



TRUSTED AI LABS  
BY DIGITALWALLONIA4.AI / SPW-RECHERCHE

*Controlled Image Augmentation Using  
Stable Diffusion*

Mohamed Benkedadra

# Mohamed Benkedadra's PhD thesis

ARIAC (Septembre 2021 – ) – Directors : Prof. Sidi Ahmed Mahmoudi, Dr. Matei Mancas

## Embedded (“Edge”) and Explainable (“XAI”) Deep Learning for Real time Dangerous Action Detection and Prediction

- Exploit 2D and 3D images and videos captured from multiple cameras to train very accurate models for dangerous action/situation detection and prediction.
- Deploy Danger Detection on Edge AI resources.
- Generalized and adaptive solutions that work on new never before seen situations.

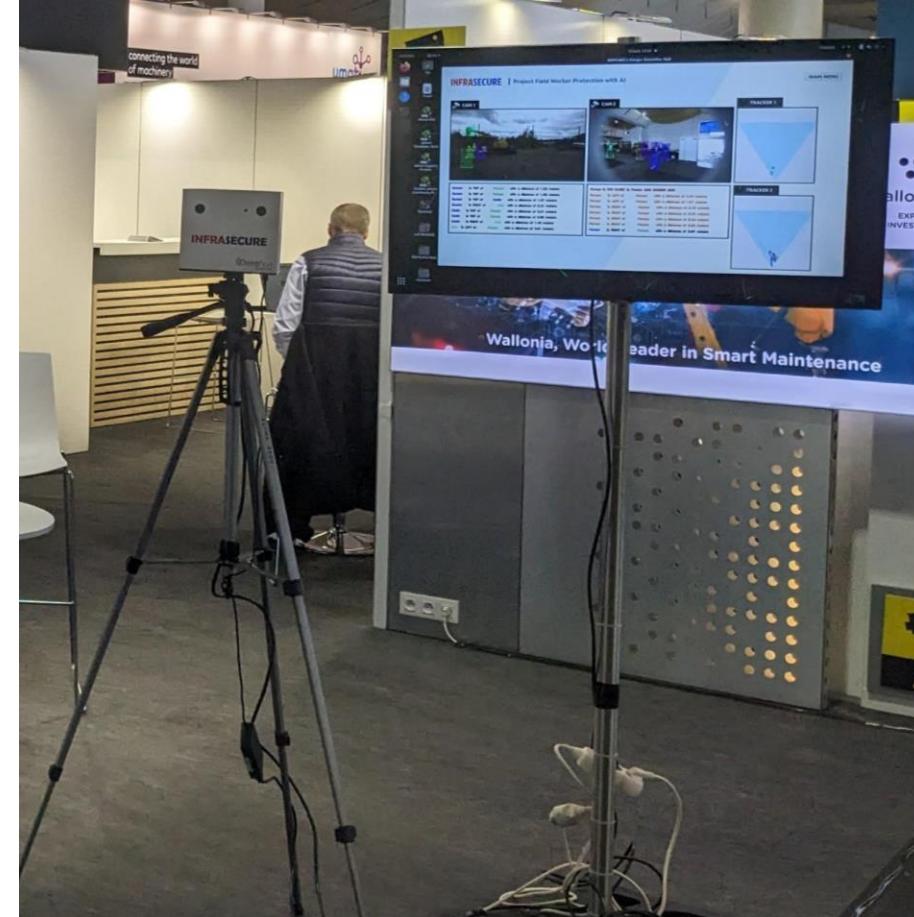


# Mohamed Benkedadra's PhD thesis

ARIAC (Septembre 2021 – ) – Directors : Prof. Sidi Ahmed Mahmoudi, Dr. Matei Mancas

## Embedded (“Edge”) and Explainable (“XAI”) Deep Learning for Real time Dangerous Action Detection and Prediction

- Exploit 2D and 3D images and videos captured from multiple cameras to train very accurate models for dangerous action/situation detection and prediction.
- Deploy Danger Detection on Edge AI resources.
- Generalized and adaptive solutions that work on new never before seen situations.



# Mohamed Benkedadra's PhD thesis

ARIAC (Septembre 2021 – ) – Directors : Prof. Sidi Ahmed Mahmoudi, Dr. Matei Mancas

## Embedded (“Edge”) and Explainable (“XAI”) Deep Learning for Real time Dangerous Action Detection and Prediction

- Exploit 2D and 3D images and videos captured from multiple cameras to train very accurate models for dangerous action/situation detection and prediction.
- Deploy Danger Detection on Edge AI resources.
- Generalized and adaptive solutions that work on new never before seen situations.



# Mohamed Benkedadra's PhD thesis

ARIAC (Septembre 2021 – ) – Directors : Prof. Sidi Ahmed Mahmoudi, Dr. Matei Mancas

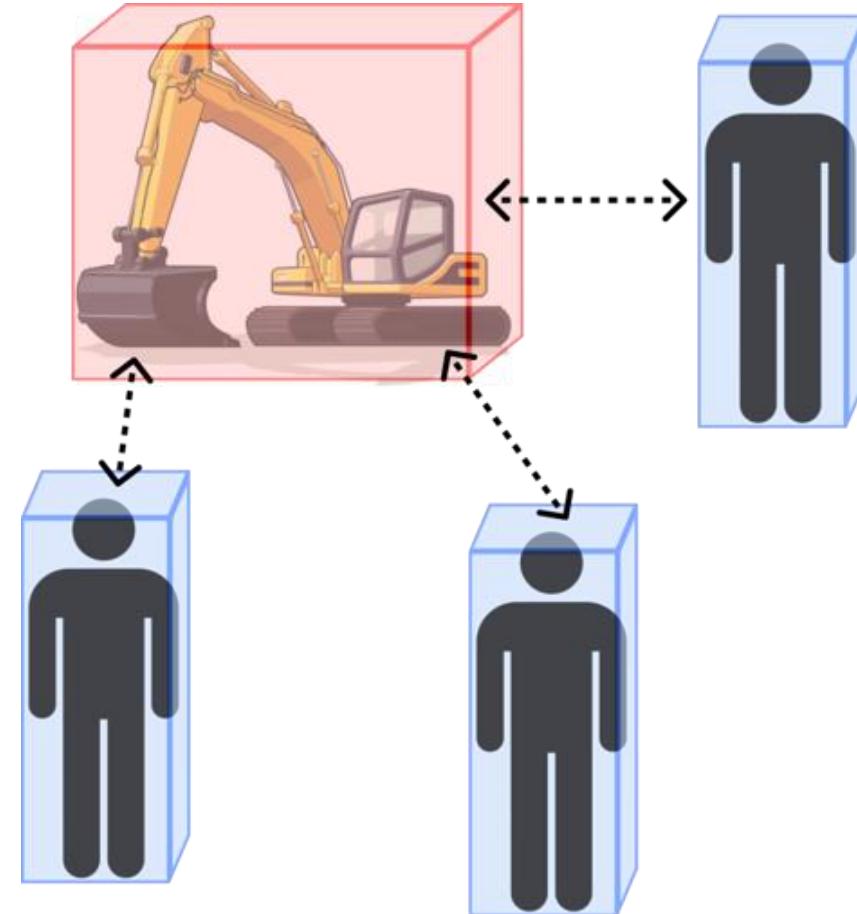
## Embedded (“Edge”) and Explainable (“XAI”) Deep Learning for Real time Dangerous Action Detection and Prediction

- Exploit 2D and 3D images and videos captured from multiple cameras to train very accurate models for dangerous action/situation detection and prediction.
- Deploy Danger Detection on Edge AI resources.
- **Generalized and adaptive solutions that work on new never before seen situations.**



# How does the current system work ?

- Detect Objects
- User can specify which objects are dangerous and which objects are safe
- User can specify safety positions and distances



# How can we learn new unknown objects ?



Already Trained Models



New Scary and Dangerous Object

# How can we learn new unknown objects ?



New Scenario

# How can we learn new unknown objects ?



# How can we learn new unknown objects ?



Optical Flow



# How can we learn new unknown objects ?

**SPARSE : Feature Specific**



**DENSE : All Pixels**



# How can we learn new unknown objects ?

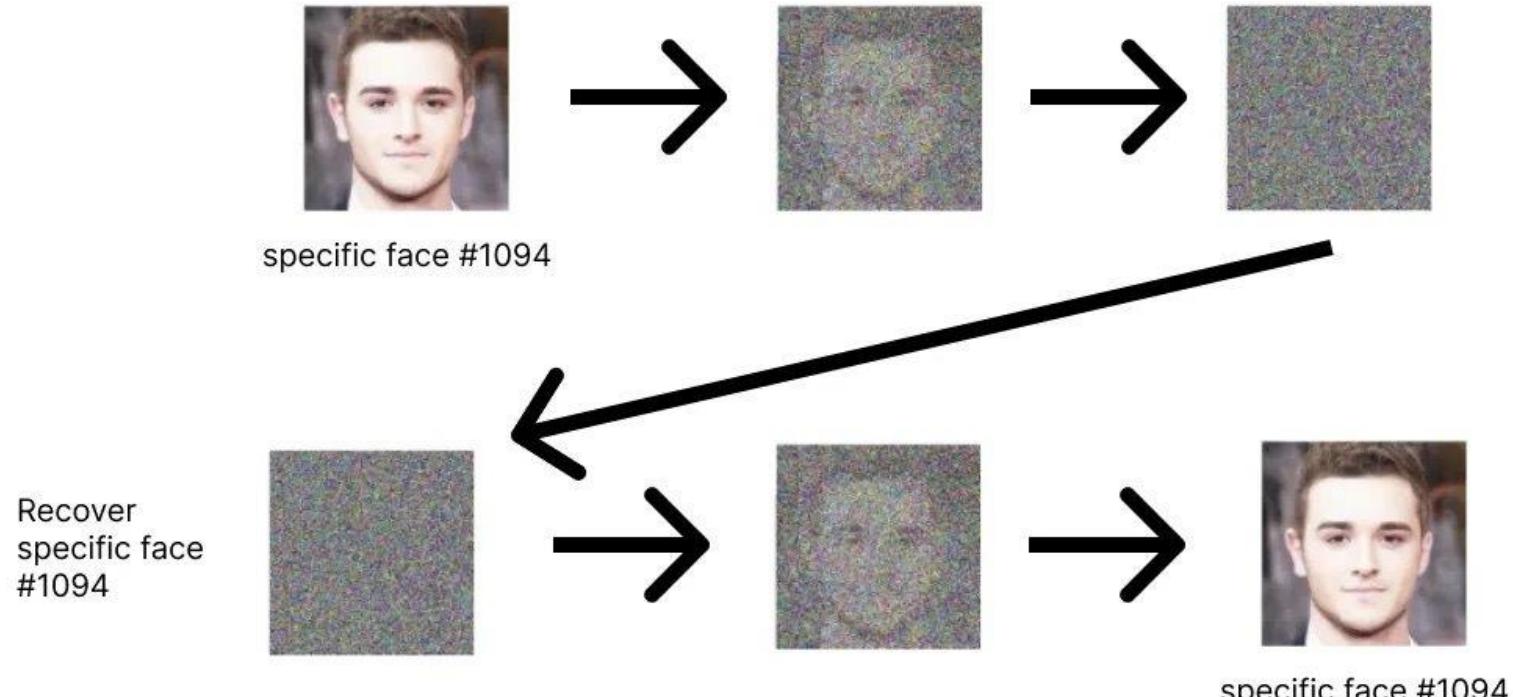


# Synthetic Data Augmentation



# Synthetic Data Augmentation

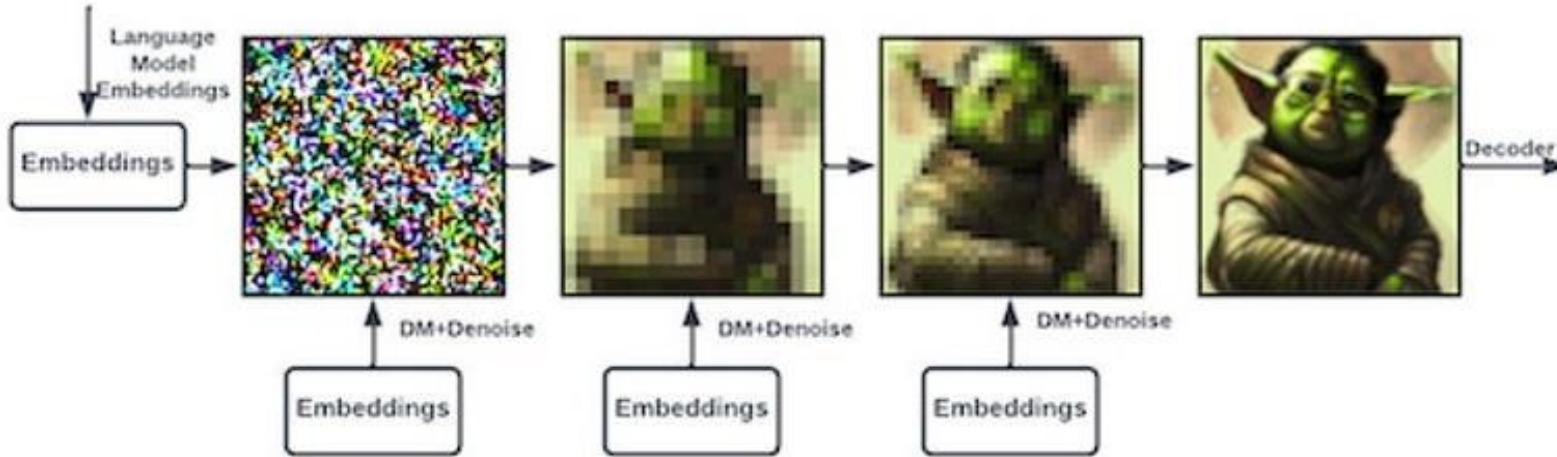
Diffusion Models learn to add more and more noise



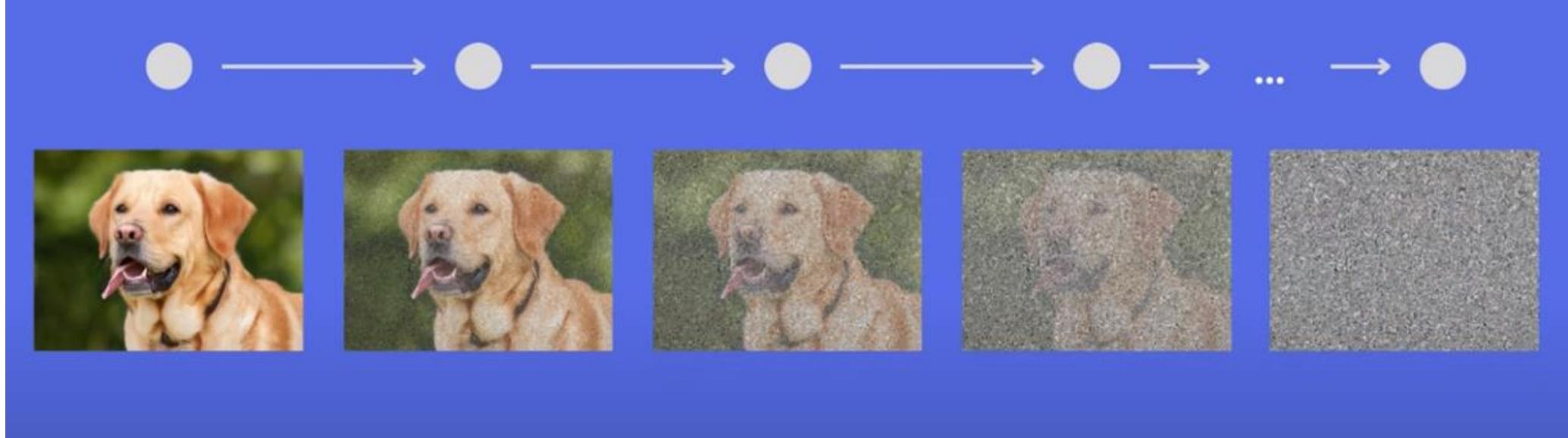
Then learn to recover the original image

# Synthetic Data Augmentation

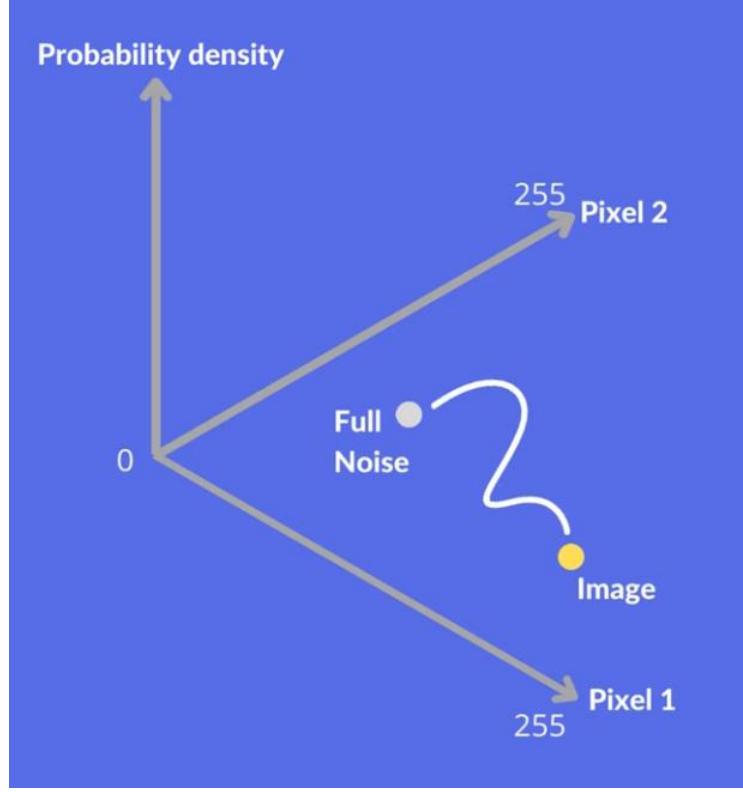
"(realbenny-t1 | yoda) person, star war, art by artgerm and greg rutkowski and alphonse mucha"



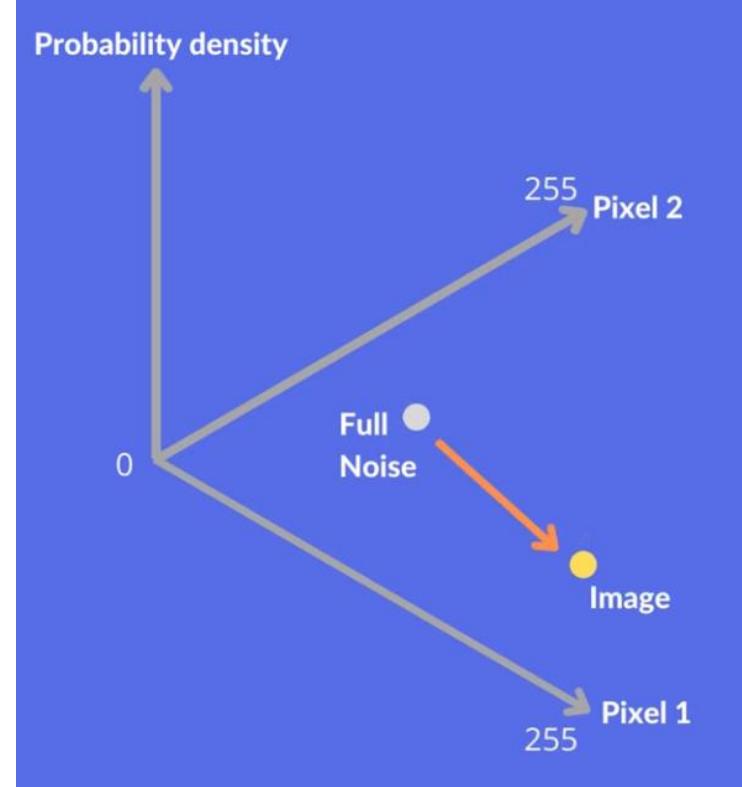
# Synthetic Data Augmentation



# Synthetic Data Augmentation

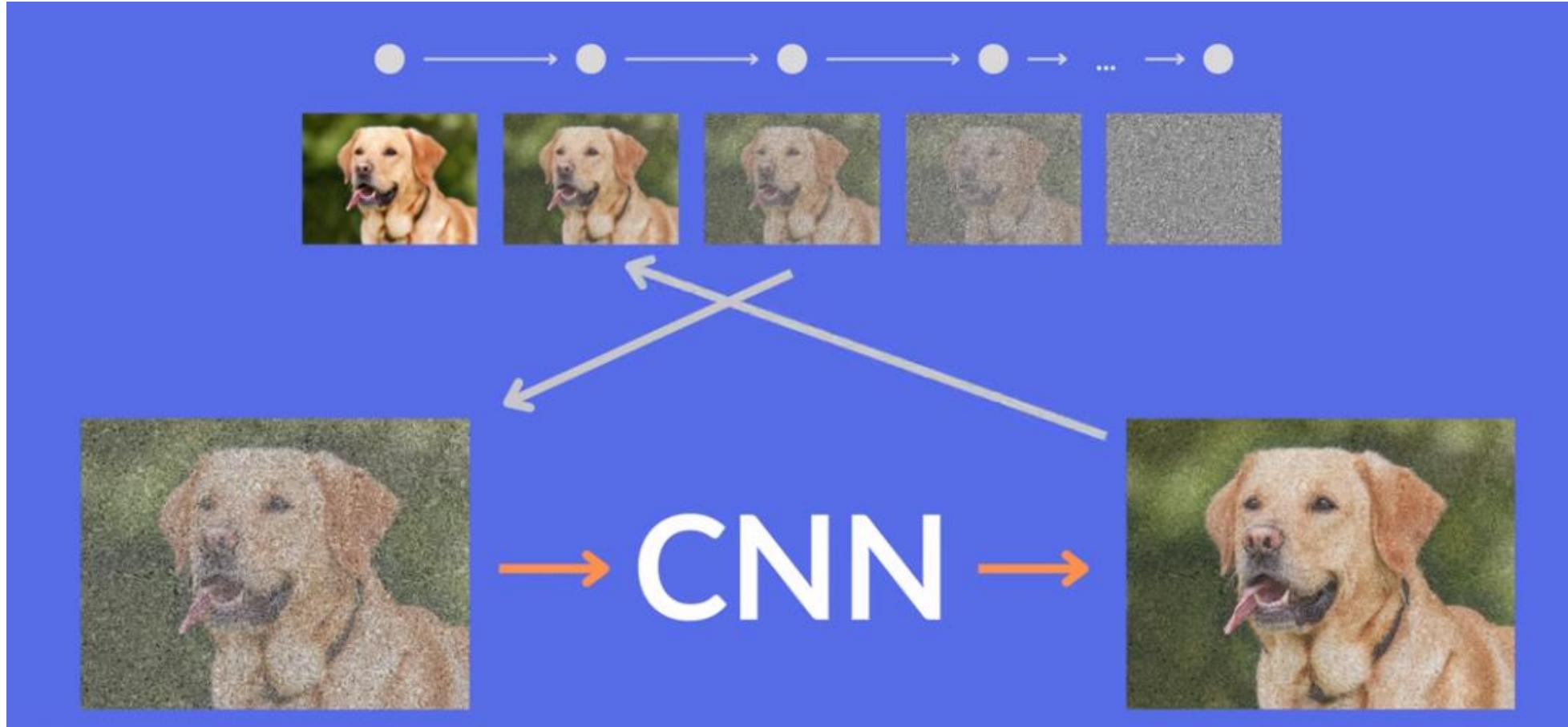


Diffusion



Reverse Diffusion

# Synthetic Data Augmentation



# Synthetic Data Augmentation



# Synthetic Data Augmentation



# Synthetic Data Augmentation (Image Space)

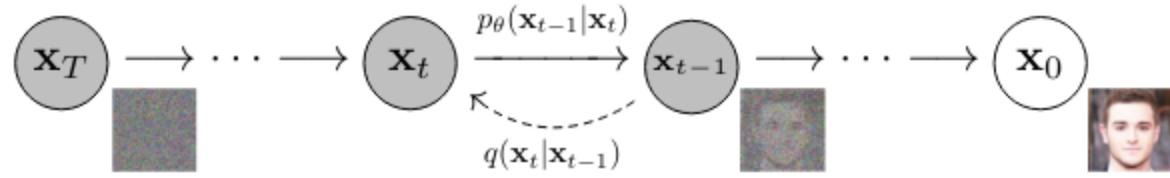
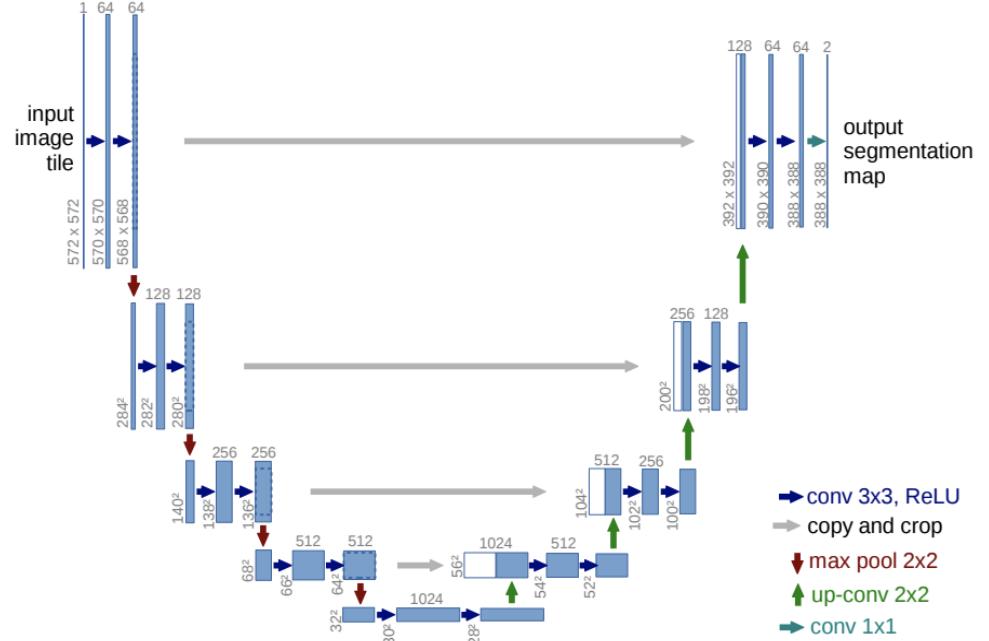
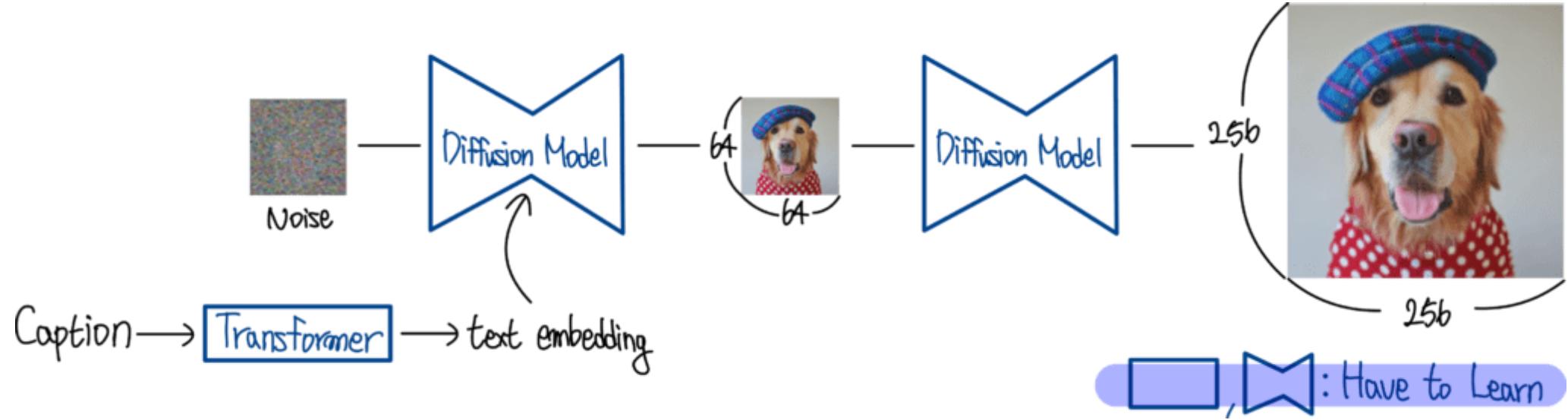


Figure 2: The directed graphical model considered in this work.



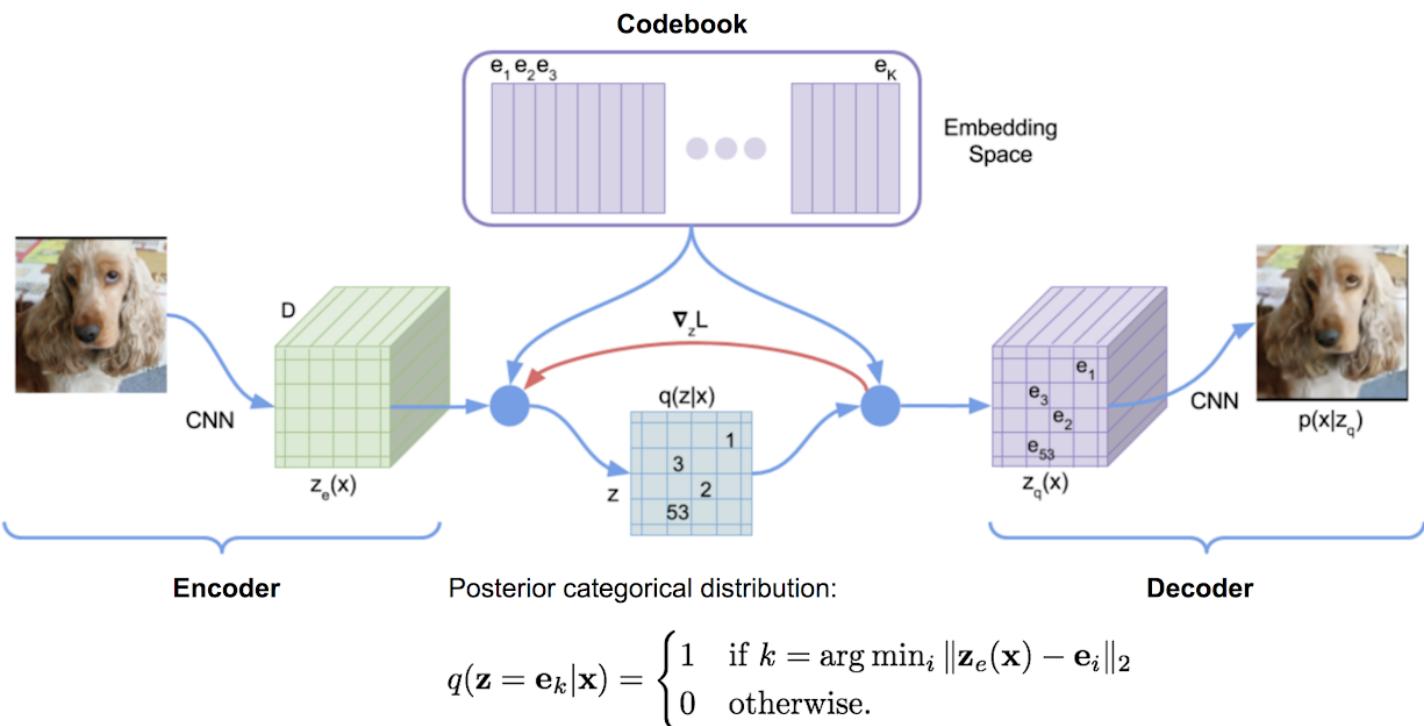
Checkout: Denoising Diffusion Probabilistic Models (2020)

# Synthetic Data Augmentation (Image Space)



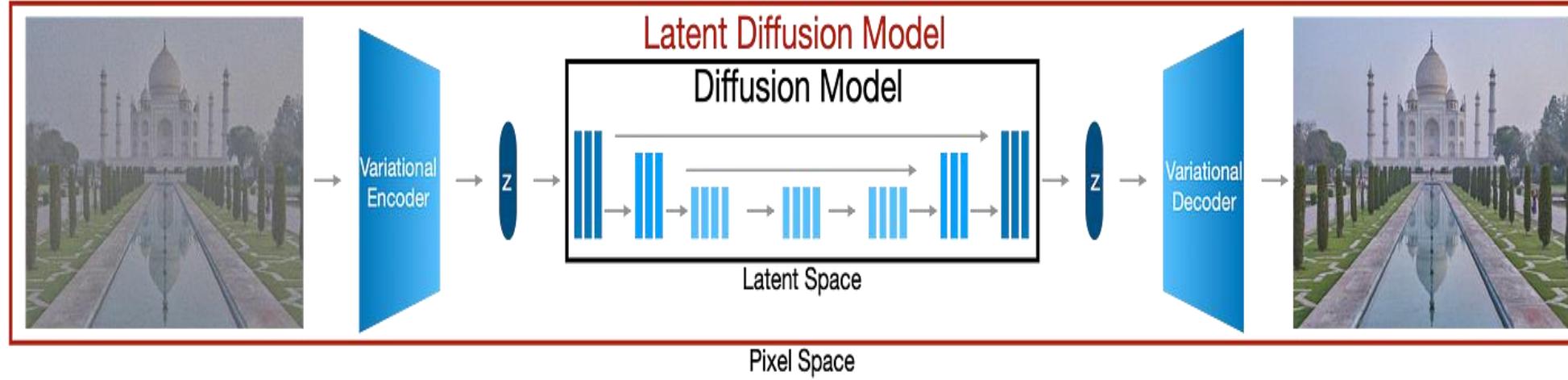
Checkout: GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models (2021)

# Synthetic Data Augmentation (Latent Space)



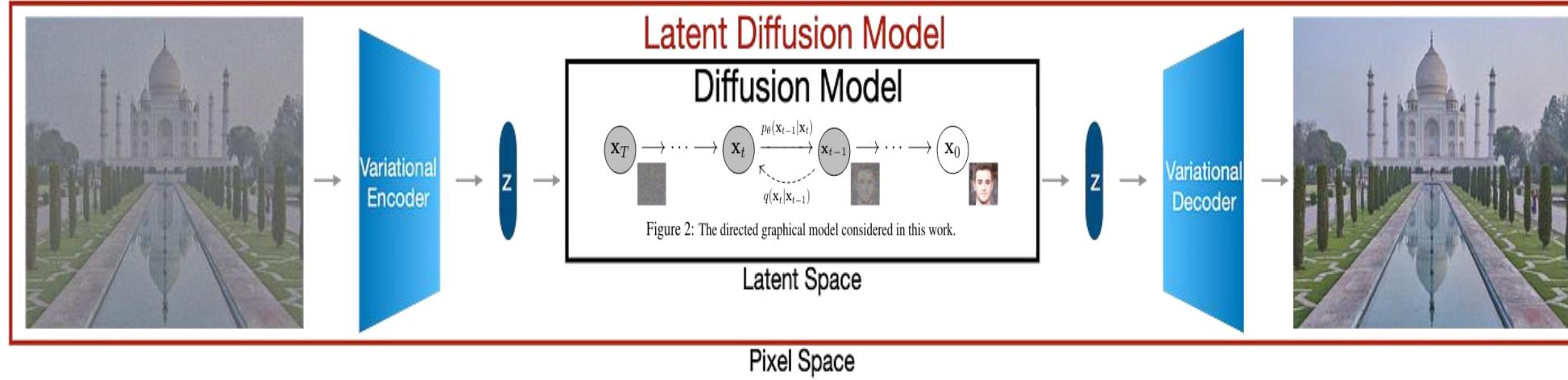
From : Dall-E - Zero-Shot Text-to-Image Generation (2021)

# Synthetic Data Augmentation (Latent Space)



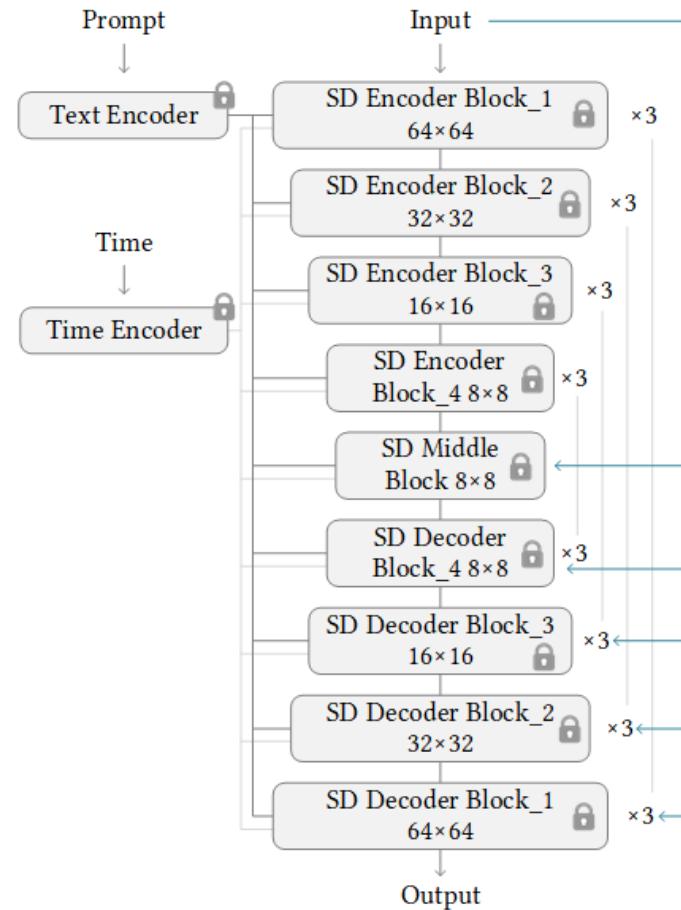
Checkout : High-Resolution Image Synthesis with Latent Diffusion Model  
(2022, First appearance of current Stable Diffusion => Latent Diffusion)

# Synthetic Data Augmentation (Latent Space)



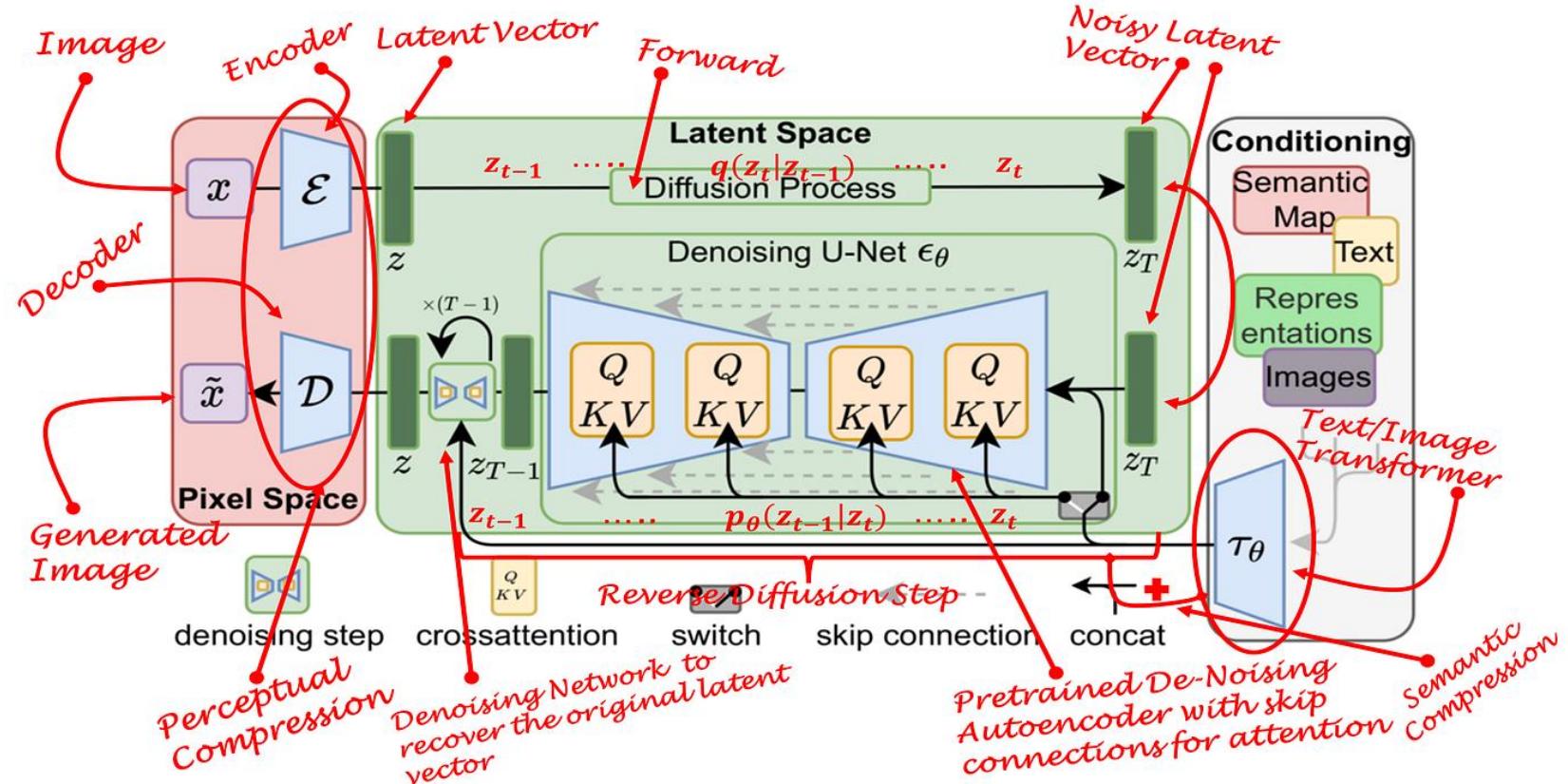
Checkout : High-Resolution Image Synthesis with Latent Diffusion Model  
(2022, First appearance of current Stable Diffusion)

# Synthetic Data Augmentation (Latent Space)



Checkout : High-Resolution Image Synthesis with Latent Diffusion Model  
(2022, First appearance of current Stable Diffusion)

# Synthetic Data Augmentation (Latent Space)



From : High-Resolution Image Synthesis with Latent Diffusion Model  
(2022, First appearance of current Stable Diffusion)

# Controlled Synthetic SD Data Augmentation



Input Canny edge



Default



"masterpiece of fairy tale, giant deer, golden antlers"



"..., quaint city Galic"



Input human pose



Default



"chef in kitchen"

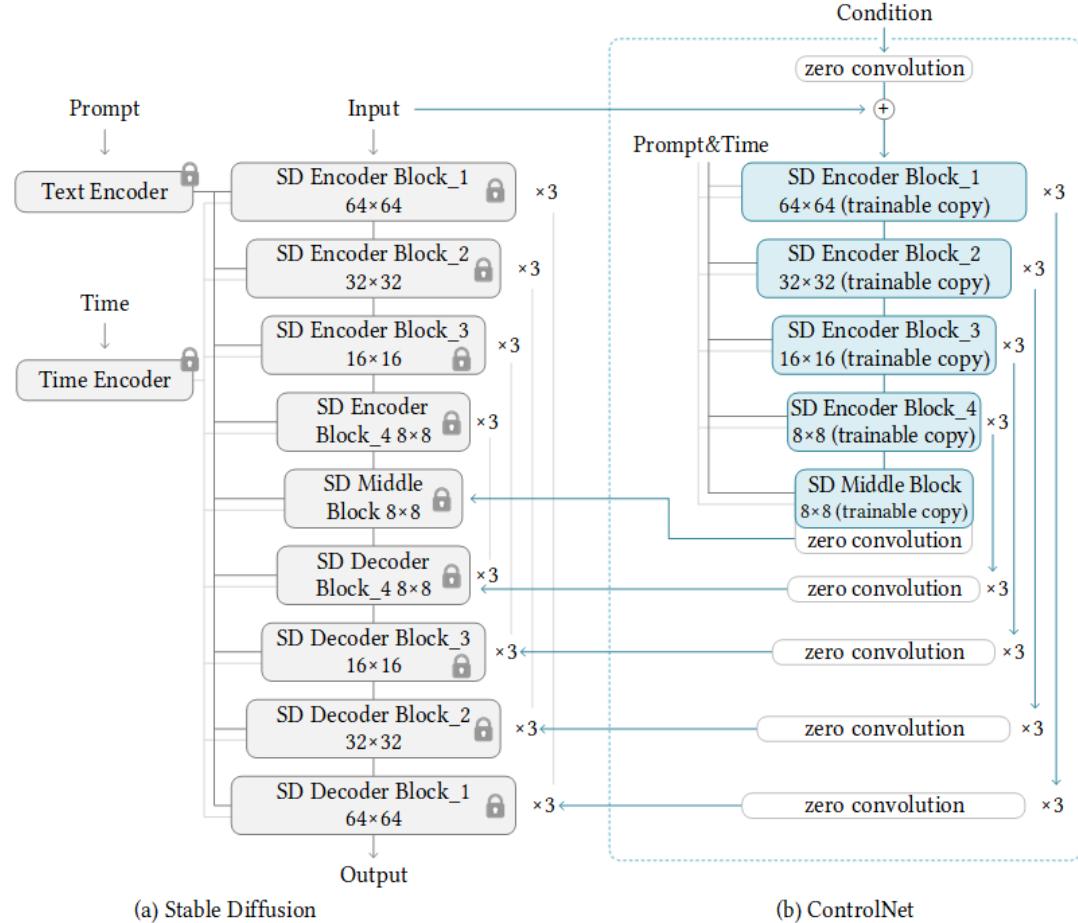


"Lincoln statue"



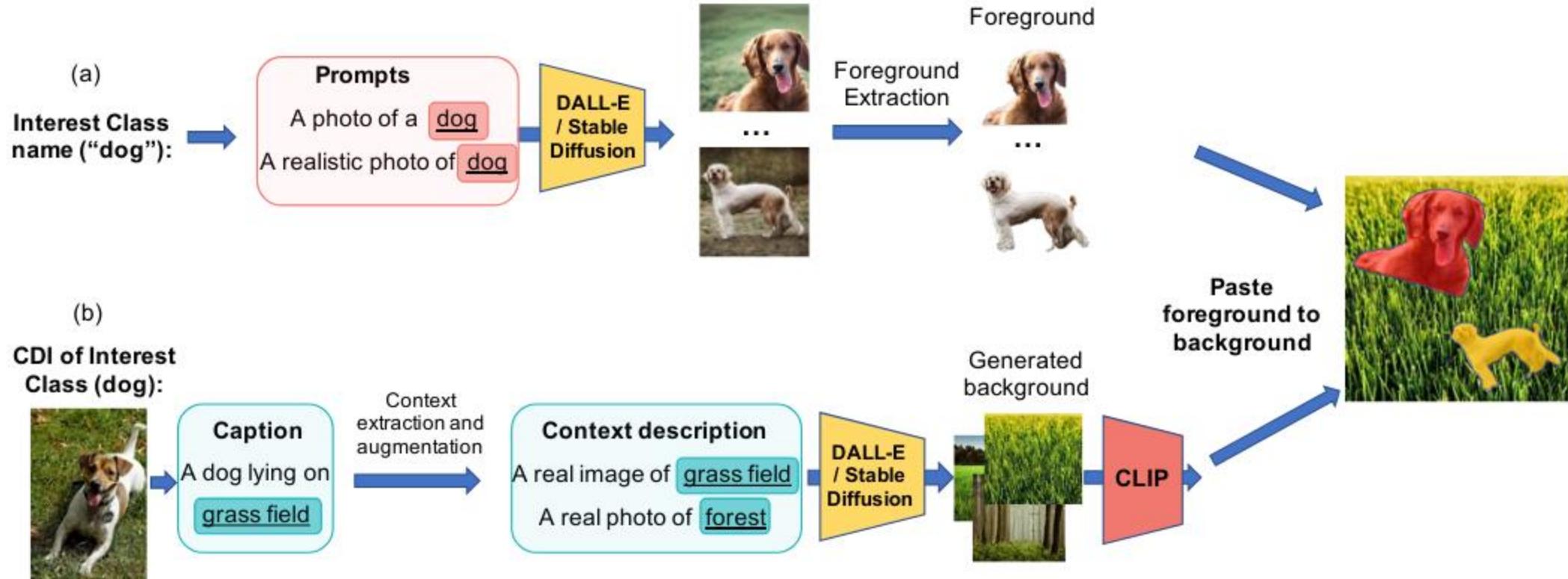
From : Adding Conditional Control to Text-to-Image Diffusion Models (2023)

# Controlled Synthetic SD Data Augmentation



From : Adding Conditional Control to Text-to-Image Diffusion Models (2023)

# Controlled Synthetic Dall-E Data Augmentation



DALL-E for Detection: Language-driven Compositional Image Synthesis for Object Detection (2022)

# Half of the Puzzle is Solved : Our Paper !

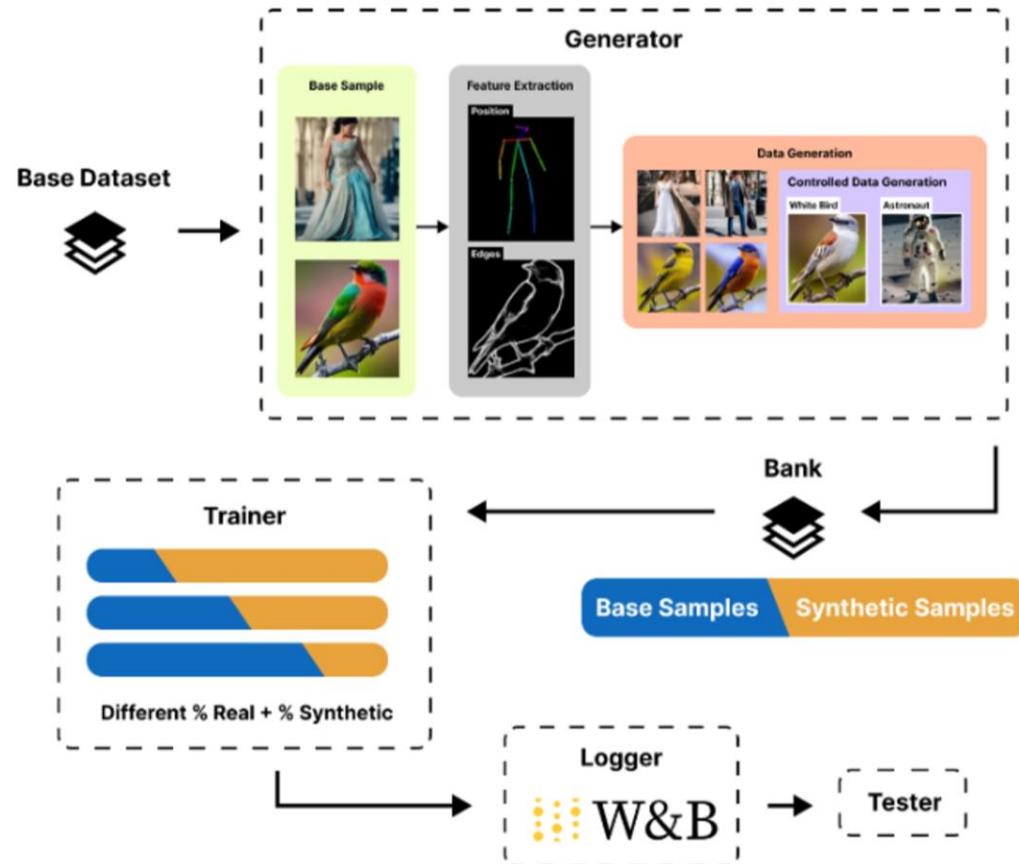


Figure 1: General Baseline Framework/Plan

CIA: Controllable Image Augmentation Framework based on Stable Diffusion (MIPR 2024)  
To be presented August 9th.

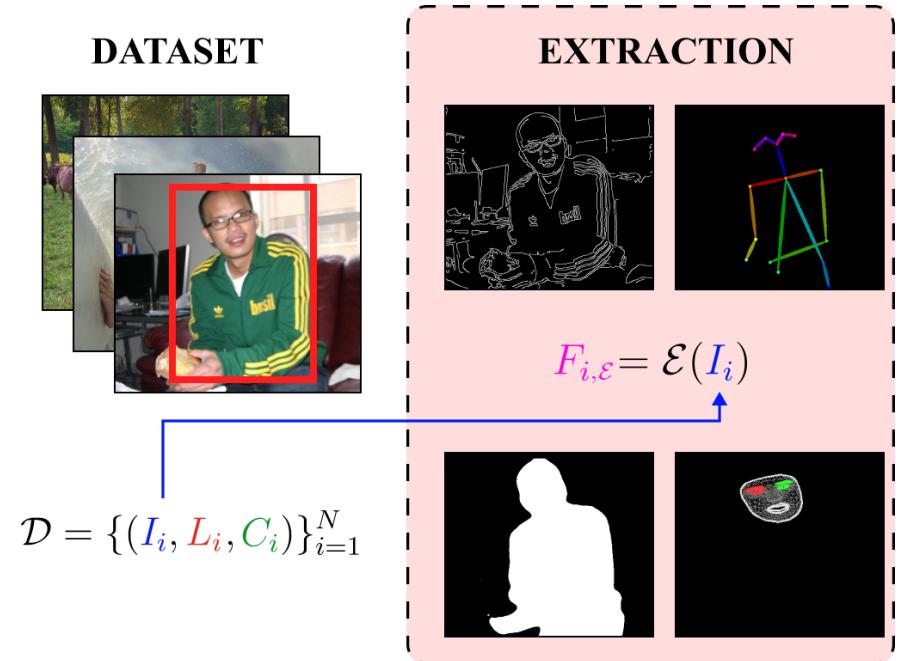
# CIA Framework

## DATASET

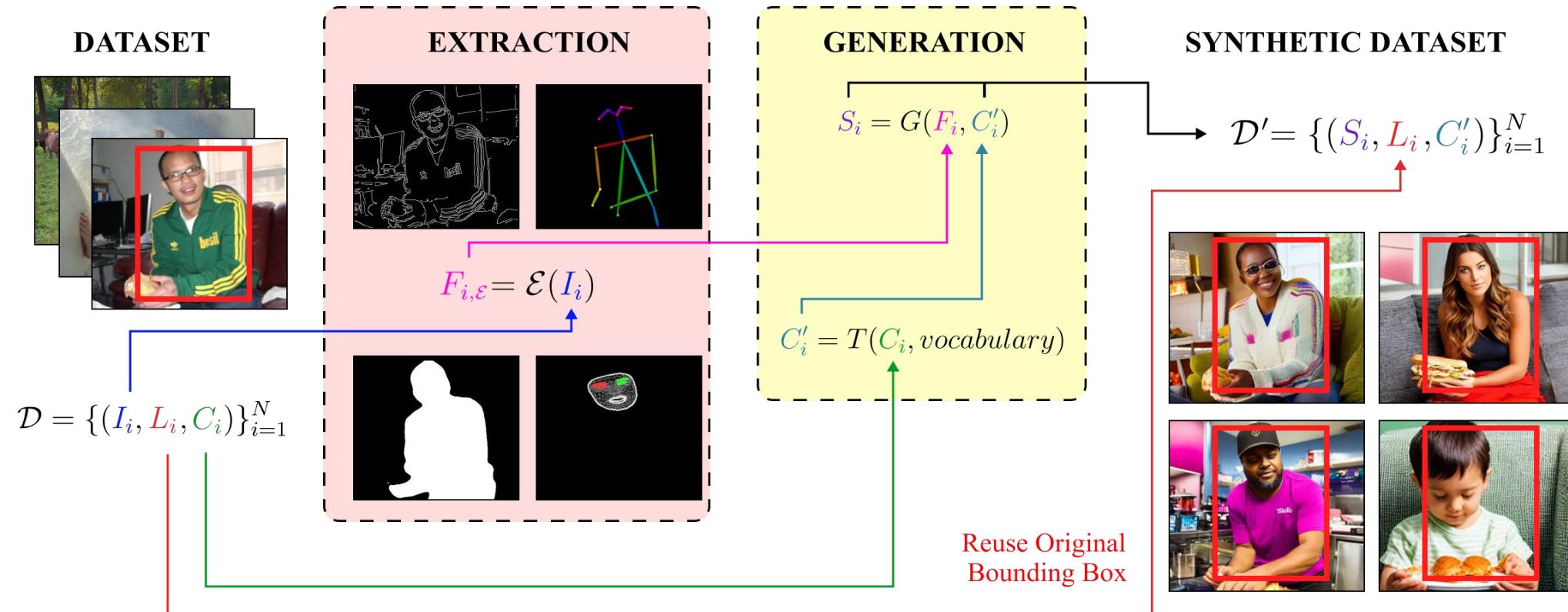


$$\mathcal{D} = \{(\textcolor{blue}{I}_i, \textcolor{red}{L}_i, \textcolor{green}{C}_i)\}_{i=1}^N$$

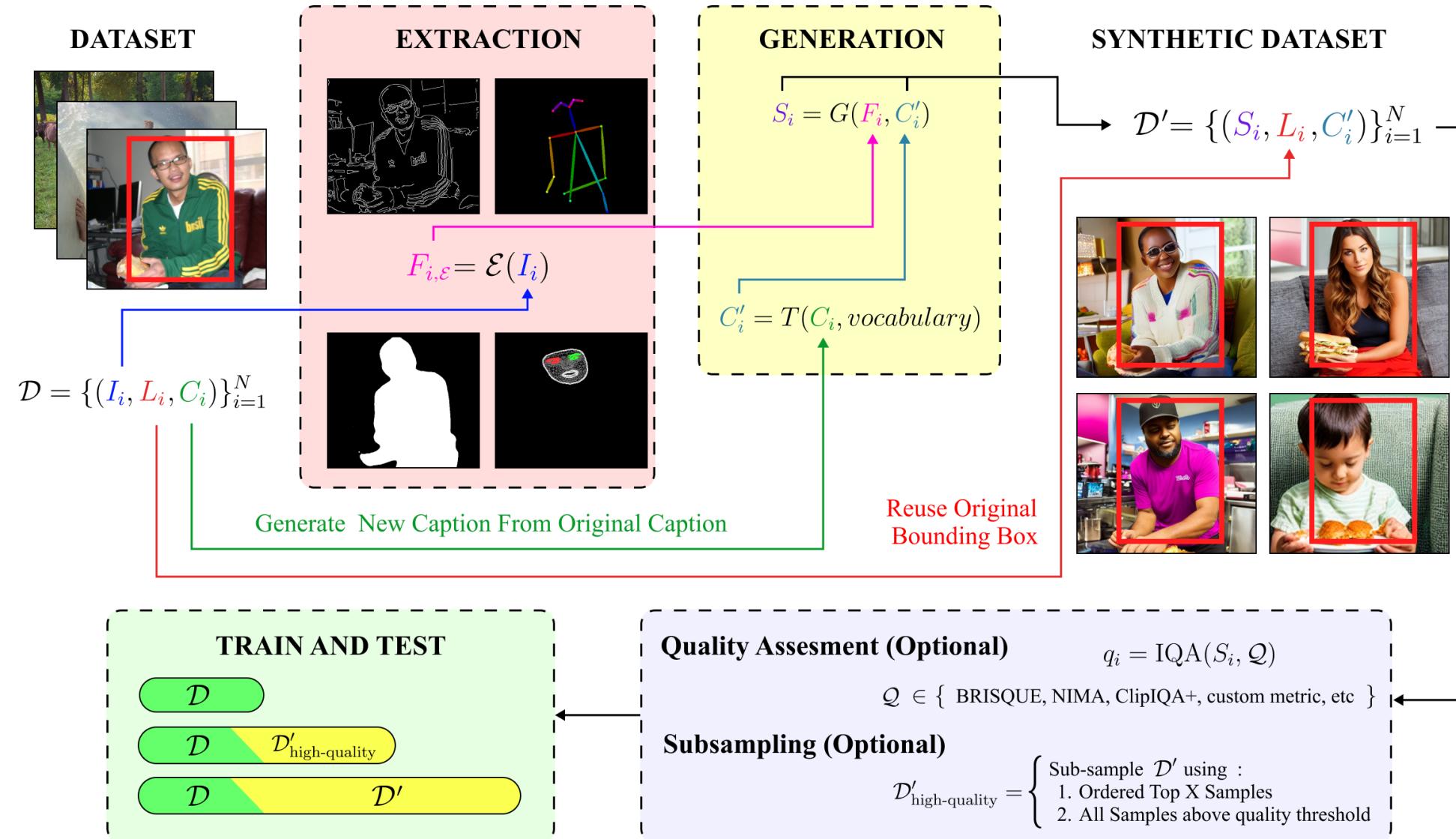
# CIA Framework



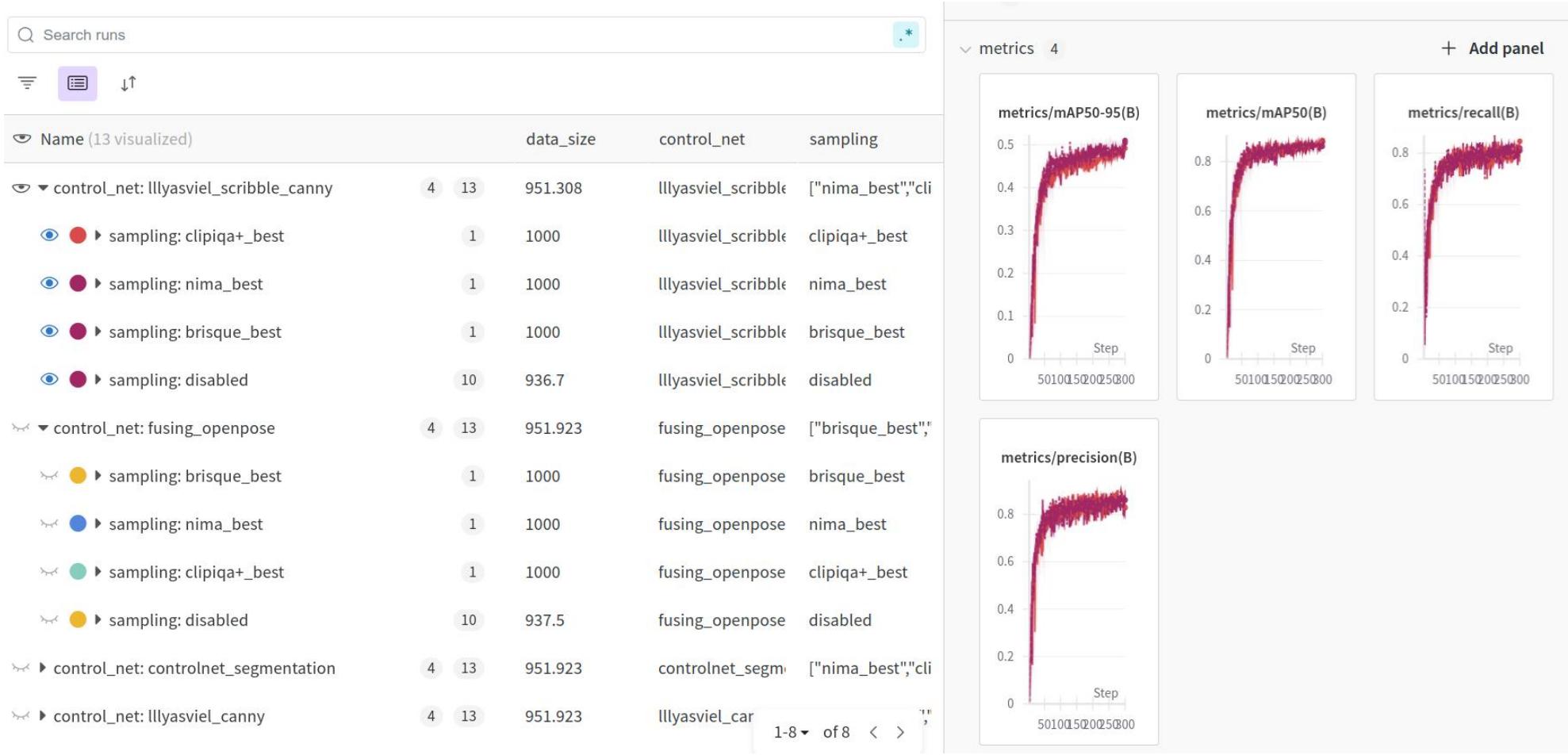
# CIA Framework



# CIA Framework



# CIA Framework : Wandb Logging



# CIA Framework : Config

```

conf > config.yaml > ...
conf > config.yaml > ...
You, 5 months ago | 2 authors (You and one other)
1  # © - 2024 Université de Mons, Multitel, Université Libre de Bruxelles, Université Catholique de Louvain
2
3  # CIA is free software. You can redistribute it and/or modify it
4  # under the terms of the GNU Affero General Public License
5  # as published by the Free Software Foundation, either version 3
6  # of the License, or any later version. This program is distributed
7  # in the hope that it will be useful, but WITHOUT ANY WARRANTY;
8  # without even the implied warranty of MERCHANTABILITY or FITNESS
9  # FOR A PARTICULAR PURPOSE. See the GNU Affero General Public License
10 # for more details. You should have received a copy of the Lesser GNU
11 # General Public License along with this program.
12 # If not, see <http://www.gnu.org/licenses/>.
13
14 data:
15   base: "data"
16   real: "real" # should contains images/, labels/, and captions/ (captions are optional)
17   generated: "generated"
18   datasets: "datasets"
19   image_formats: ["jpeg", "jpg"]
20
21 # put every parameter related to machine learning here: dataset size,
22 # ratio between train and test, learning rate ...
23 ml:
24   val_nb: 300
25   test_nb: 300
26   train_nb: 250
27   augmentation_percent: 0.1 # controls all aug percents parameters everywhere
28   augmentation_percent_baseline: 0
29   baseline: True
30   epochs: 300
31   sampling:
32     enable: false
33     metric: brisque # brisque (smaller is better), dbcnn (bigger is better), ilniqe (smaller is better)
34     sample: best # to take smaller or bigger values is decided depending the metric
35
36 wandb:
37   entity: sdcn-nantes
38   project: sdcn-shit-testing
39   download:
40     list_all: false
41     list_finished: true
42     list_running: false
43

```

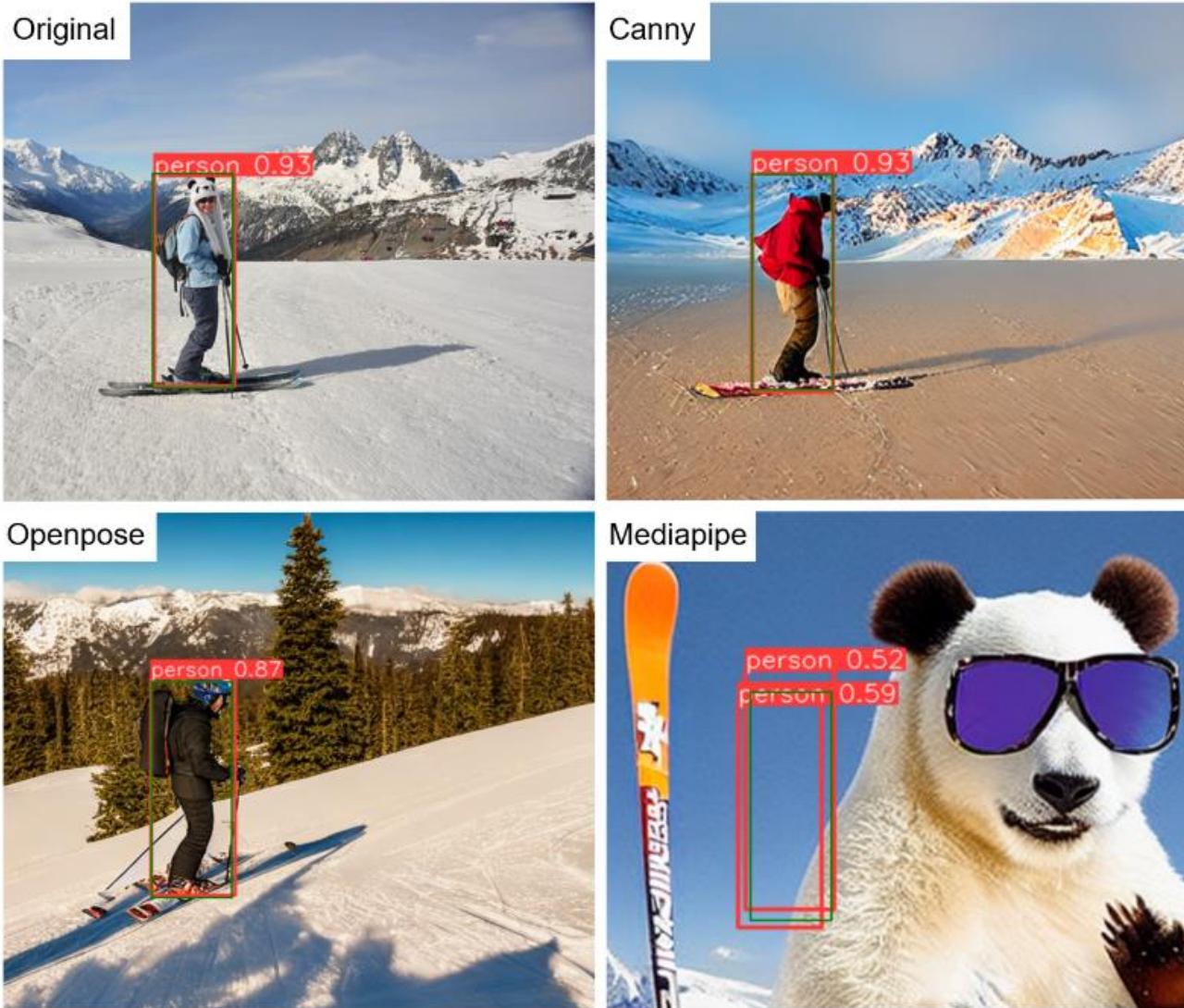
# CIA Framework : Prompt Loader / Generator

```

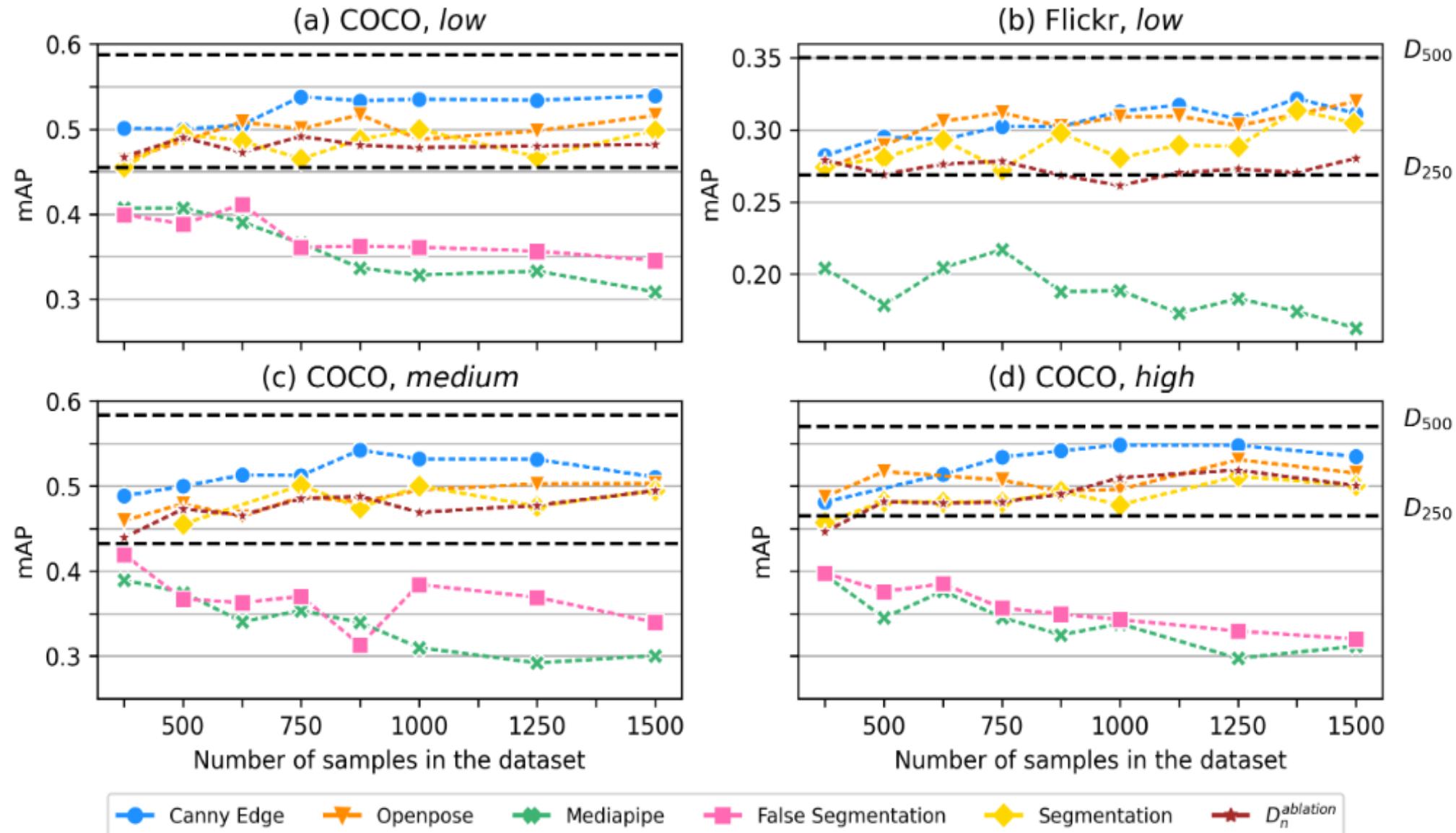
prompt:
  template: vocabulary
  modify_captions: 1
  generation_size: 10
  base: ["Sandra Oh", "Kim Kardashian", "rihanna ", "taylor swift"]
  quality: "showing emotion, great realistic face, best quality, extremely detailed,"
  modifier: "Happy man smiling"
  negative:
    [
      "monochrome, lowres, bad anatomy, worst quality, low quality, cartoon, unrealistic, bad proportion,",
      "distortion, bad quality, lowres, cropped, bad focus, blurry, ad compression, bad artifact,",
      "bad pixel, deformed iris, deformed pupils, semi-realistic, cgi, 3d, render, sketch, drawing, anime:1.4),",
      "close up, cropped, out of frame, jpeg artifacts, ugly, duplicate, morbid, mutilated,",
      "extra fingers, mutated hands, poorly drawn hands, poorly drawn face, mutation, deformed, blurry, dehydrated,",
      "extra limbs, cloned face, disfigured, gross proportions, malformed limbs, missing arms, missing legs,",
      "extra arms, extra legs, fused fingers, too many fingers, long neck, no cars, no people, illustration, painting,",
      "drawing, art, sketch, anime, deformation, distortion",
    ]
  negative_simple: "monochrome, lowres, bad anatomy, worst quality, low quality, cartoon, unrealistic, bad proportion, disfigure
  ...

```

# Tests & Results



# Tests & Results





## MeetIA - WP4



Thank you all for your attention!  
For questions :  
[Mohamed.benkedadra@umons.ac.be](mailto:Mohamed.benkedadra@umons.ac.be)