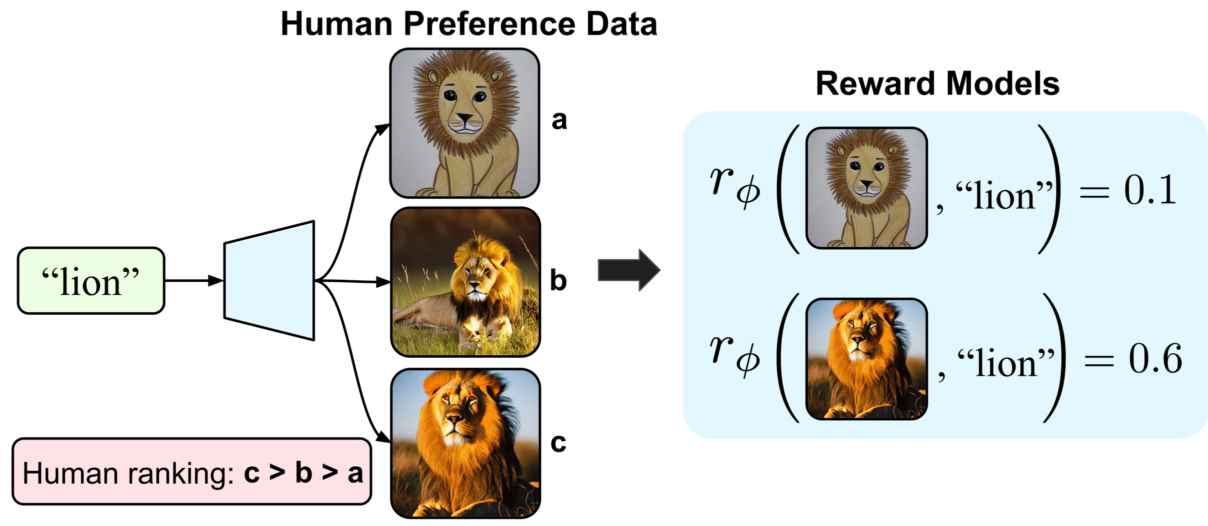
# Directly Fine-Tuning Diffusion Models on Differentiable Rewards

Kevin Clark*, Paul Vicol*, Kevin Swersky, David J. Fleet
Google DeepMind, *Equal Contribution

## Motivation




- Modeling the data distribution exactly often does not align with desired behavior (e.g., generating images with certain aesthetic qualitites) → motivates fine-tuning.
- Human preference data can be used to train reward models parameterized by neural networks.

## Overview

- Direct Reward Fine-Tuning (DRaFT) is a simple and effective method for fine-tuning diffusion models to maximize differentiable reward functions, such as scores from human preference models.
- **Key idea:** Use gradient-based optimization to maximize differentiable rewards:

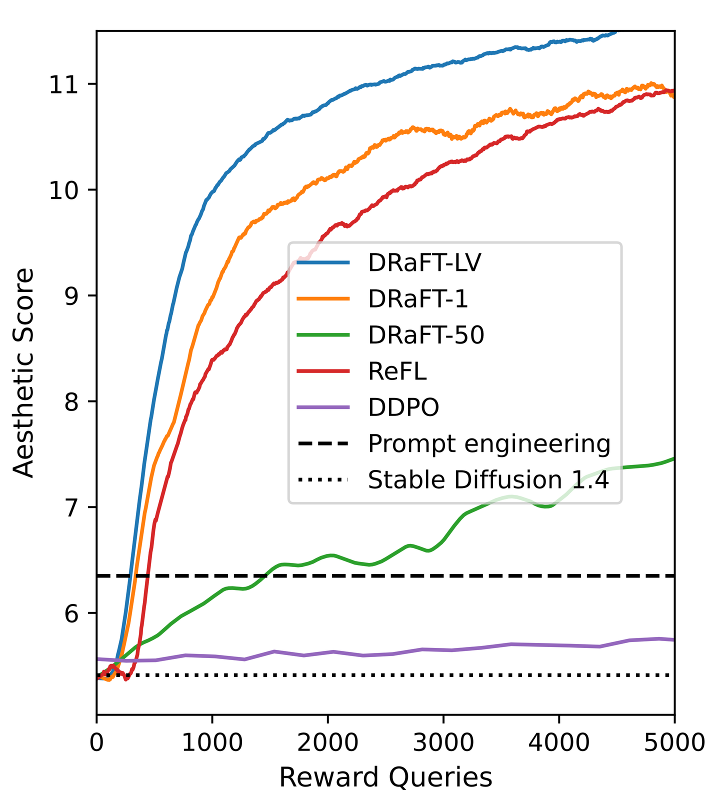
$$J(\theta) = \mathbb{E}_{\mathbf{c} \sim p_\mathbf{c}, \mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[ r(\text{sample}(\theta, \mathbf{c}, \mathbf{x}_T), \mathbf{c}) \right]$$

### DRaFT Ingredients

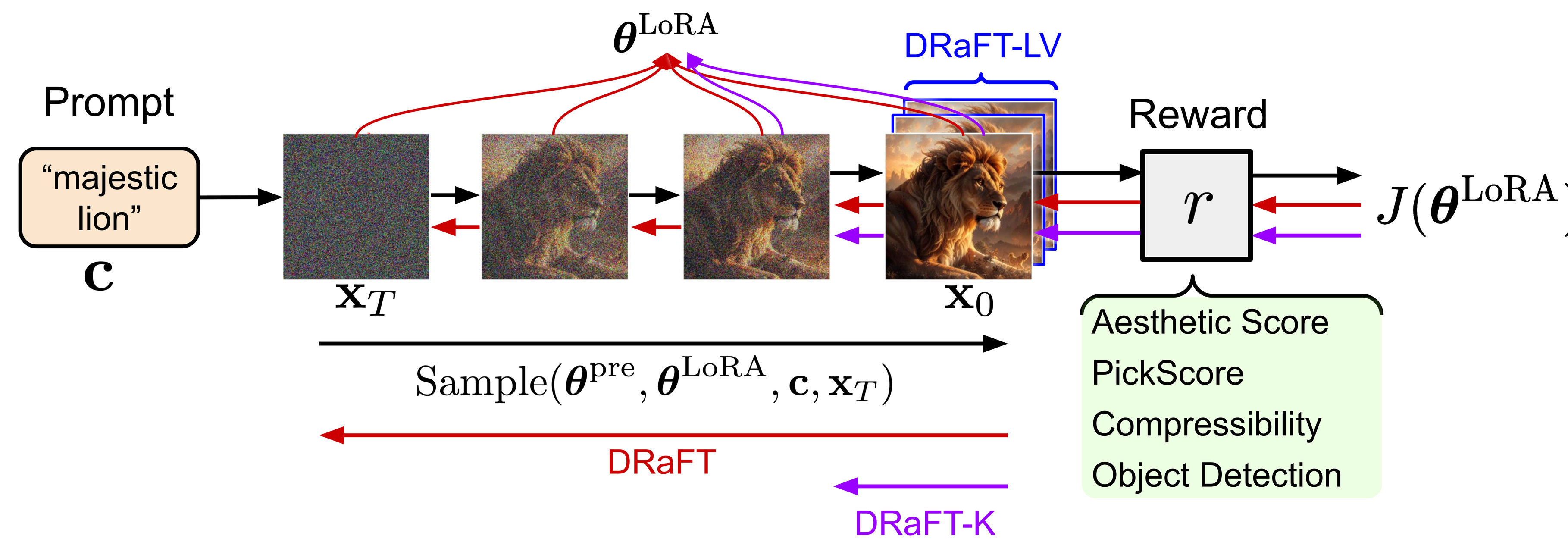
1. **LoRA:** Reduces the # of parameters to optimize, reducing the memory cost of fine-tuning.
2. **Gradient Checkpointing:** store the input latent and re-materialize the UNet activations during backprop.
3. **Truncated Backpropagation Through Sampling**

## Fine-Tuning for Aesthetic Quality

- Sample efficiency of methods maximizing scores from the LAION Aesthetic Classifier.
- DRaFT is significantly more sample-efficient than RL-based methods.



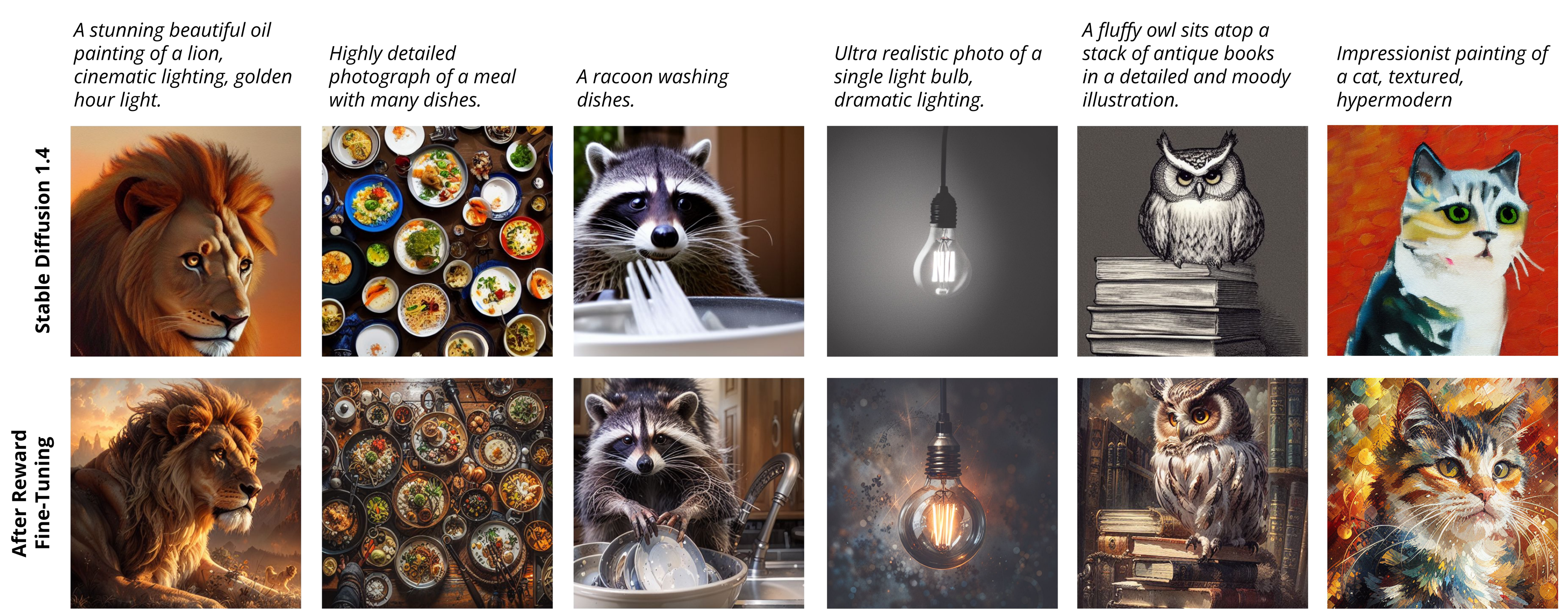

## Direct Reward Fine-Tuning (DRaFT)




- **Vanilla DRaFT:** Backpropagate through the full 50-step sampling chain.
- **DRaFT-$K$:** Backpropagate through the last $K$ steps of sampling.
- **DRaFT-LV:** Variant of DRaFT-1 with lower-variance gradient estimates. Noise the generated image $n$ times, and sum the reward gradient over these examples.
  - Using $n = 2$ is around 2× more efficient than DRaFT-1 while adding ∼ 10% overhead.

## Improving the Aesthetic Quality of Stable Diffusion

- DRaFT yields more detailed and stylized images than baseline Stable Diffusion

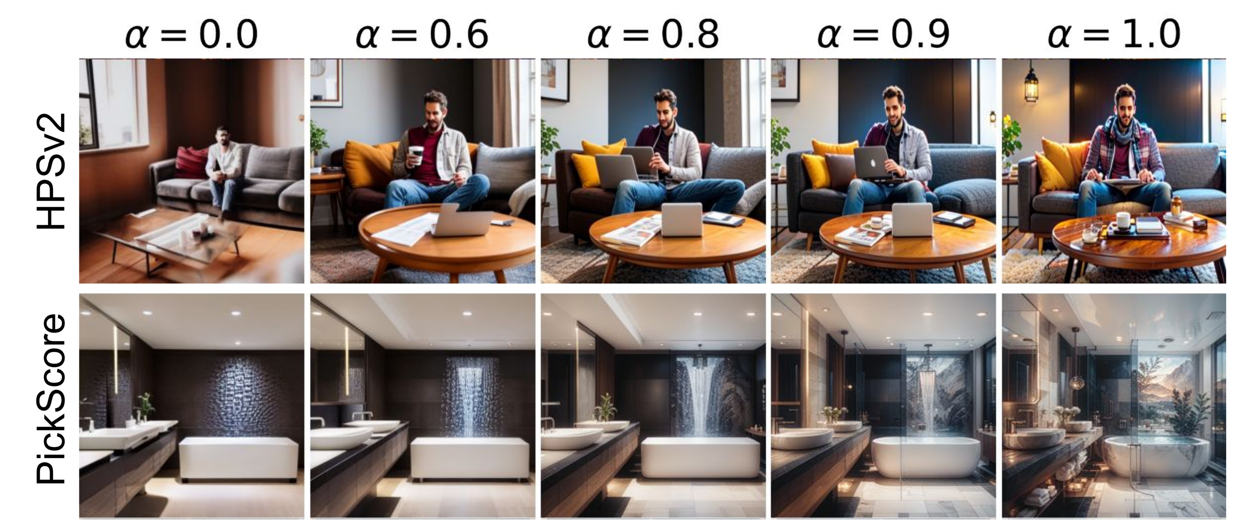



## Qualitative Comparison of Reward Functions

- Comparison of generations from DRaFT models fine-tuned on the LAION Aesthetic Classifier, Human Preference Score v2, and PickScore reward functions.


## LoRA Scaling & Interpolation

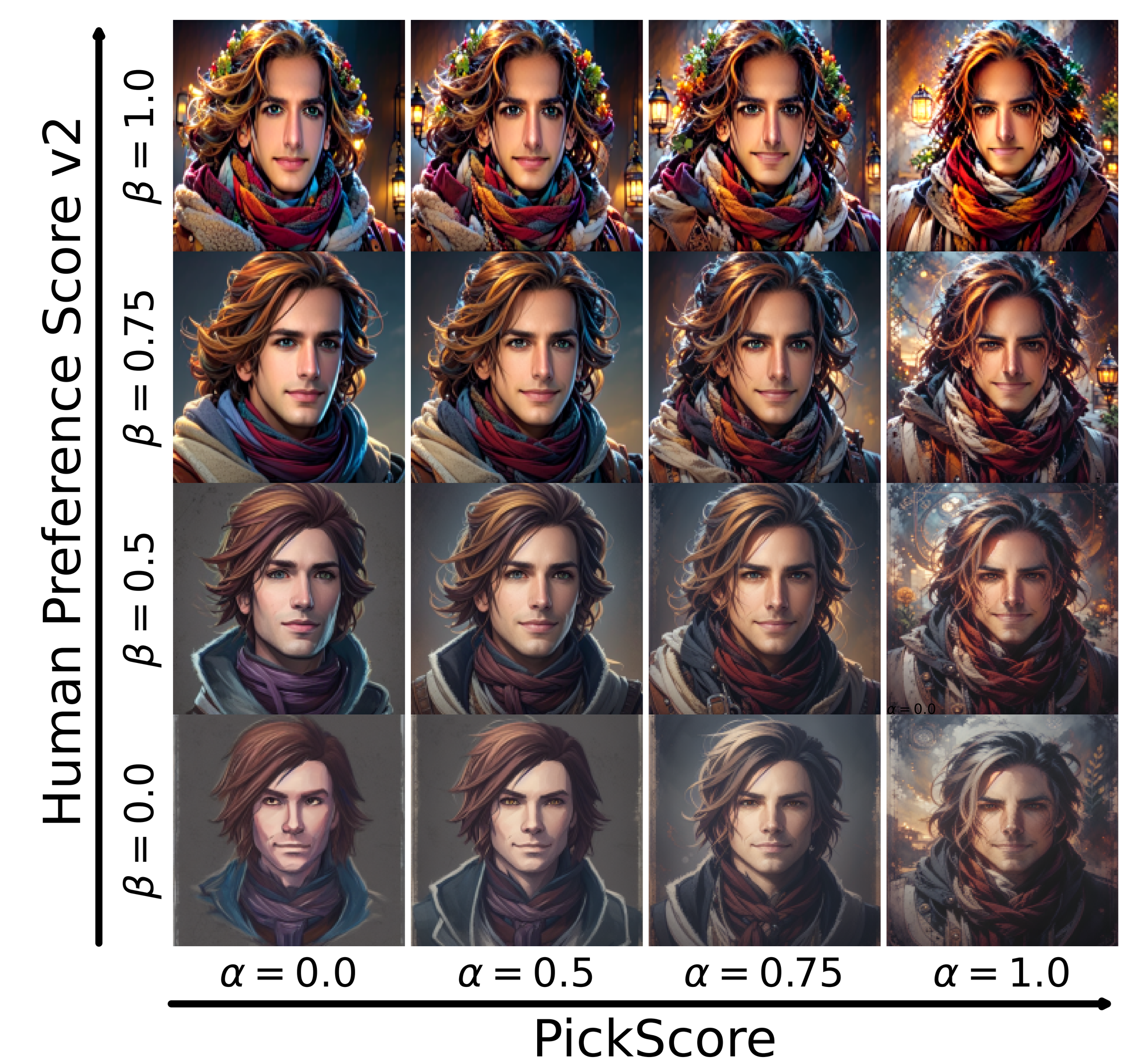
- Scaling LoRA parameters yields semantic interpolations between the pre-trained and fine-tuned models




## Mixing LoRA Parameters

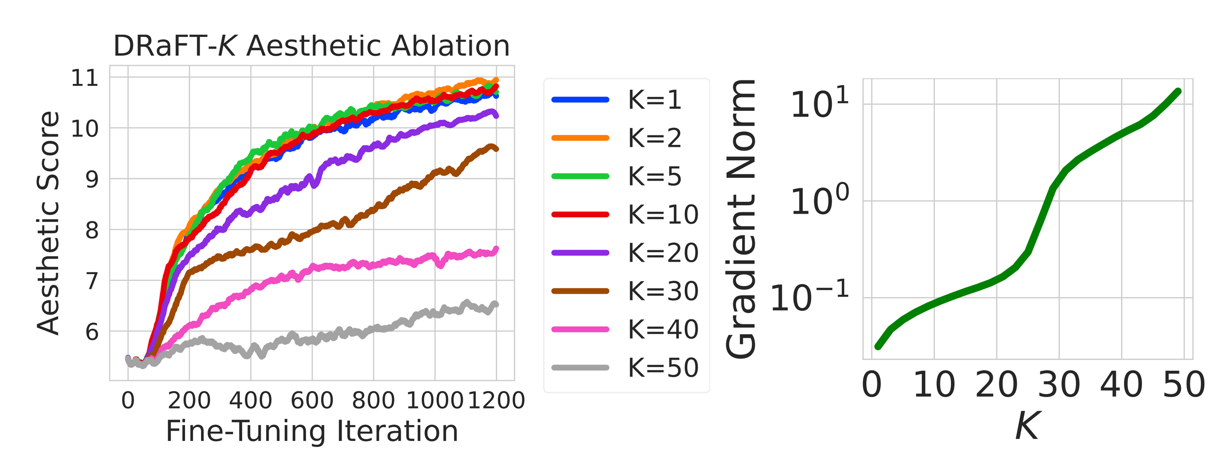
- We can combine the effects of different rewards by taking linear combinations of LoRA parameters

$$\theta^{\text{pre}} + \alpha \theta^{\text{PickScore}}_{\text{LoRA}} + \beta \theta^{\text{HPSv2}}_{\text{LoRA}}$$

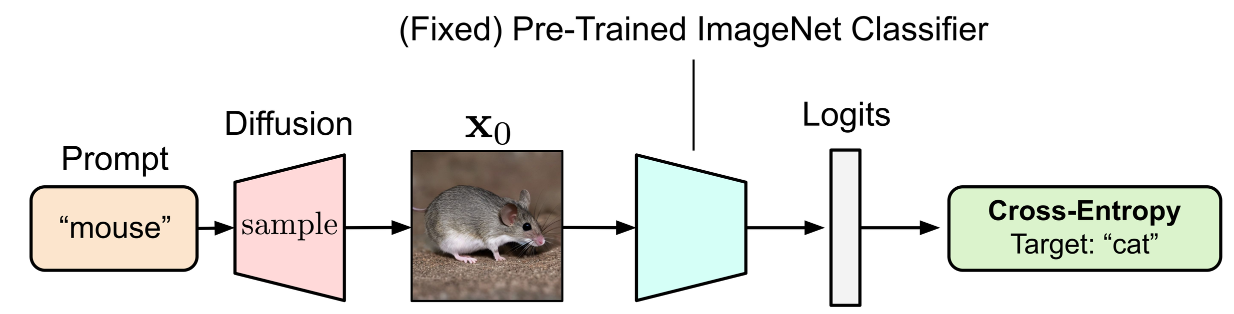



## Impact of $K$ in DRaFT-$K$

- Gradients tend to explode as $K$ increases.
- Even with small $K$, LoRA params affect all sampling steps.

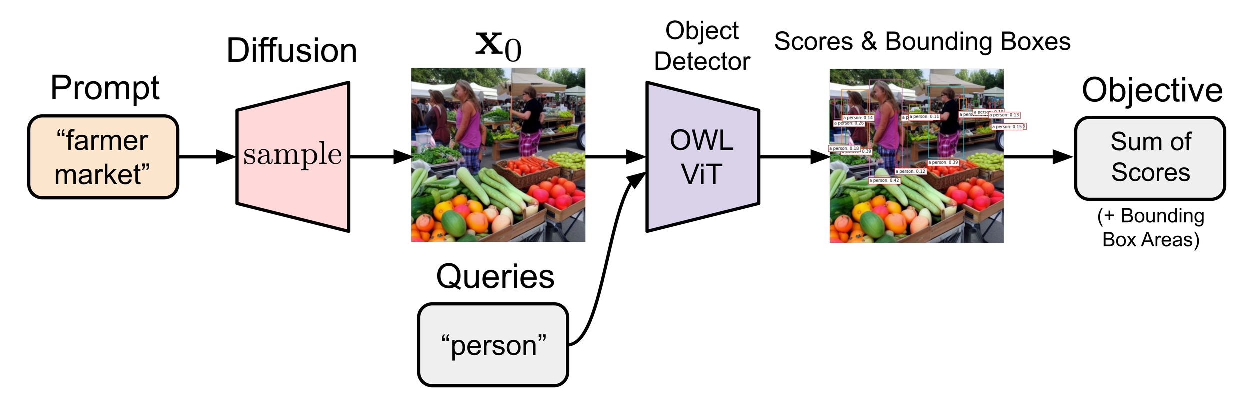



## Diffusion Adversarial Examples




- Fine-tune a diffusion model such that images generated based on a prompt for a class $y$ (e.g., "mouse") are classified as a different class $y'$ (e.g., "cat") by a ResNet-50 pretrained on ImageNet.
- The classifier is texture-biased: fine-tuned images have cat textures while keeping the animal shapes intact.


## Object Detection & Addition/Removal




- We can maximize or minimize scores from an object detection model to add or remove certain object classes.

**Object Addition:** Fruit Bowl + Strawberries