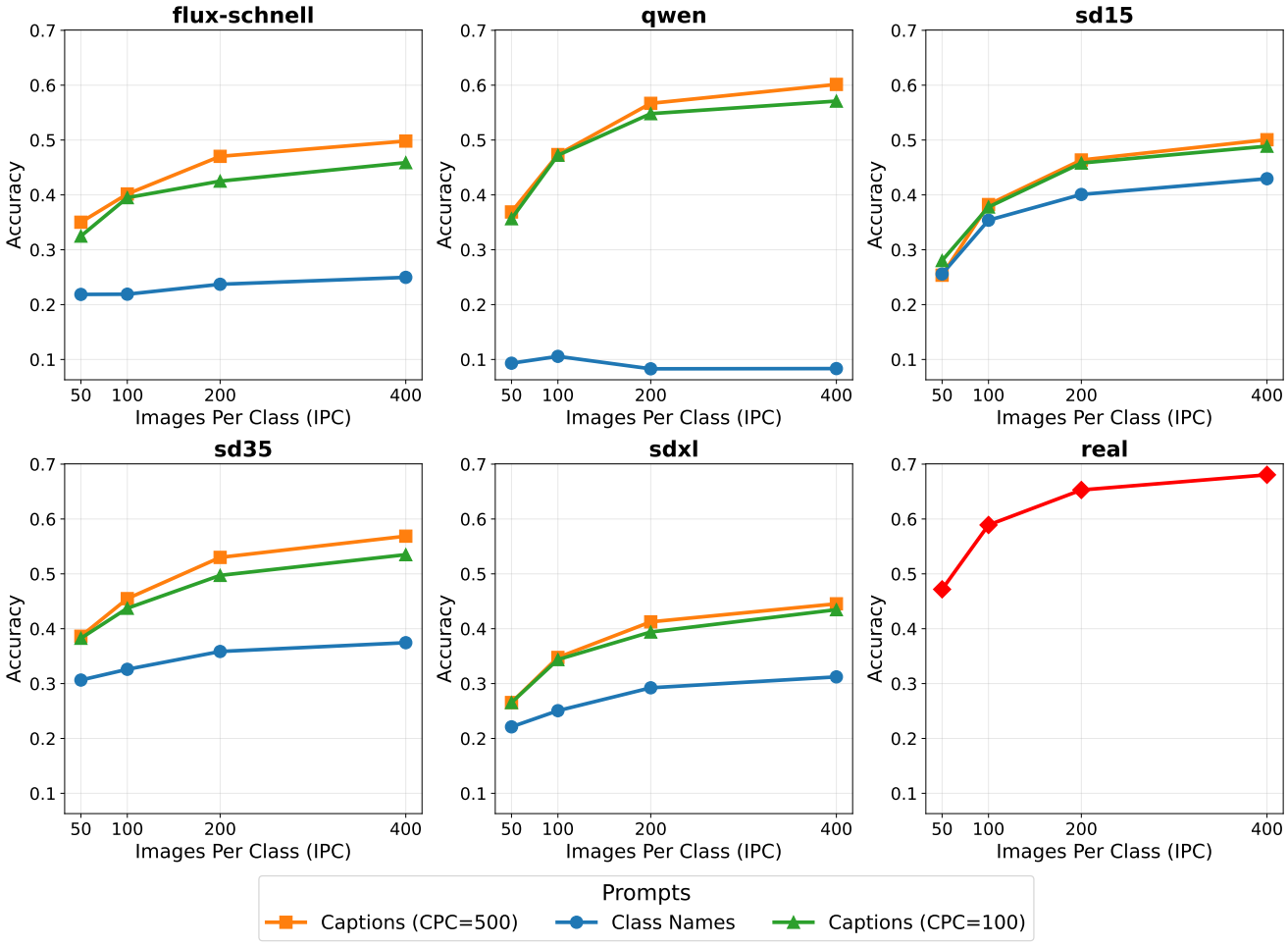
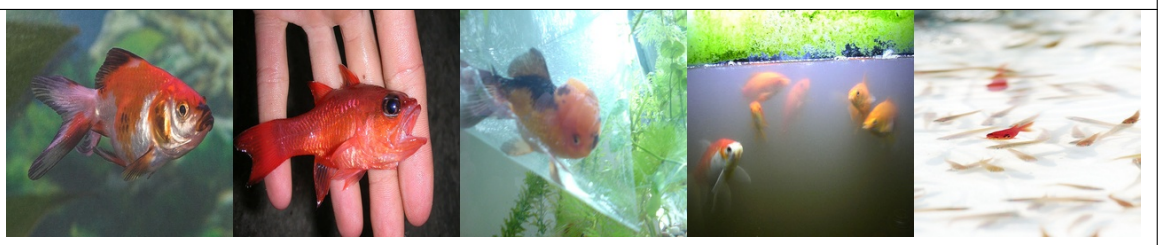
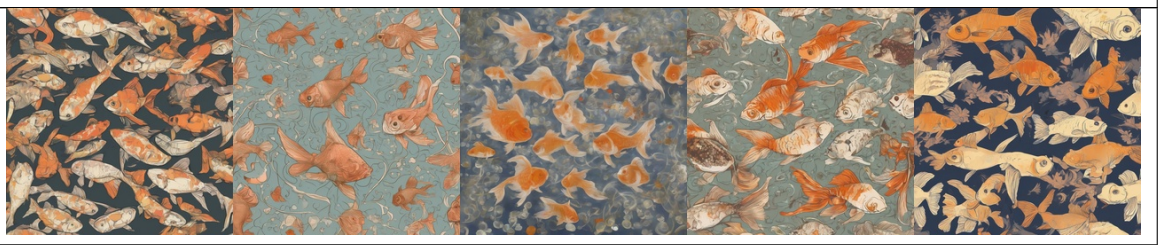


Figure 7. We investigate the scaling of the performance as a function of the number of Images per Class (IPC). We compare class-name prompts to image captions. We additionally test whether generating multiple images from the same caption impacts scaling. To test this, we also add a case in which the number of captions per class (CPC) is reduced to 100 (we generate 5 images per prompt). We observe that models which perform well with captions (qwen, flux-schnell, sd35). Exhibit a significant disparity between the scaling law of class name prompts and caption prompts. Additionally, we observe that when the number of unique prompts is reduced, those models also exhibit a bigger gap in scaling than models performing well with class names. This highlights the importance of using a diverse set of detail prompts for generation with recent models.

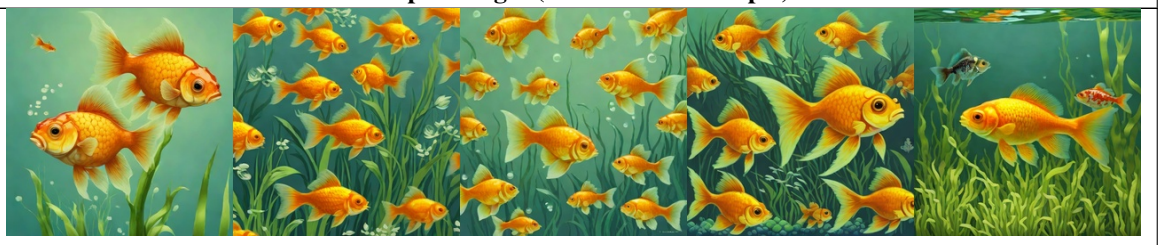
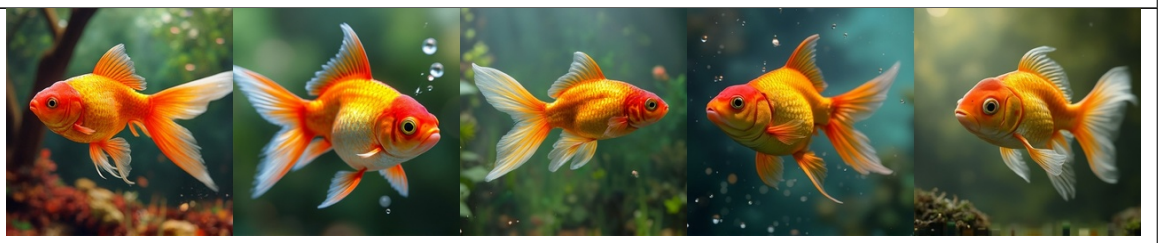
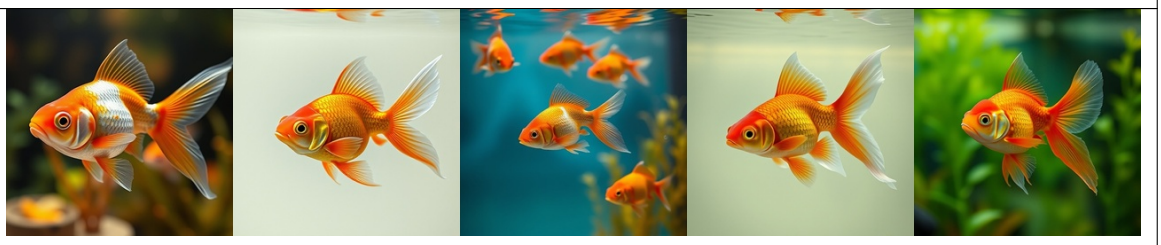
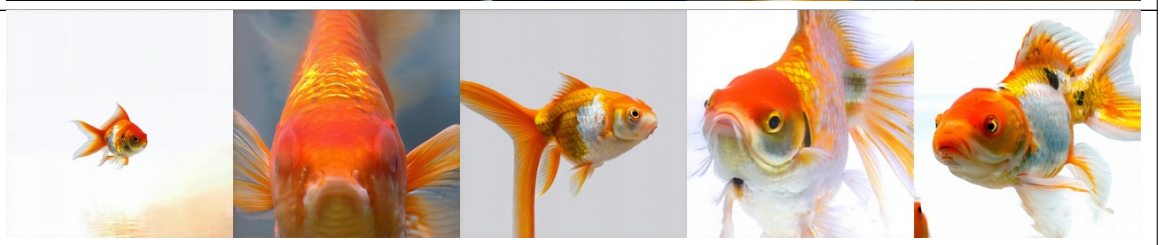
E. Synthetic Image Samples

Table 5. We visualize random samples from selected classes in the synthetic training dataset. Models are sorted by release date. Many of the tested models exhibit a distinct "visual style" that differs from real data. For example, select models (SDXL and pixelart) are biased towards stylistic, art-like images by default, rather than photorealistic images when the prompt is underspecified. Newer models tend to produce high-fidelity images; however, backgrounds often become blurry or plain, with the main object clearly visible and centered in the frame.

Data Source	Sample Images (Class Name Prompts)
	Goldfish
ImageNet	
sd15	
sd21	
sdxl	
pixelart	

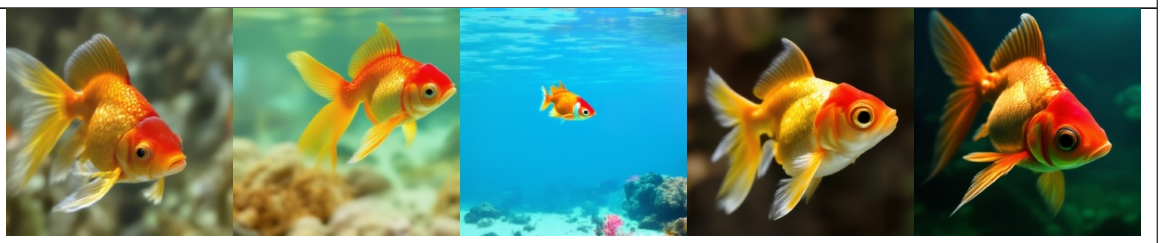
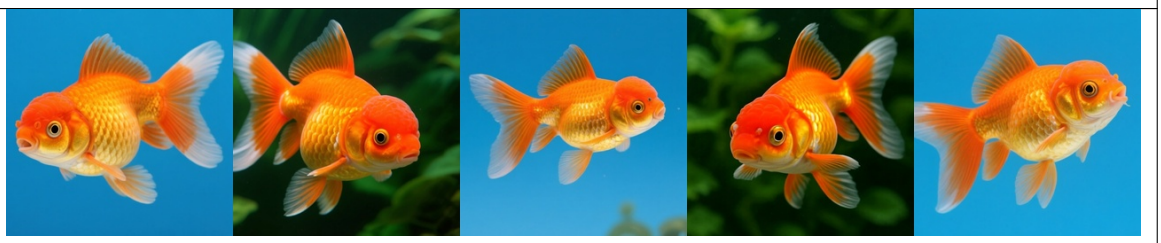
Continued on next page

Table 5 – *Continued from previous page*

Data Source	Sample Images (Class Name Prompts)
sdxl-turbo	
sd30	
flux-dev	
flux-schnell	
sd35	
sd35-large	

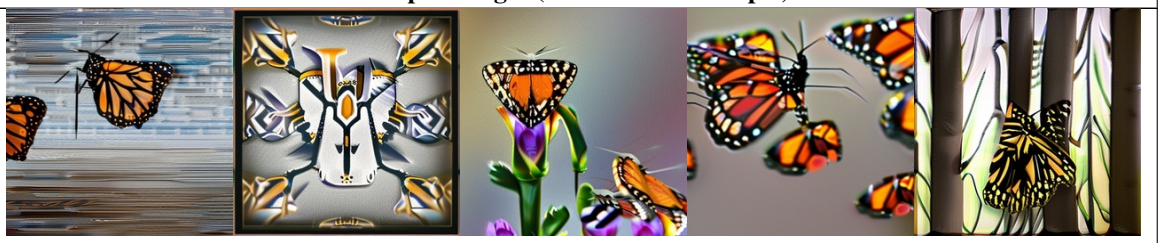
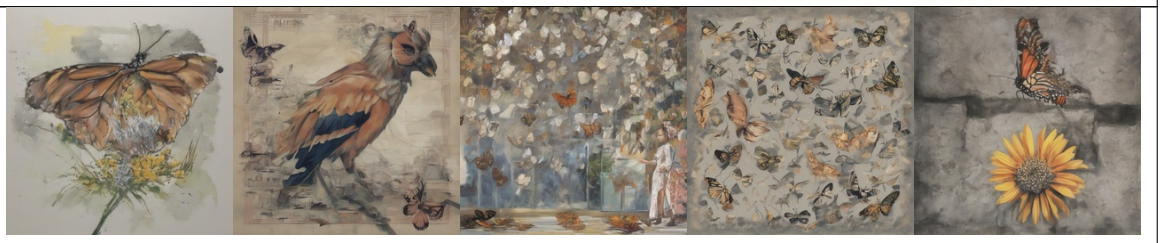
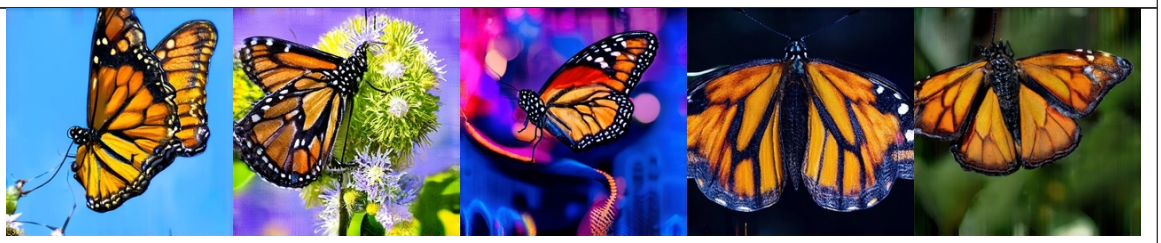
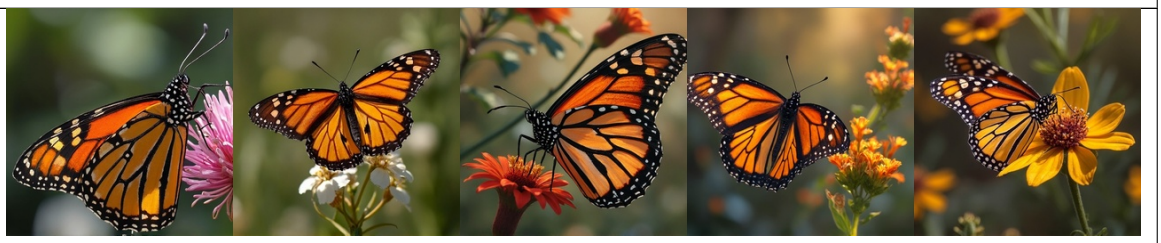
Continued on next page

Table 5 – *Continued from previous page*

Data Source	Sample Images (Class Name Prompts)
sd35-turbo	
sana	
lumina2	
qwen	
Monarch	
ImageNet	
sd15	

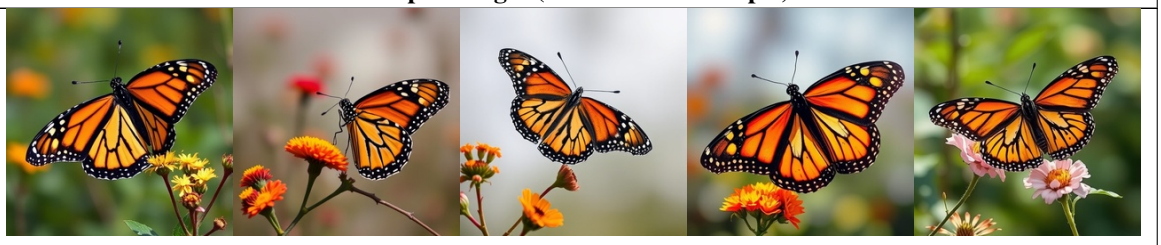
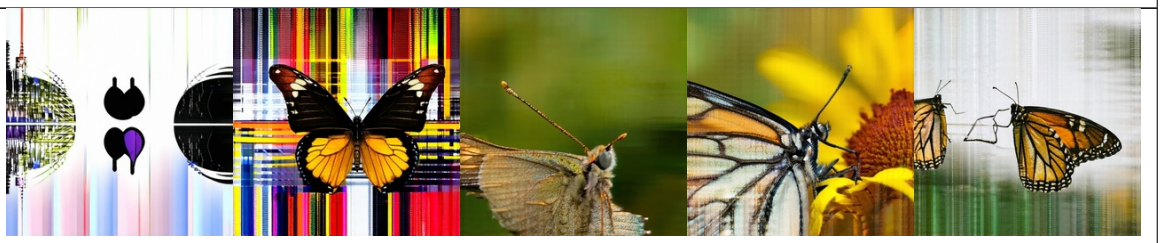
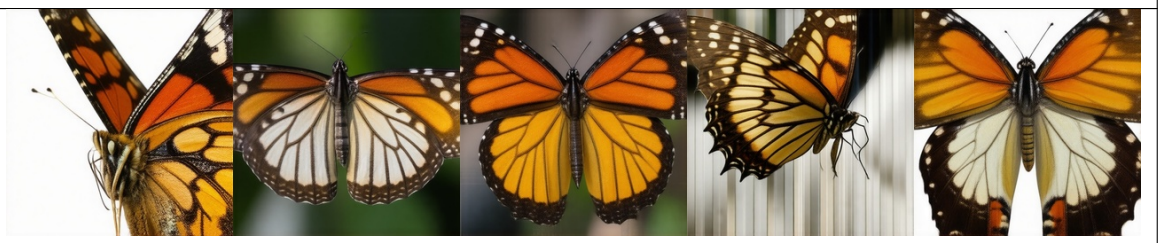
Continued on next page

Table 5 – *Continued from previous page*

Data Source	Sample Images (Class Name Prompts)
sd21	
sdxl	
pixelart	
sdxl-turbo	
sd30	
flux-dev	

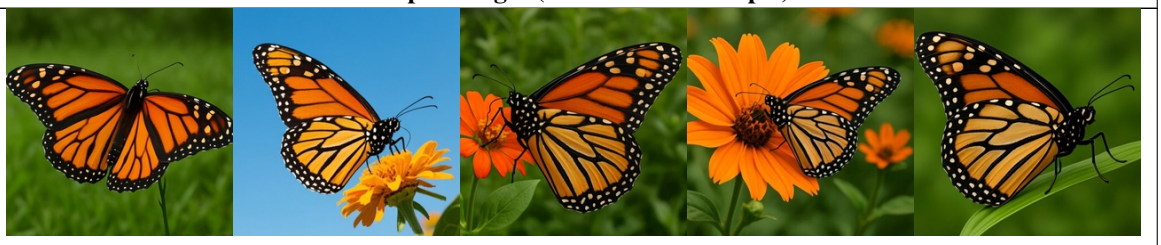
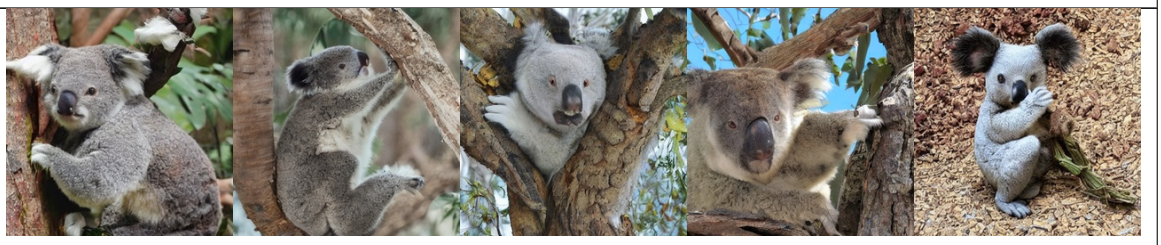
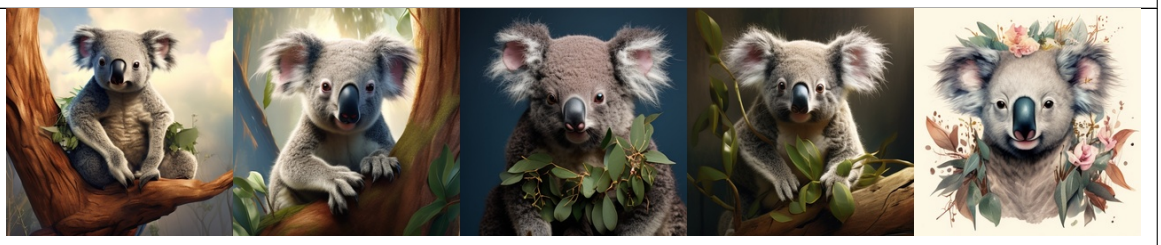
Continued on next page

Table 5 – *Continued from previous page*

Data Source	Sample Images (Class Name Prompts)
flux-schnell	
sd35	
sd35-large	
sd35-turbo	
sana	
lumina2	

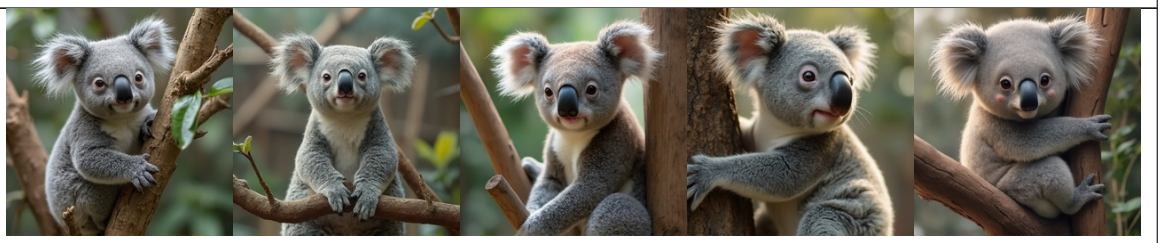
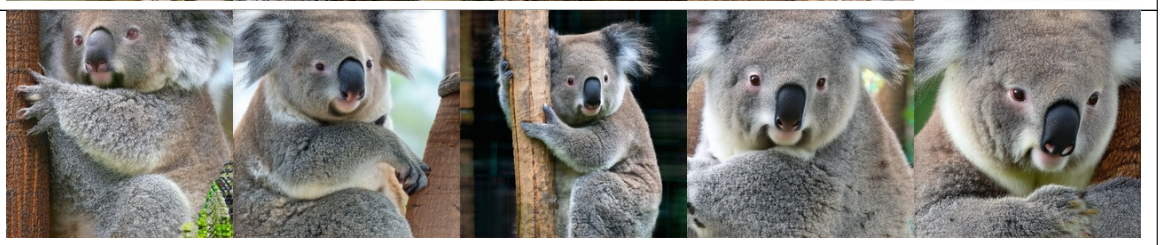
Continued on next page

Table 5 – *Continued from previous page*

Data Source	Sample Images (Class Name Prompts)
qwen	
Koala	
ImageNet	
sd15	
sd21	
sdxl	
pixelart	

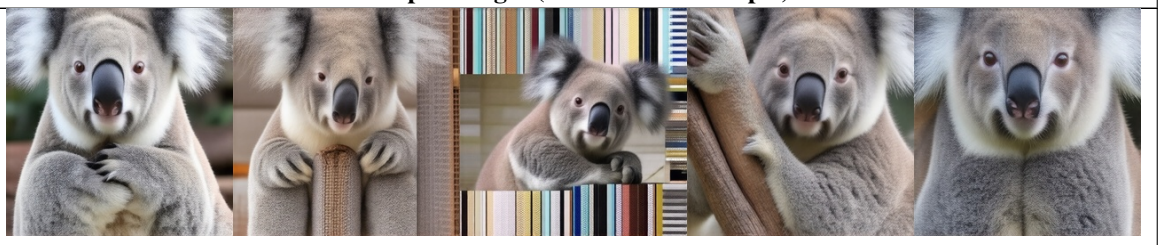
Continued on next page

Table 5 – *Continued from previous page*

Data Source	Sample Images (Class Name Prompts)
sdxl-turbo	
sd30	
flux-dev	
flux-schnell	
sd35	
sd35-large	

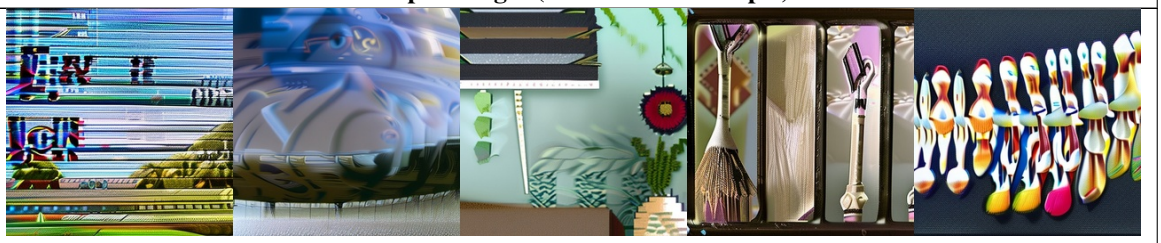
Continued on next page

Table 5 – *Continued from previous page*

Data Source	Sample Images (Class Name Prompts)
sd35-turbo	
sana	
lumina2	
qwen	
Broom	
ImageNet	
sd15	

Continued on next page

Table 5 – *Continued from previous page*

Data Source	Sample Images (Class Name Prompts)
sd21	
sdxl	
pixelart	
sdxl-turbo	
sd30	
flux-dev	

Continued on next page

Table 5 – *Continued from previous page*

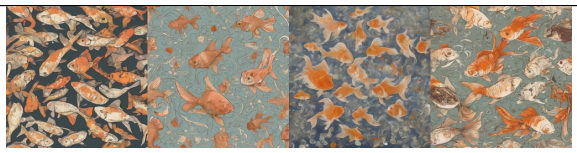
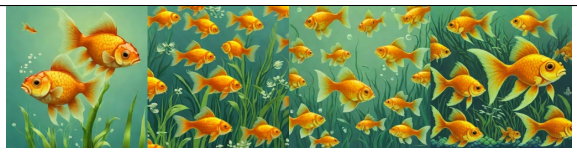
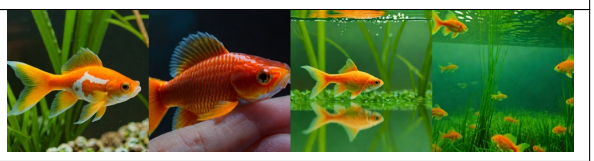
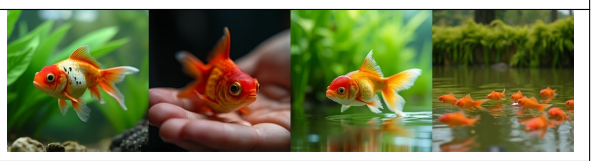
Data Source	Sample Images (Class Name Prompts)
flux-schnell	
sd35	
sd35-large	
sd35-turbo	
sana	
lumina2	

Continued on next page

Table 5 – *Continued from previous page*

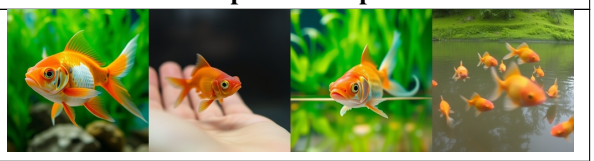
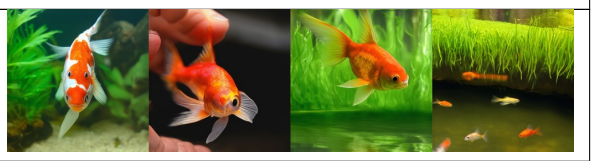
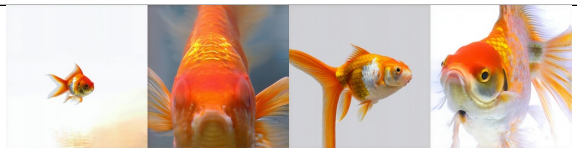
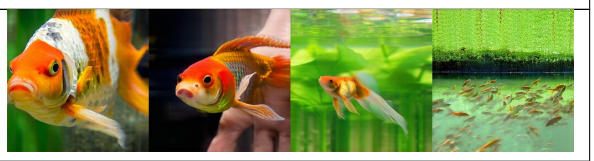
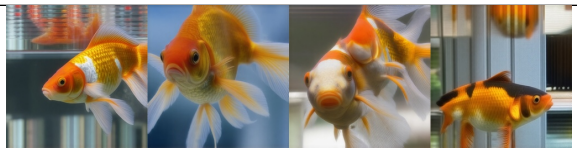
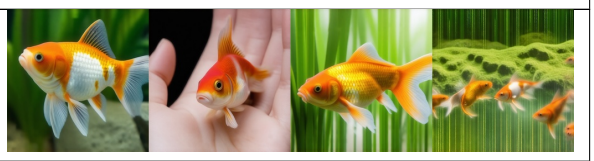
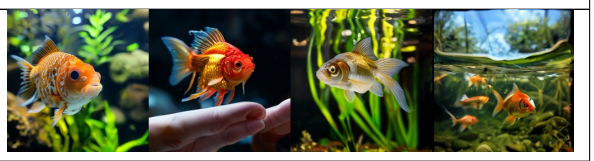
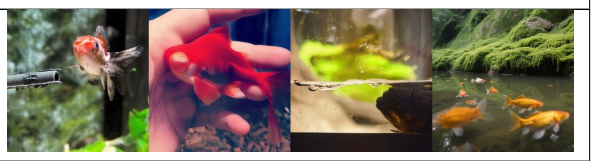
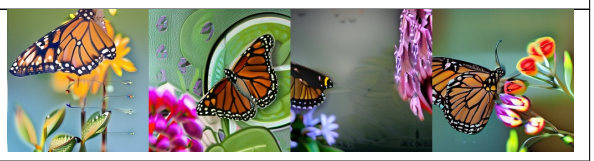
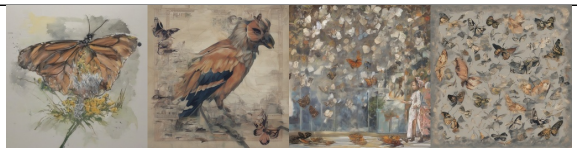
Data Source	Sample Images (Class Name Prompts)
qwen	

Table 6. We visualize samples generated from images with class name and caption prompts. We see that using detailed captions significantly improves the diversity and realism of the synthetic data, matching the observed boost in model performance.

Model	Class Name Prompt	Caption Prompt
	goldfish	
sd15		
sd21		
sdxl		
pixelart		
sdxl-turbo		
sd30		
flux dev		

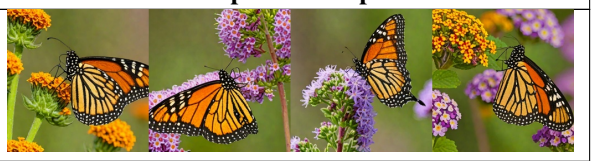
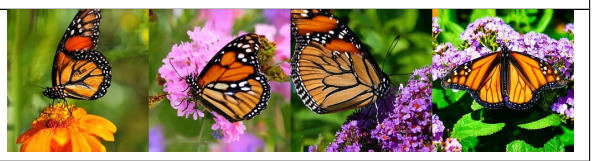
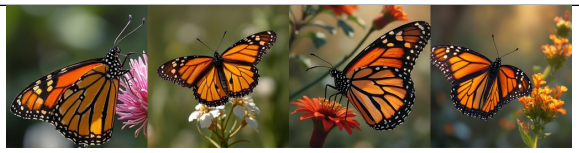
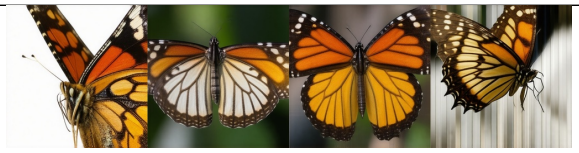
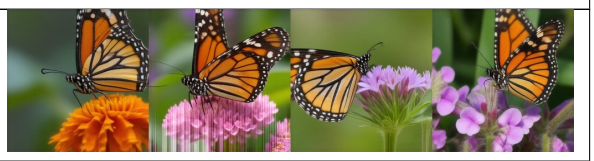
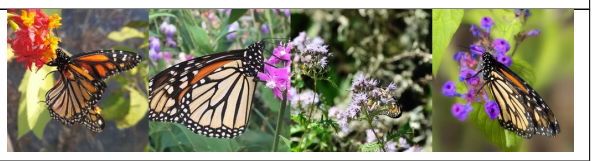
Continued on next page

Table 6 – *Continued from previous page*

Model	Class Name Prompt	Caption Prompt
flux-schnell		
sd35		
sd35-large		
sd35-turbo		
sana		
qwen		
monarch		
sd15		
sd21		
sdxl		
pixelart		

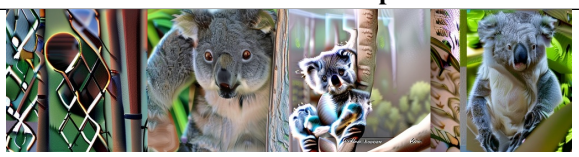
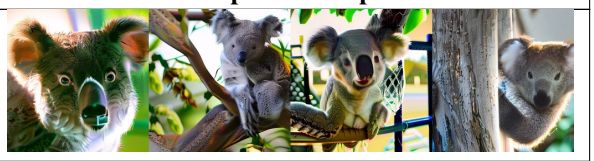
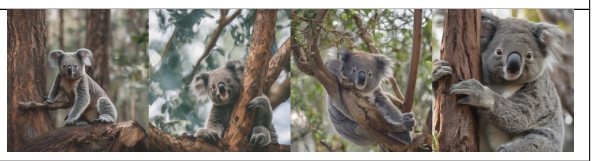
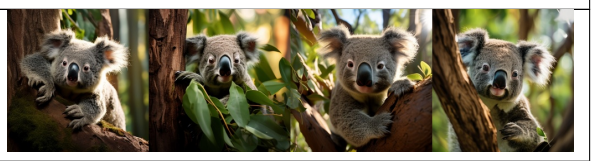
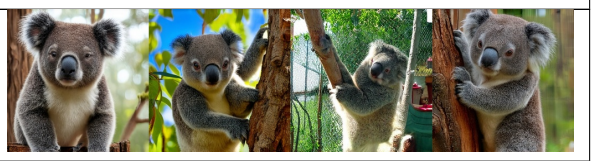
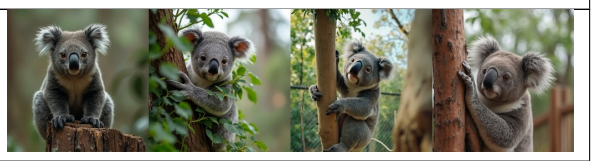
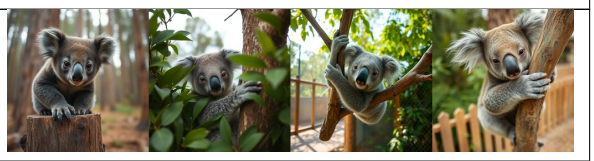
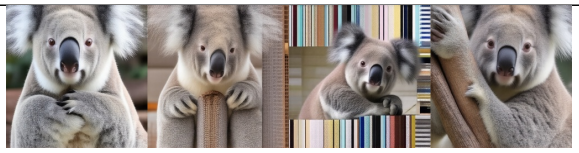
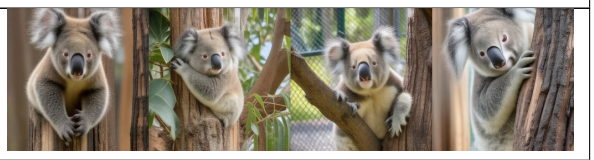
Continued on next page

Table 6 – *Continued from previous page*

Model	Class Name Prompt	Caption Prompt
sdxl-turbo		
sd30		
flux dev		
flux-schnell		
sd35		
sd35-large		
sd35-turbo		
sana		
qwen		
koala		
sd15		

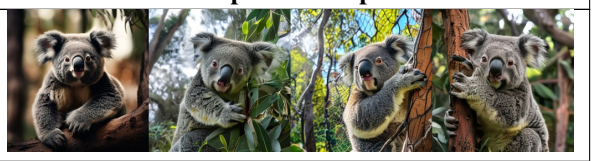
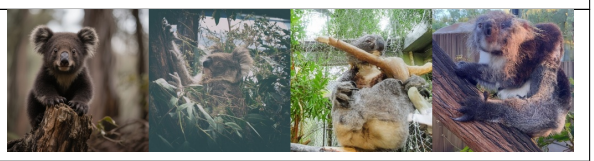
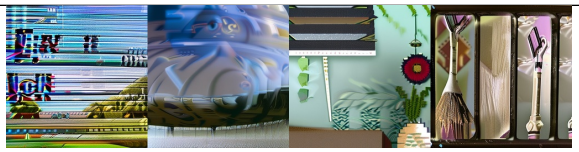
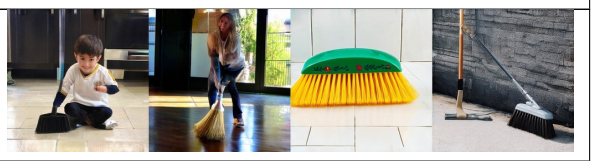
Continued on next page

Table 6 – *Continued from previous page*

Model	Class Name Prompt	Caption Prompt
sd21		
sdxl		
pixelart		
sdxl-turbo		
sd30		
flux dev		
flux-schnell		
sd35		
sd35-large		
sd35-turbo		

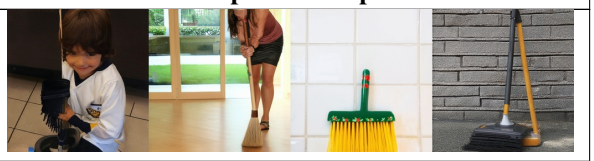
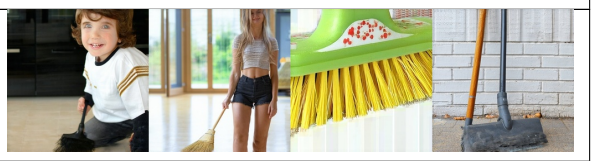
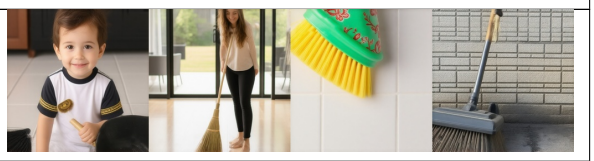
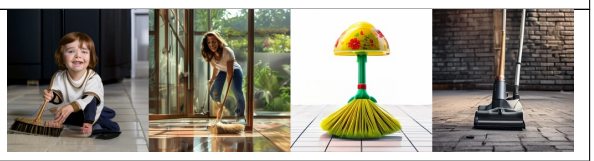
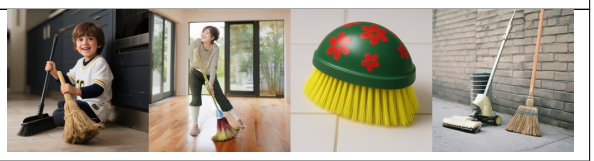
Continued on next page

Table 6 – *Continued from previous page*

Model	Class Name Prompt	Caption Prompt
sana		
qwen		
broom		
sd15		
sd21		
sdxl		
pixelart		
sdxl-turbo		
sd30		
flux dev		
flux-schnell		

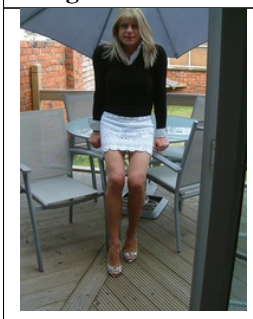
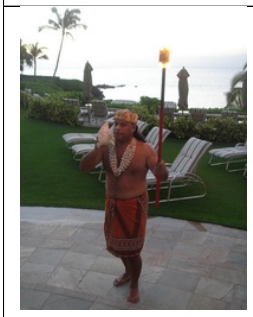
Continued on next page

Table 6 – *Continued from previous page*

Model	Class Name Prompt	Caption Prompt
sd35		
sd35-large		
sd35-turbo		
sana		
qwen		

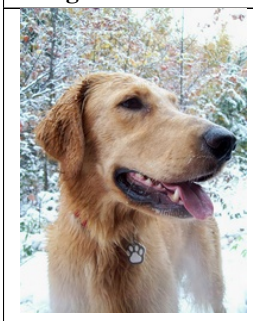
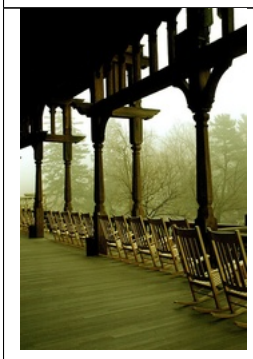
F. Image Caption Samples

Table 7. The table below contains sample captions that were used as prompts in the paper. Captions were obtained through prompting the GPT4-nano model.

Image	Caption
	The image features a woman standing on a wooden patio, framed by an open door in the foreground. She is holding a large umbrella against the backdrop of a brick house and a small garden area, with outdoor furniture and a grayish sky indicating an overcast day. The camera angle captures her from a straight-on perspective, focusing on her entire body and the surrounding outdoor setting.
	The image captures an indoor setting with stair steps viewed from a slightly elevated angle, focusing on the black metal handrail and the dark stairs below. In the background, a glass railing and the railing's supporting structure are visible, contributing to a modern architectural aesthetic. The lighting is subdued, emphasizing the metallic and glass materials in the scene.
	The image features a religious display set against a richly decorated interior with vertical, multicolored striped curtains and ornate wood paneling. In the foreground, a pedestal supports a statue of Jesus Christ on the cross, flanked by two golden lion-like sculptures with outstretched paws. The camera angle is slightly upward, emphasizing the statue and the intricate details of the backdrop.
	The image displays a dense layer of golden-brown pretzels filling the entire frame, creating a textured background. The foreground is dominated by the uniform, shiny pretzels, shot from a close-up, slightly overhead camera angle that emphasizes their shape and glossiness. There are no other distinct objects or elements visible in the background.
	A rugby match takes place outdoors on a lush green field, with a blurred background of trees indicating a park-like setting. In the foreground, two players are engaged in a tense tackle, with one player wearing a white shirt and gray shorts holding the rugby ball, while the other, in a red and black jersey, attempts to challenge him. The camera angle is slightly tilted and close-up, capturing the intensity of the moment and emphasizing the players' dynamic movements.
	The image features a lush, well-maintained garden area overlooking a body of water, with a few tall palm trees in the background. In the foreground, a traditionally dressed performer with a lei around his neck holds a lit torch or flame, standing on a paved surface. The camera angle captures him at eye level, emphasizing his cultural attire and the scenic outdoor setting behind him.

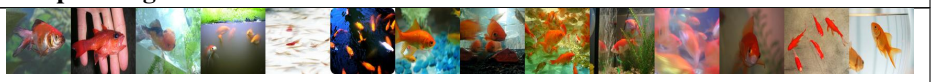
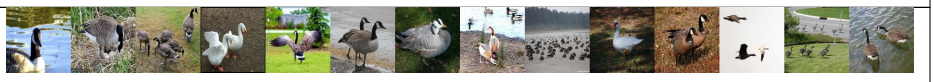
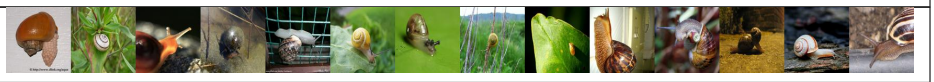
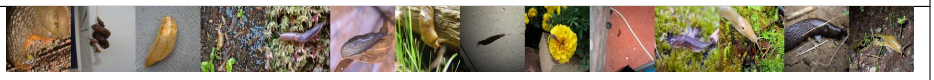
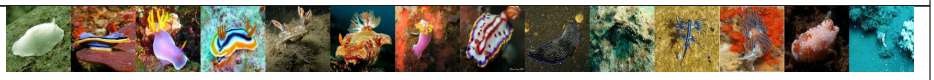
Continued on next page

Table 7 – *Continued from previous page*

Image	Caption
	The image features a close-up side profile of a golden retriever with a friendly expression, its mouth slightly open and tongue visible. The dog is wearing a collar with a paw print tag and is captured from a slightly lower angle, highlighting its head and upper body. In the background, there is a snowy landscape with trees and falling snowflakes, creating a wintery atmosphere.
	The image depicts a covered porch or veranda with a row of wooden rocking chairs arranged along the railing, facing outward. The background features a misty, foggy landscape with leafless trees, creating a serene and slightly mysterious ambiance. The camera angle is level with the chairs, capturing the perspective of someone standing on the porch looking toward the landscape.
	A young man is sitting on a boat or dock near the water, wearing a bright yellow life jacket. He is smiling and looking toward the camera, with an ocean or large body of water and an overcast sky in the background. The image is taken from a slightly low angle, capturing his upper body and part of his legs, with the water stretching out behind him.
	The image features a close-up view of a fire salamander with striking black and yellow coloration, positioned amidst lush green grass. The camera angle captures the salamander from a top-down perspective, highlighting its elongated body, glossy skin, and distinctive markings. The background consists of dense, vibrant grass blades, providing a natural and contrasting setting for the amphibian.

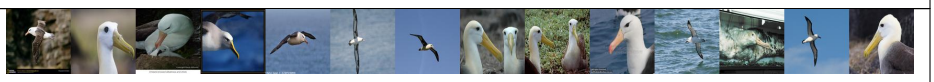
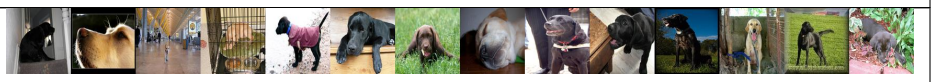
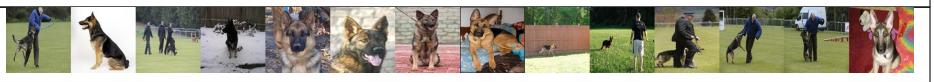
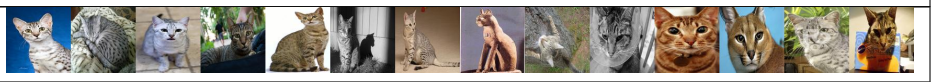
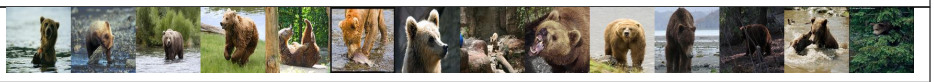
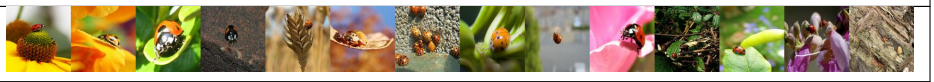
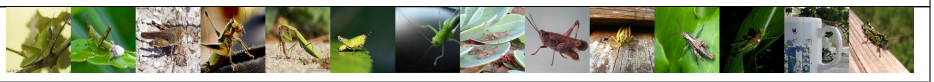
G. Imagenet-1k Subset

Table 8. We visualize image samples from the subset of the imagenet-1k (200 classes, 500 IPC) that we use as real data in our experiments

Class Name	Sample Images
goldfish	
european fire salamander	
bullfrog	
tailed frog	
american alligator	
boa constrictor	
trilobite	
scorpion	
black widow	
tarantula	
centipede	
goose	
koala	
jellyfish	
brain coral	
snail	
slug	
sea slug	
american lobster	

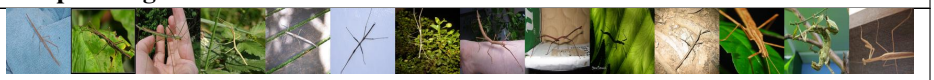
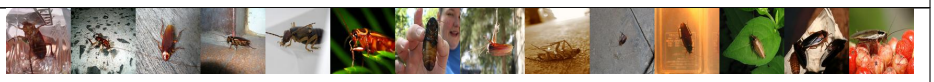
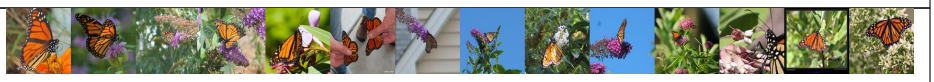
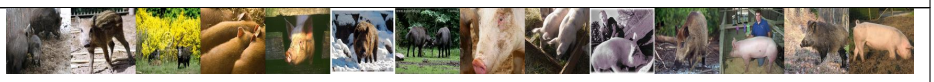
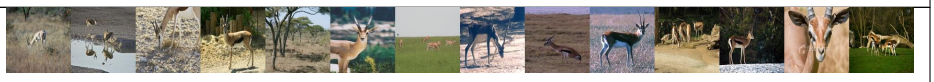
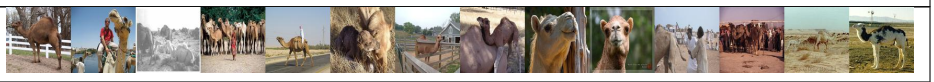
Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
spiny lobster	
black stork	
king penguin	
albatross	
dugong	
chihuahua	
yorkshire terrier	
golden retriever	
labrador retriever	
german shepherd	
standard poodle	
tabby cat	
persian cat	
egyptian cat	
cougar	
lion	
brown bear	
ladybug	
fly	
bee	
grasshoppe	

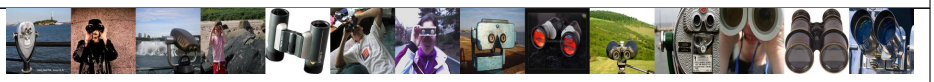
Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
walking stick	
cockroach	
mantis	
dragonfly	
monarch	
sulphur butterfly	
sea cucumber	
guinea pig	
hog	
ox	
bison	
bighorn	
gazelle	
arabian camel	
orangutan	
chimpanzee	
baboon	
african elephant	
panda	
abacus	
academic gown	

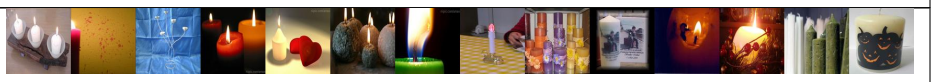
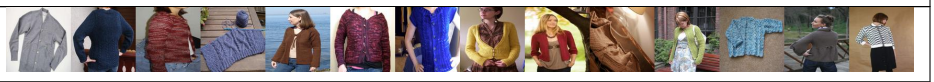
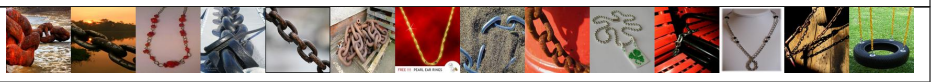
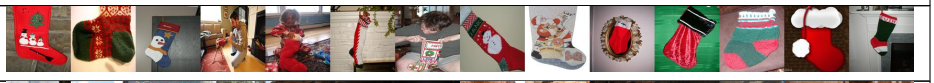
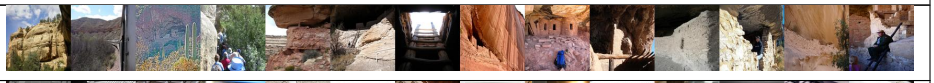
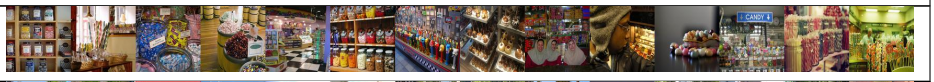
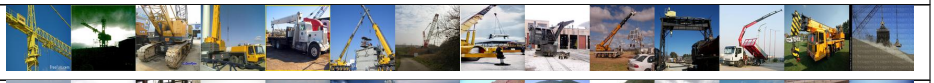
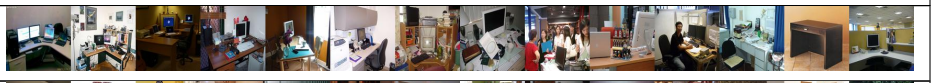
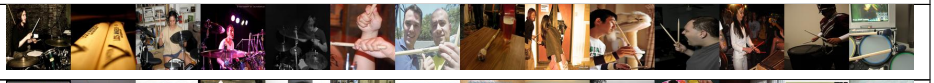
Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
altar	
apron	
backpack	
bannister	
barbershop	
barn	
barrel	
basketball	
bathtub	
beach wagon	
beacon	
beaker	
beer bottle	
bikini	
binoculars	
birdhouse	
bow tie	
brass	
broom	
bucket	
bullet train	

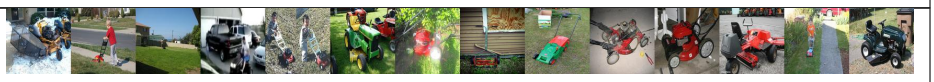
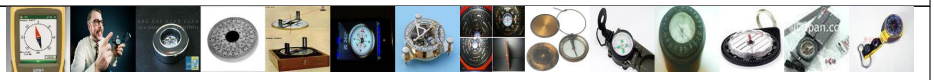
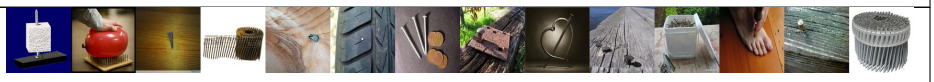
Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
butcher shop	
candle	
cannon	
cardigan	
cash machine	
CD player	
chain	
chest	
christmas stocking	
cliff dwelling	
computer keyboard	
confectionery	
convertible	
crane	
dam	
desk	
dining table	
drumstick	
dumbbell	
flagpole	
fountain	

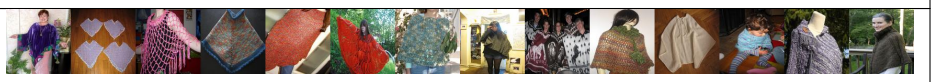
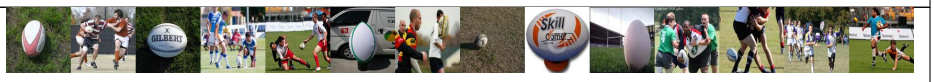
Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
freight car	
frying pan	
fur coat	
gasmask	
go-kart	
gondola	
hourglass	
iPod	
jinrikisha	
kimono	
lampshade	
lawn mower	
lifeboat	
limousine	
magnetic compass	
maypole	
military uniform	
miniskirt	
moving van	
nail	
neck brace	

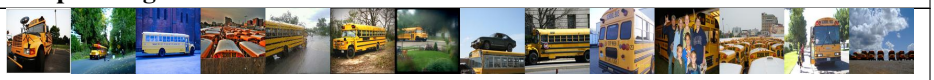
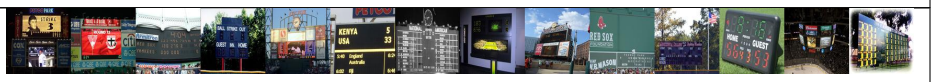
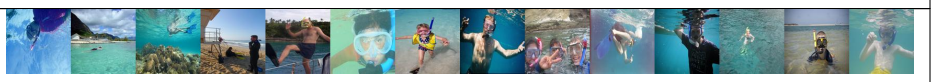
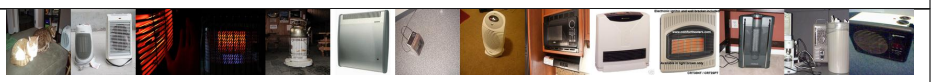
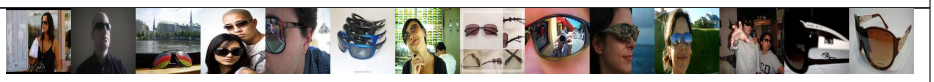
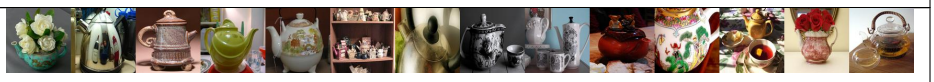
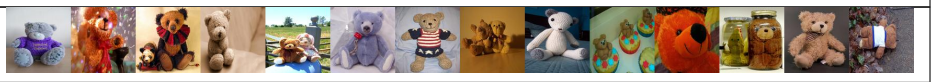
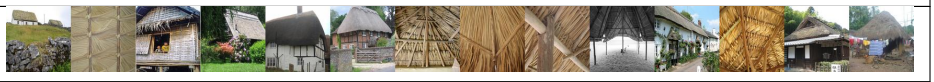
Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
obelisk	
oboe	
pipe organ	
parking meter	
pay-phone	
picket fence	
pill bottle	
plunger	
pole	
police van	
poncho	
pop bottle	
potter's wheel	
projectile	
punching bag	
reel	
refrigerator	
remote control	
rocking chair	
rugby ball	
sandal	

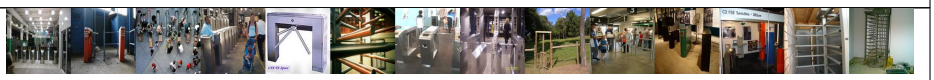
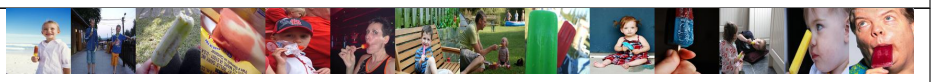
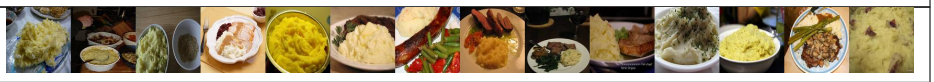
Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
school bus	
scoreboard	
sewing machine	
snorkel	
sock	
sombrero	
space heater	
spider web	
sports car	
steel arch bridge	
stopwatch	
sunglasses	
suspension bridge	
swimming trunks	
syringe	
teapot	
teddy bear	
thatched roof	
torch	
tractor	
triumphal arch	

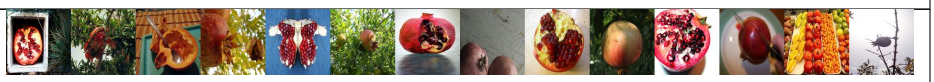
Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
trolleybus	
turnstile	
umbrella	
vestment	
viaduct	
volleyball	
water jug	
water tower	
wok	
wooden spoon	
comic book	
plate	
guacamole	
ice cream	
lollipop	
pretzel	
mashed potato	
cauliflower	
bell pepper	
mushroom	
orange	

Continued on next page

Table 8 – *Continued from previous page*

Class Name	Sample Images
lemon	
banana	
pomegranate	
meat loaf	
pizza	
potpie	
espresso	
alp	
cliff	
coral reef	
lakeside	
seashore	
acorn