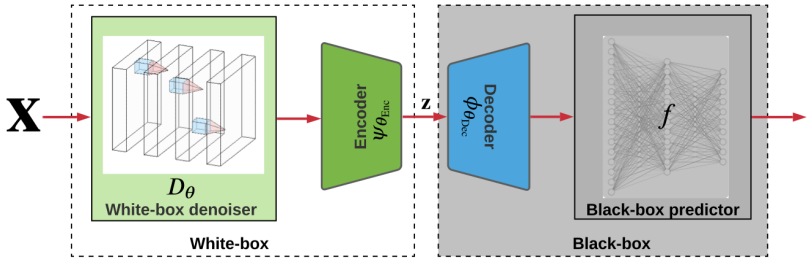
# Yimeng's Selected Publications

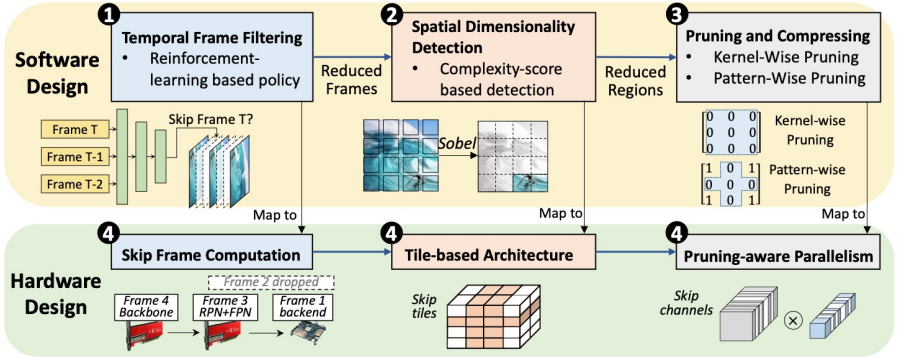
## ICLR'22 (Spotlight)




**How to Robustify Black-Box ML Models?**
**A Zeroth-Order Optimization Perspective**

DARPA

## ASP-DAC'23




**Data-Model-Circuit Tri-Design**
**for Ultra-Light Video Intelligence on Edge Devices**

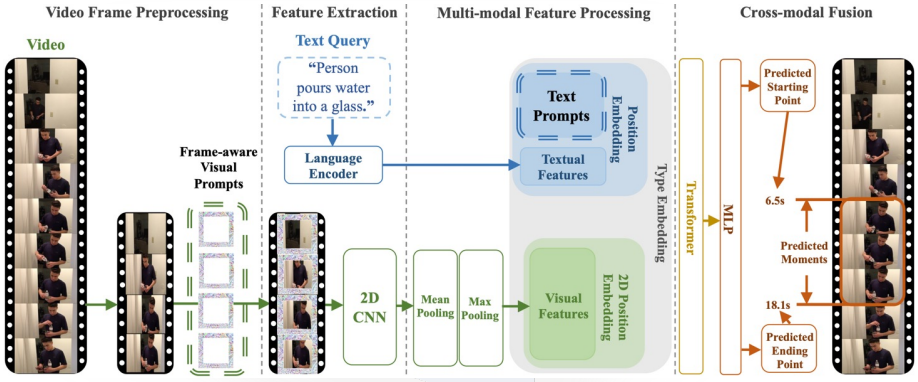
DARPA

MICHIGAN STATE UNIVERSITY

OPTML

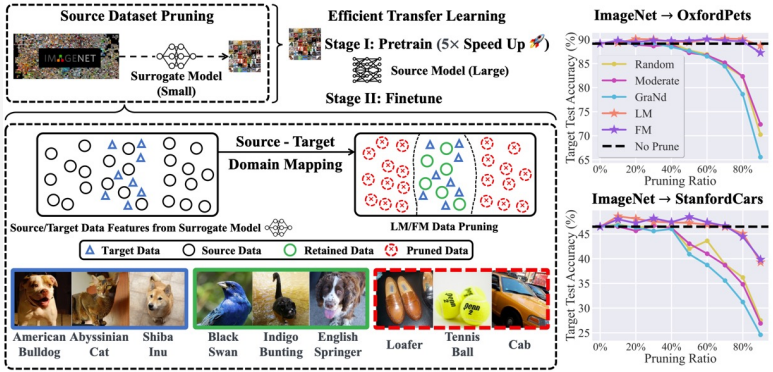
# Yimeng's Selected Publications

## CVPR'23




**Text-Visual Prompting for Efficient 2D Temporal Video Grounding**

## NeurIPS'23




**Selectivity Drives Productivity:**
**Efficient Dataset Pruning for Enhanced Transfer Learning**

MICHIGAN STATE UNIVERSITY

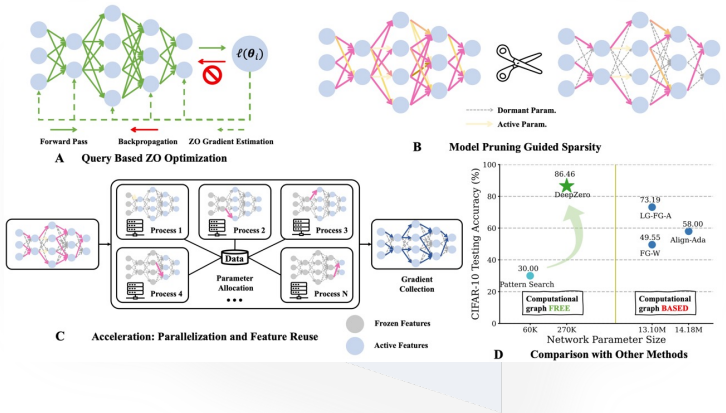
intel

NSF

OPTML

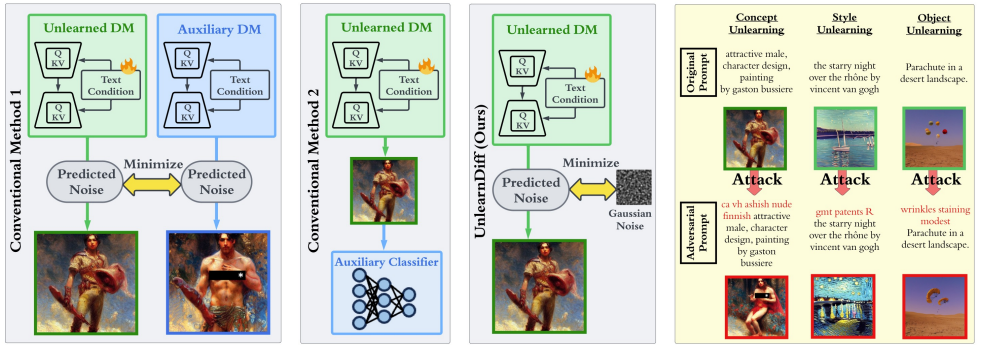
# Yimeng's Selected Publications

## ICLR'24




**DeepZero: Scaling up Zeroth-Order Optimization for Deep Model Training**

## Under Review




(a) W/ auxiliary DM (b) W/ classifier (c) Auxiliary model-free (d) `UnlearnDiff` attack demonstrations

**To Generate or Not? Safety-Driven Unlearned Diffusion Models Are Still Easy To Generate Unsafe Images ... For Now**

MICHIGAN STATE UNIVERSITY

Lawrence Livermore National Laboratory

intel

OPTML

# [ ICLR'22 ] How to Robustify Black-Box ML Models? A Zeroth-Order Optimization Perspective

**Motivations:**

- Nearly all existing works ask a defender to perform over white-box ML models. However, the white-box assumption may restrict the defense application in practice.
- Zeroth-Order (ZO) Optimization for high-dimension variables suffers **high variance**.

**Zeroth-Order Optimization for high-dimension variables suffers high variance** ⚠ ⚠ ⚠

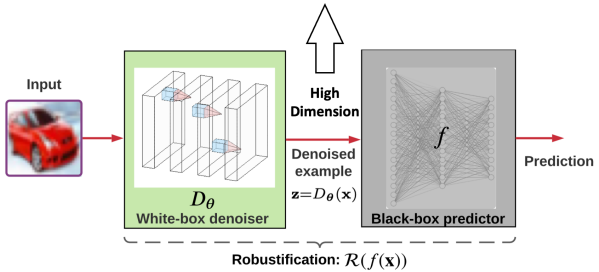
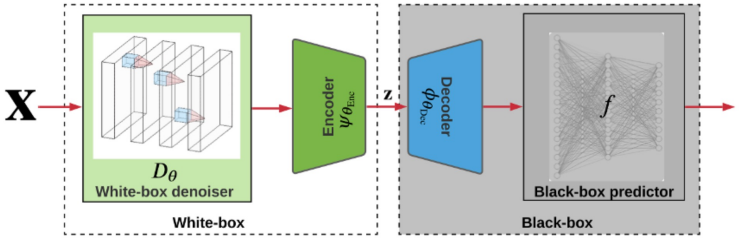



Figure 2: DS-based black-box defense.

$D_\theta$ : white-box denoiser with parameter $\theta$
$f$ : black-box predictor
$x$ : input

## ZO-AE-DS Model Architecture




## Random Gradient Estimate (RGE)

$$\hat{\nabla}_{\mathbf{w}}\ell(\mathbf{w}) = \frac{1}{q}\sum_{i=1}^{q}\left[\frac{d}{\mu}\left(\ell(\mathbf{w}+\mu\mathbf{u}_i)-\ell(\mathbf{w})\right)\mathbf{u}_i\right]$$

## Coordinate-wise Gradient Estimate (CGE)

$$\hat{\nabla}_{\mathbf{w}}\ell(\mathbf{w}) = \sum_{i=1}^{d}\left[\frac{\ell(\mathbf{w}+\mu\mathbf{e}_i)-\ell(\mathbf{w})}{\mu}\mathbf{e}_i\right]$$

## ZO gradient estimate of reduced dimension

$$\nabla_{\boldsymbol{\theta}}\mathcal{R}_{\text{new}}(f(\mathbf{x})) \approx \frac{d\phi_{\boldsymbol{\theta}_{\text{Enc}}}(D_{\boldsymbol{\theta}}(\mathbf{x}))}{d\boldsymbol{\theta}}\ \hat{\nabla}_{\mathbf{z}}f'(\mathbf{z})\,|_{\mathbf{z}=\phi_{\boldsymbol{\theta}_{\text{Enc}}}(D_{\boldsymbol{\theta}}(\mathbf{x}))}$$

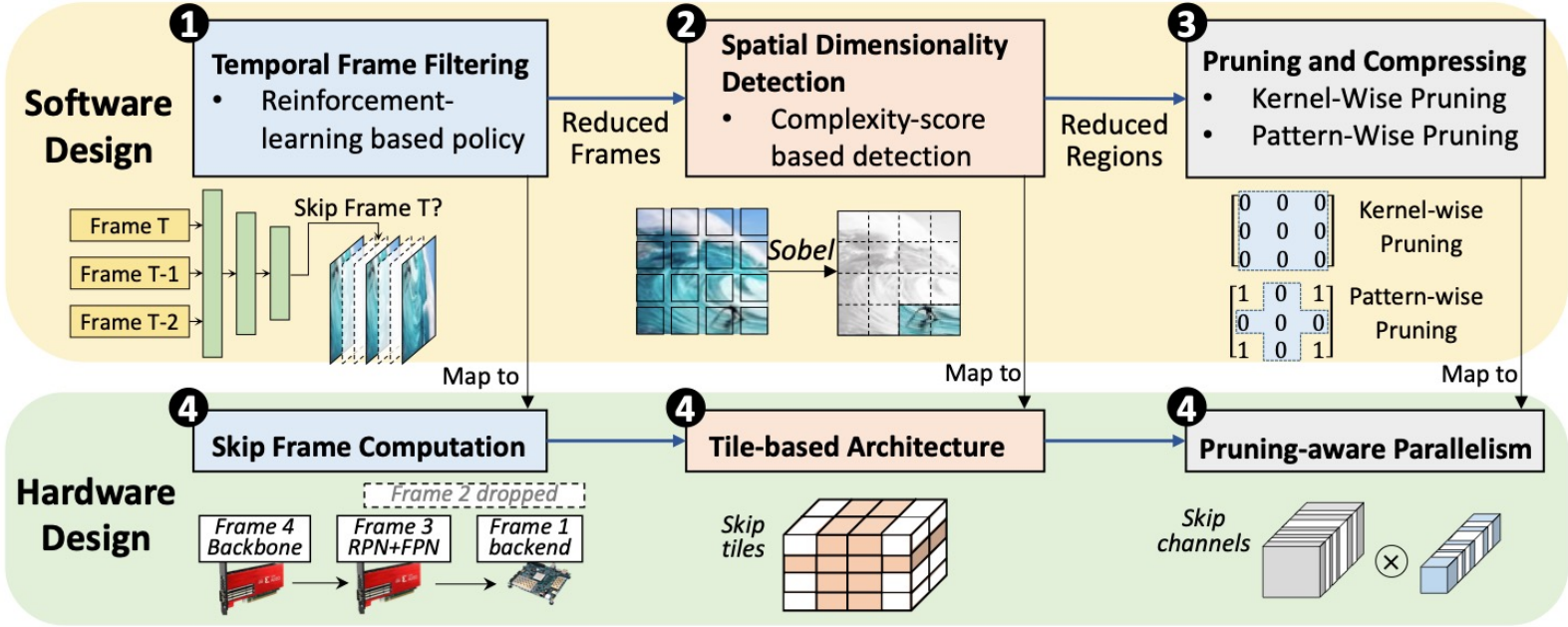
FO Gradient (Backpropogation) | ZO Gradient Estimation

MICHIGAN STATE UNIVERSITY

OPTML

**Task:**

Efficient implementation of multi-object tracking (MOT) on the edge devices for HD video processing by fully utilizing data- and model-level sparsity.




MICHIGAN STATE UNIVERSITY

OPTML

# [ CVPR'23 ] Text-Visual Prompting for Efficient 2D Temporal Video Grounding

## Task:

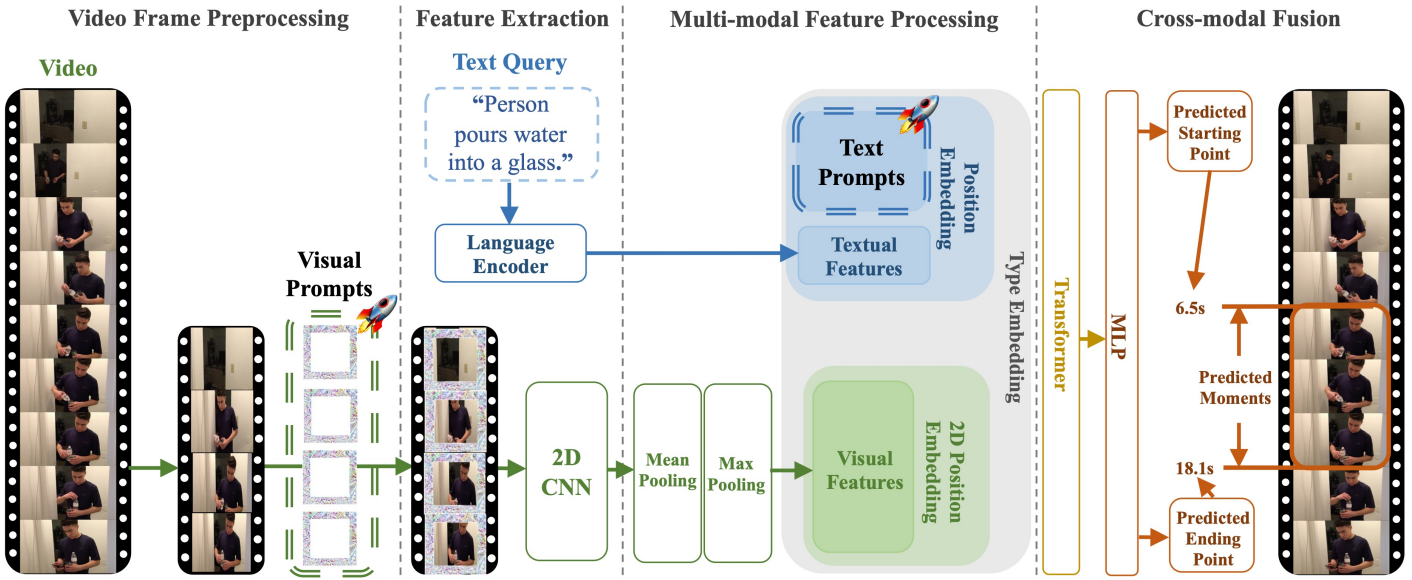
TVG is to predict the starting/ending time points of moments described by a text sentence within a long untrimmed video.

## Motivation:

High complexity of 3D CNNs makes extracting dense 3D visual features time- consuming, which calls for intensive memory and computing resources.

## Challenges:

How to advance 2D TVG methods so as to achieve comparable results to 3D TVG methods?




MICHIGAN STATE UNIVERSITY

OPTML

# [ NeurIPS'23 ] Selectivity Drives Productivity: Efficient Dataset Pruning for Enhanced Transfer Learning

## Task:

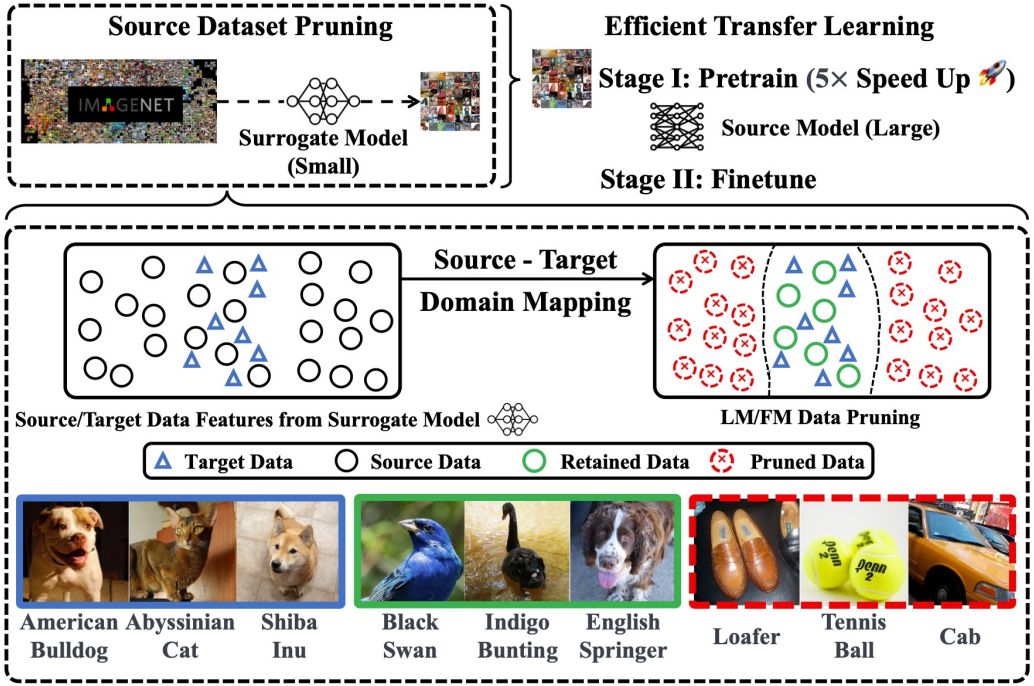
dataset pruning for transfer learning
→ Find a subset of source data for pretraining

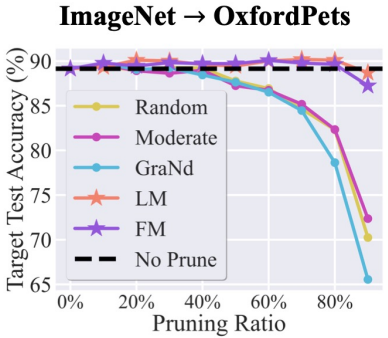
## Motivation:

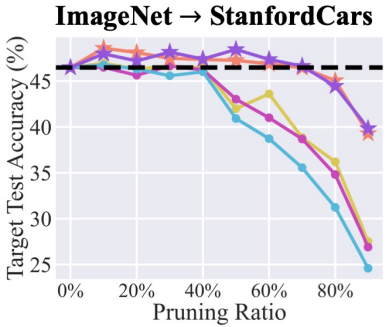
Some source data could make a harmful influence in the downstream performance.

## Rationales:

Source data similar to downstream data intend to contribute more during the transfer process




MICHIGAN STATE UNIVERSITY

OPTML

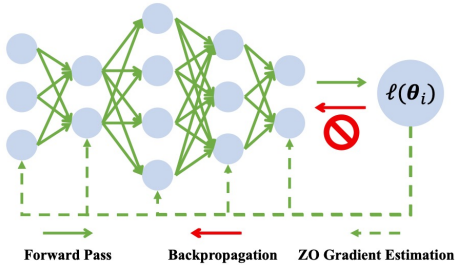
# [ ICLR'24 ] DeepZero: Scaling up Zeroth-Order Optimization for Deep Model Training

## Task:

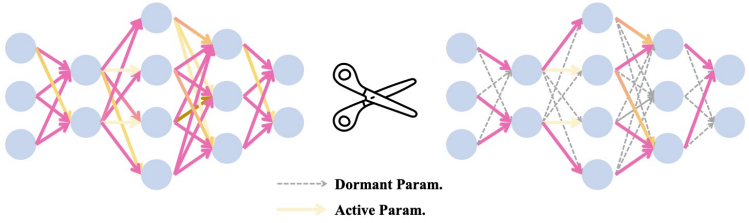
How to scale up ZO optimization for training deep models real-world circumstances where FO gradients are difficult to obtain?
(e.g., physics-informed DL tasks)

## Challenges:

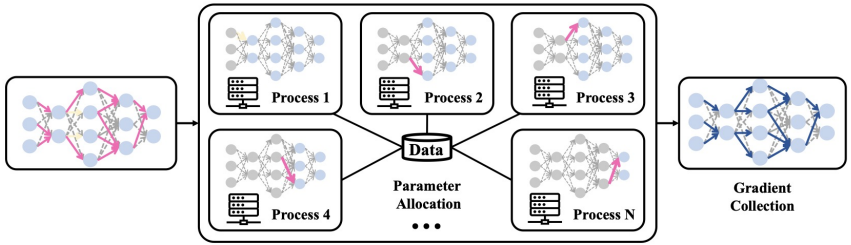
ZO finite difference-based gradient estimates are biased estimators of FO gradients, and the bias becomes more pronounced in higher-dimensional spaces



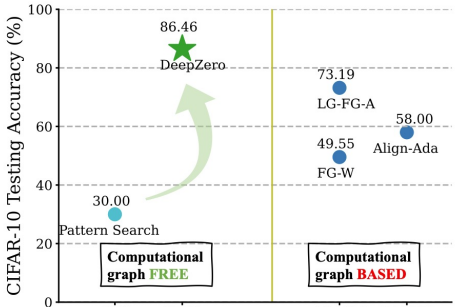

A Query Based ZO Optimization




B Model Pruning Guided Sparsity




C Acceleration: Parallelization and Feature Reuse




D Comparison with Other Methods

MICHIGAN STATE UNIVERSITY

OPTML

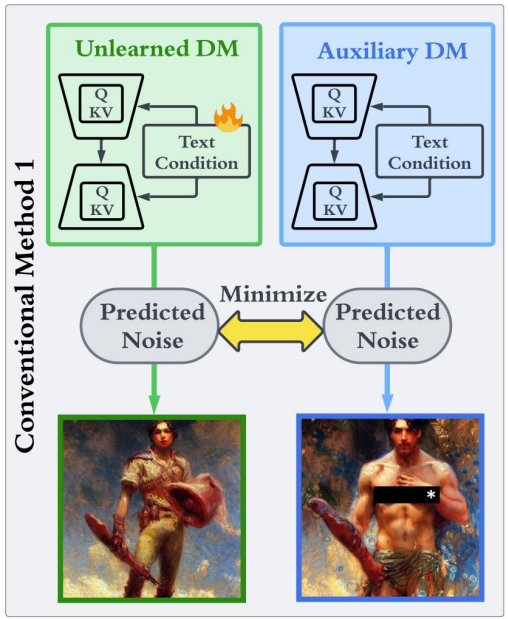
# [ Under Review ] To Generate or Not?
# Safety-Driven Unlearned Diffusion Models Are Still Easy To Generate Unsafe Images ... For Now

## Task:

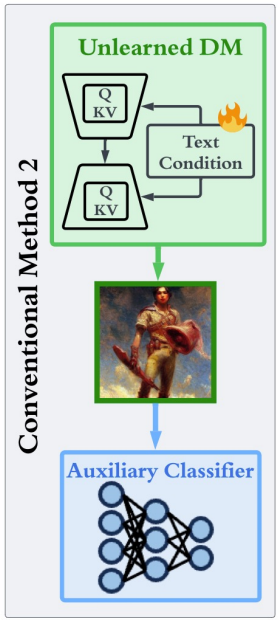
investigate the robustness of stateful unlearned diffusion models (DMs) in effectively eliminating undesired concepts, styles, and objects by crafting adversarial prompts.

## Method:

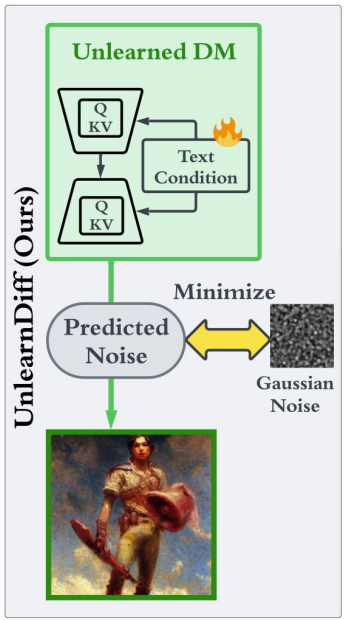
develop a novel adversarial prompt generation method called *UnlearnDiff*, which leverages the inherent classification capabilities of DMs, simplifying the generation of adversarial prompts for generative modeling as much as it is for image classification attacks.



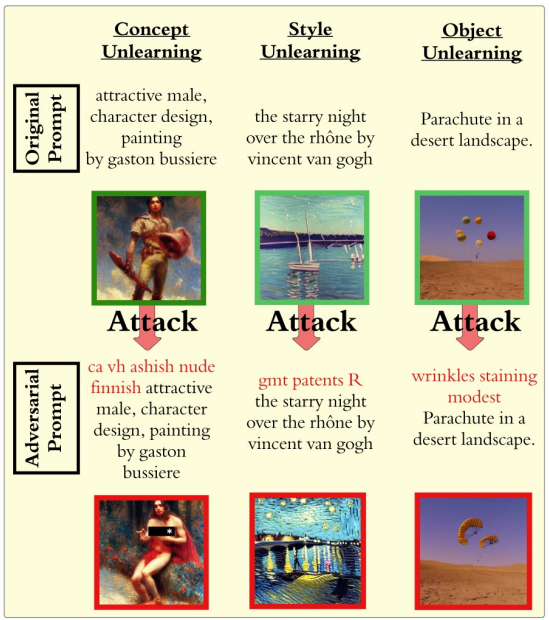

(a) W/ auxiliary DM




(b) W/ classifier




(c) Auxiliary model-free




(d) `UnlearnDiff` attack demonstrations