Deep Generative Modeling for Automated Image Manipulation by Interpreting Text-Guided Prompts with Natural Language Instructions

Principal Investigator

Er. Dinesh Baniya Kshatri

Assistant Professor

Institute of Engineering, Thapathali Campus

Undergraduate Students

Nishan Khanal

Prashant Raj Bista

Nimesh Gopal Pradhan

Abhinav Chalise

Institute of Engineering, Thapathali Campus

Collaborators

Dr. Madhav Prasad Ghimire

Associate Professor & Head

Advanced Material Research Laboratory, Central Department of Physics, Tribhuvan University

Bishartha Manandhar

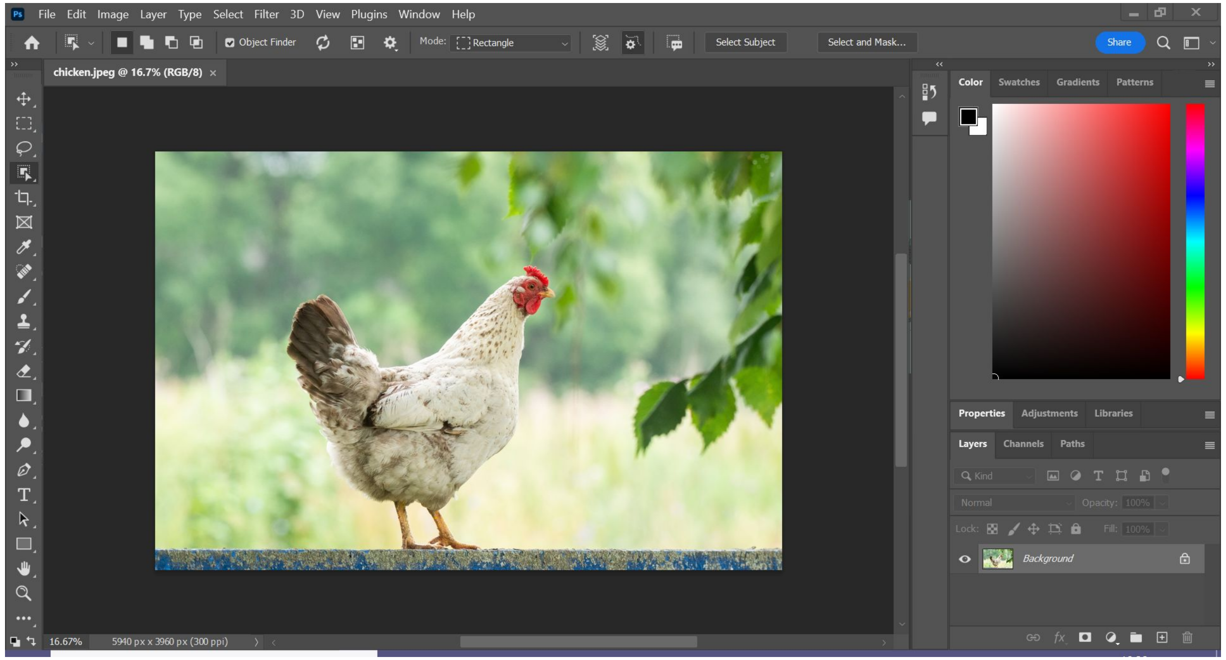
Graduate Student

Department of Physics, Tribhuvan University

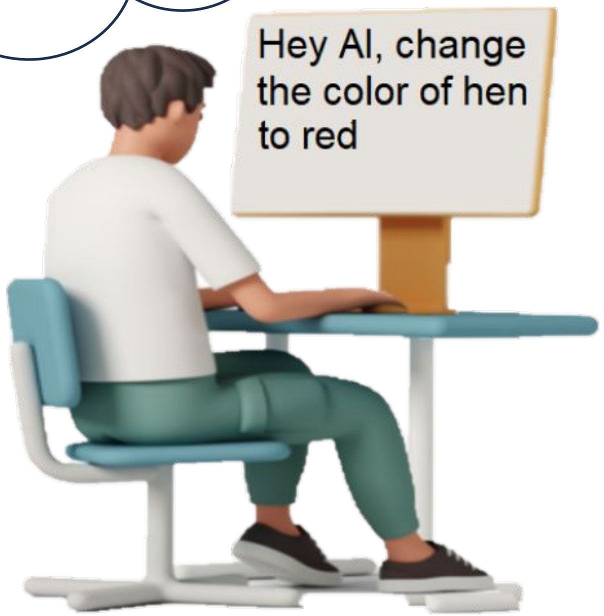
Presentation Outline

- Motivation
- Introduction
- Research Objectives
- Applications
- Scope of Research
- Literature Review
- Metrics Interpretation
- Few Results of Literature Review
- Dataset
- Research Gaps & Research Questions
- Theoretical Background
- Mathematical Formulation
- System Block Diagram
- Hardware and Software Requirements
- Concept Figures
- Expected Outputs
- Deliverables
- Budget Breakdown
- Gantt Chart
- Appendix
- References

Motivation



Learning editing tools like Photoshop requires time and effort



Introduction

- Simplify image editing by guiding edits with everyday language instructions.
- Empower non-experts to craft personalized visuals through simple text commands.
- Blend natural language processing and diffusion techniques for expressive image transformations.
- Simultaneously edit multiple objects within an image.
- Capable of undoing and redoing edits with minimal artifacts.

Research Objectives

- General objectives
  - Advance Image Editing With Text Prompt.
  - Introducing Multi-Object Editing through single prompt.

- Specific objectives
  - Editing on up to 3 multiple objects
  - Performing edits such as change in color, change in spatial location
  - Adding and Removal of foreign objects
  - Undoing and Redoing of Edits up to 2 levels.

Applications

- Content Creators
  - Visual content creators, can easily prototype ideas, generate visuals from descriptions, and explore new creative designs.
- Accessibility for Non-Experts
  - Individuals without advanced editing skills can easily modify their images.
- Filmmaking
  - Filmmakers can visualize scenes based on scripts, making pre-production more efficient and cost-effective.
- Photographers
  - Photographers can use this technology to correct errors or enhance images based on textual descriptions.

Scope of Research

- Capabilities
  - Performing edits on real and synthetic image
  - Perform editing such as changing color, spatial location for multiple objects.
  - Adding and removal of objects.
  - Undoing and redoing edits without creating image distortions, blurred objects.
- Limitations
  - Performing edits such as structural changes, resolution changes.
  - Undoing, redoing more than two levels.
  - Performing edits on more than 3 objects simultaneously

Literature Review - [1]
(Base papers)

Title	Year	Author
Iterative Multi-granular Image Editing using Diffusion Models	Sep 2023	K J Joseph, Prateksha Udhayanan, Tripti Shukla, Aishwarya Agarwal, Srikrishna Karanam, Koustava Goswami, Balaji Vasan Srinivasan
Negative-prompt Inversion: Fast Image Inversion for Editing with Text-guided Diffusion Models	May 2023	Daiki Miyake, Akihiro Iohara, Yu Saito, Toshiyuki Tanaka
Prompt Tuning Inversion for Text-Driven Image Editing Using Diffusion Models	May 2023	Wenkai Dong, Song Xue, Xiaoyue Duan, Shumin Han
Null-text Inversion for Editing Real Images using Guided Diffusion Models	Nov 2022	Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, Daniel Cohen-Or

Literature Review - [2]
(Comparative Analysis)

Title	Method	Results	Merits	Demerits
Iterative Multi-granular Image Editing using Diffusion Models	• InstructPix2Pix diffusion model • EMILIE (Iterative Multi-granular Image Editor)	• Designed IMIE-Bench metric for iterative editing. • Average CLIP score of 0.311 • Average BLIP score of 0.620	• Introduce concept of iterative editing in existing diffusion models editing capabilities in • Spatial control over edits	• Cannot effectively handle negative edits
Negative-prompt Inversion: Fast Image Inversion for Editing with Text-guided Diffusion Models	• Stable Diffusion • Negative-Prompt Inversion	• PSNR score of 23.38 ± 0.66 • LPIPS score of 0.1603 ± 0.0155 • Processing speed of 4.627 ± 0.020 seconds.	• About 30 times faster compared to other image editing methods	• Failure when reconstruction image of people
Prompt Tuning Inversion for Text-Driven Image Editing Using Diffusion Models	• Stable Diffusion • Prompt Tuning Inversion	• PSNR score of 25.71 • SSIM score of 0.8501	• Generalizable to large domain • No need of source-prompt or mask	• Fails when editing multiple objects in image
Null-text Inversion for Editing Real Images using Guided Diffusion Models	• Stable Diffusion • Null-text inversion	• User Evaluation used as metric • 61.7% users choose their method	• Introduction of new inversion method • High fidelity to original image	• High Inference time • Artifacts in human faces

Metrics Interpretation - [1]

- CLIP Score (Contrastive Language Image Pre-Training)
  - Evaluate alignment between text and image
  - Calculated Using Cosine Similarity
  - $Cosine\ Similarity = \frac{A.B}{||A||||B||}$
  - $A = embedding\ vector\ of\ image,\ B = embedding\ vector\ for\ text$
  - The cosine similarity score ranges from -1 to 1
  - Score close to 1 indicates high degree of similarity, 0 implies little to no relevance, close to -1 implies high degree of dissimilarity

Metrics Interpretation - [2]

- BLIP (Brier Score Improvement Percentage)
  - Evaluates improvement of predictive model over baseline model
  - BLIP score is calculated using Brier score
  - Brier Score is mean squared difference between predicted probability and actual outcome
  - $BLIP = \left(1 - \frac{BS_{model}}{BS_{baseline}}\right) \times 100\%$
  - $BS(Brier\ Score) = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$
  - $BS_{model} = Brier\ score\ of\ model,\ BS_{baseline} = Brier\ score\ of\ baseline\ model$
  - $N = number\ of\ predictions,\ p_i = predicted\ probability,\ o_i = actual\ outcome$
  - A positive BLIP score indicates improvement over baseline model and negative indicates model performs worse than the baseline

Metrics Interpretation - [3]

- PSNR (Peak Signal to Noise Ratio)
  - Measures quality of reconstructed or compressed image
  - Calculated Using Mean Squared Error (MSE)
  - $PSNR = 20 \times \log_{10}(\frac{MAX_I}{\sqrt{MSE}})$
  - $MSE = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}[I(x,y) - K(x,y)]^2$
  - $MAX_I = Maximum\ possible\ pixel\ value\ of\ the\ image$
  - $M = Width\ of\ image$, $N = Height\ of\ image$
  - $x, y = pixel\ coordinate$, $I = Intensity\ of\ original\ Image$
  - $K = intensity\ of\ reconstructed\ or\ compressed\ image$
  - High PSNR indicates better quality

Metrics Interpretation - [4]

- SSIM (Structural Similarity Index Measure)
  - Measures similarity between images
  - SSIM is calculated for small overlapping windows
  - $SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$
  - $x, y = windows\ in\ image\ 1\ and\ image\ 2$
  - $\mu_x,\ \mu_y = average\ pixel\ intensities\ of\ window\ x\ and\ y$
  - $\sigma_x^2,\ \sigma_y^2 = variances\ of\ pixel\ intensities\ in\ window\ x\ and\ y$
  - $\sigma_{xy} = covariance\ of\ windows\ x\ and\ y$
  - $C_1 = (k_1 L)^2, C_2 = (k_2 L)^2 = constants\ to\ avoid\ instabiliy\ due\ to\ small\ denominator$
  - $L = dynamic\ range\ of\ pixel\ values, k_1 = 0.01,\ k_2 = 0.03$
  - SSIM value ranges from 1 to -1 with 1 indicating perfect similarity and -1 indicating dissimilar images

Metrics Interpretation - [5]

- LPIPS (Learned Perceptual Image Patch Similarity)
  - Uses deep learning to measure perceptual similarity between images
  - $LPIPS(A, B) = \sum_l D^l(A, B)$
  - $D^l(A, B) = \ w_l \times \left\| \hat{F}_A^l \hat{F}_B^l \right\|_2^2$
  - $\hat{F}_A^l = \frac{F_A^l}{\left\|F_A^l\right\|_2}, \hat{F}_B^l = \frac{F_B^l}{\left\|F_B^l\right\|_2}$
  - $A, B = given\ images$
  - $F_A^l,\ F_B^l = feature\ maps\ at\ l^{th}\ layer\ of\ network\ for\ image\ A\ and\ B$
  - $\hat{F}_A^l, \hat{F}_B^l = L2\ normalized\ feature\ maps$
  - LPIPS generally ranges from 0 to 1 but can exceed 1 depending on deep neural network used
  - Score close to 0 indicates images are perceptually indistinguishable
  - Score close to 1 indicates images are perceptually very different

Metrics Interpretation - [6]

- Normalized Blur Metric (NBM)
  - Metric used to quantify the amount of blur in an image.
  - It considers both blur and contrast information.
  - It is calculated based on the image gradients.
  - $NBM = \frac{1}{N}\sum_{i=1}^{N}\frac{G_i}{C_i}$
  - $N = total\ number\ of\ pixels\ in\ the\ image$
  - $G_i = gradient\ magnitude\ at\ pixel\ i$
  - $C_i = local\ contrast\ at\ pixel\ i$
  - G_i represents strength of local gradient
  - C_i measures the difference in intensity between the pixel and its surroundings.
  - High value of NBM indicates high blur and vice versa

Metrics Interpretation - [7]

- Normalized Cross-Correlation
  - It is used to identify halos in images.
  - Halos are bands of light that follow edges in an image.
  - Halos often occur as artifacts when an object is added or removed from an image.
  - $NCC(I,J) = \frac{\sum_{i,j}(I(i,j)-\bar{I})(J(i,j)-\bar{J})}{\sqrt{\sum_{i,j}(I(i,j)-\bar{I})^2 \sum_{i,j}(J(i,j)-\bar{J})^2}}$
  - $I(i,j), J(i,j) = Pixel\ values\ at\ location\ (i,j)\ in\ Images\ I\ and\ J$
  - $\bar{I}, \bar{J} = mean\ pixel\ values\ of\ images\ I\ and\ J$

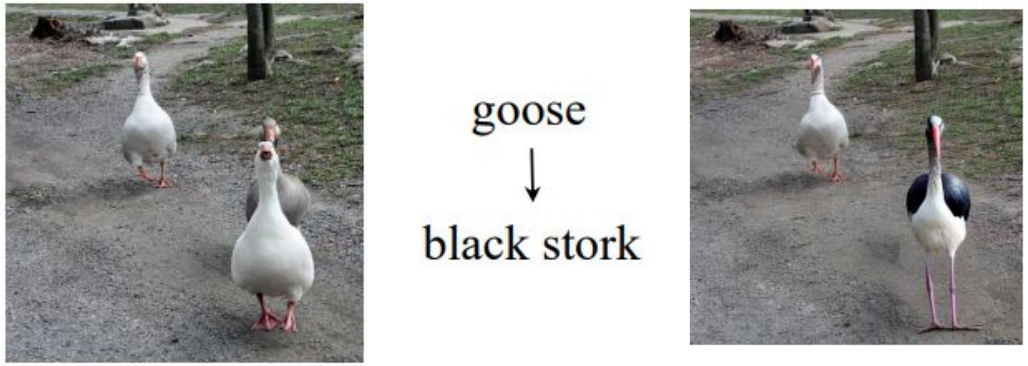
Metrics Interpretation - [8]

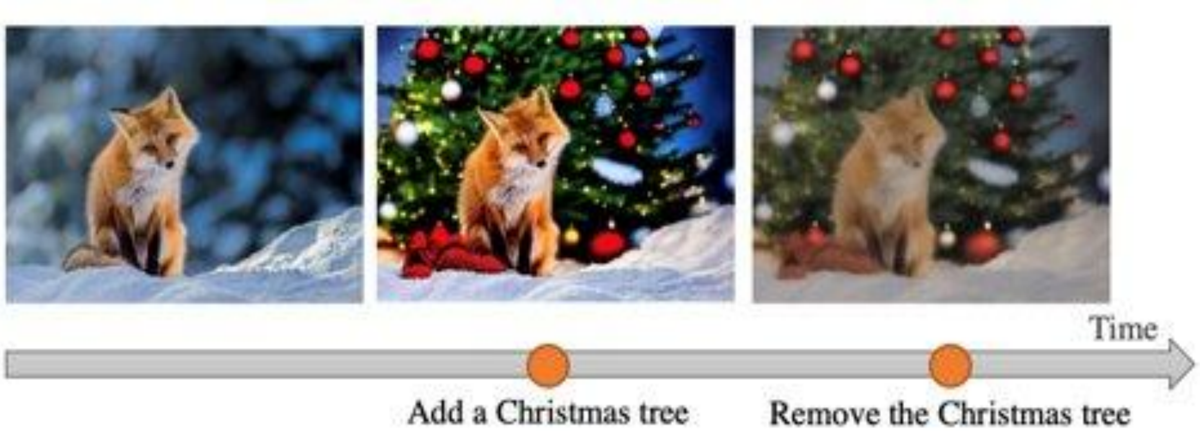
- Normalized Cross-Correlation
  - The range of the NCC metric is from -1 to 1.
  - A value of 1 indicates a perfect match or similarity between the two compared images.
  - A value of 0 indicates no correlation or similarity.
  - A value of -1 indicates perfect dissimilarity.

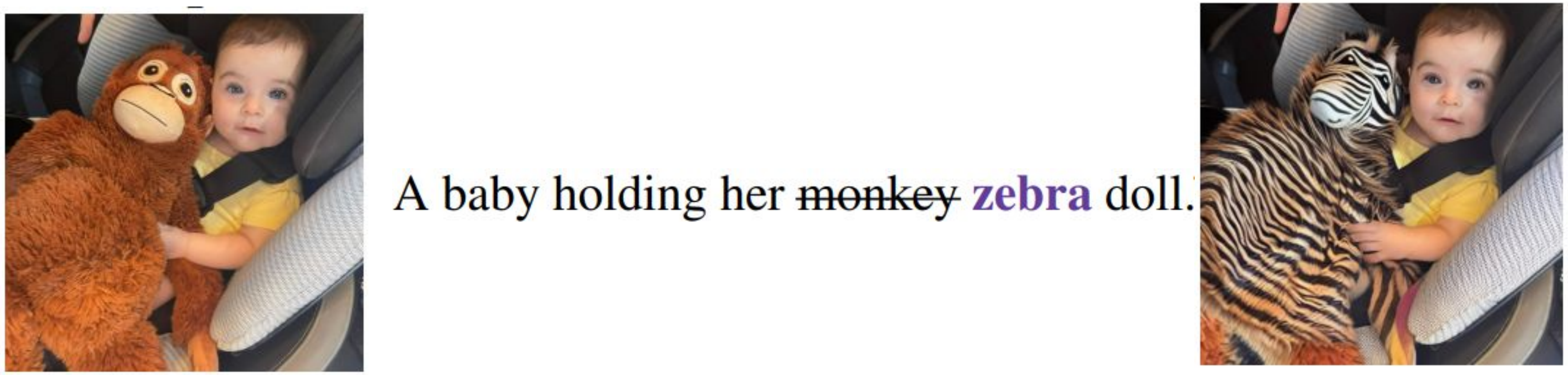
Few Results of Literature Review

Input Image

Output Image







Input Image

Output Image

Dataset

- COCO (Common Objects in Context)
  - Large-scale image recognition dataset
  - Used for object detection, segmentation, captioning tasks
  - Contains over 33,0000 images, 80 object categories and 5 caption per image
  - Among 80 objects categories some categories are car, bird, motorcycle
  - Number of images in categories car, bird, motorcycle are 12786, 3362, 3661 respectively

Dataset
(Sample Images)

Image of motorcycles

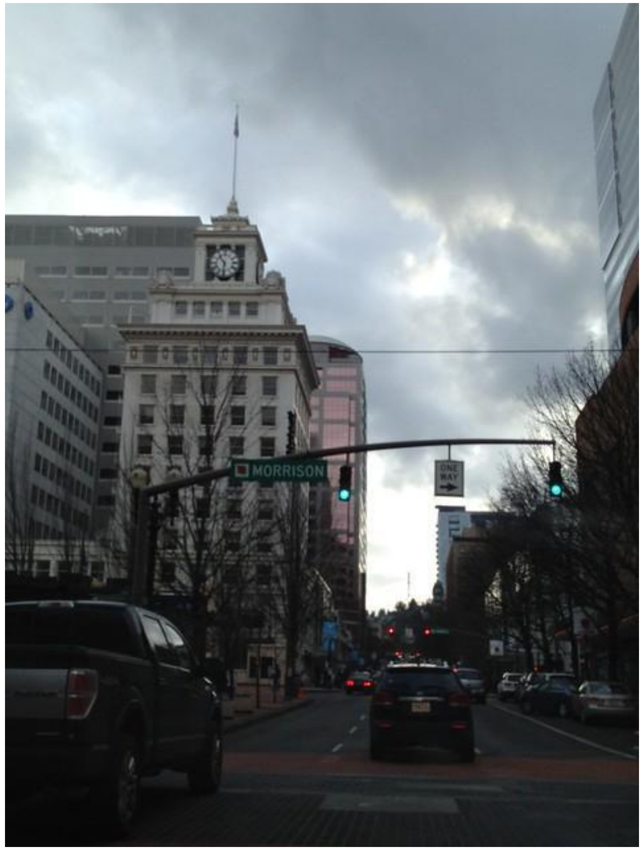


Image of cars

Image of birds

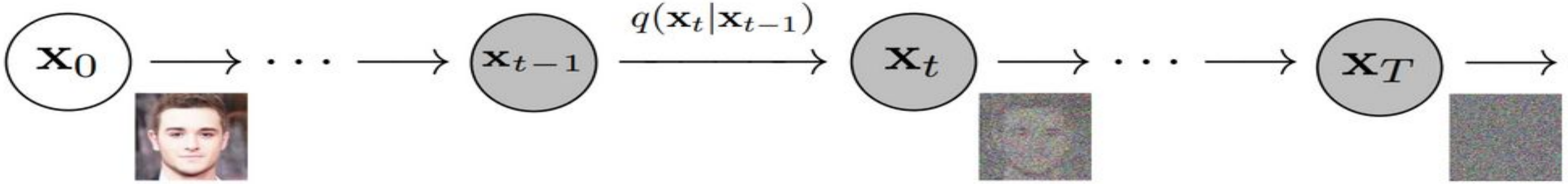
Research Gaps & Research Questions

- Research Gaps
  - Handling negative edits
  - Undoing and redoing of edits without image distortion, blurring
  - Editing multiple objects simultaneously
  - Editing on human faces without distortion
- Research Questions
  - Rq1: How can diffusion model be used to effectively manipulate multiple objects within a single editing cycle?
  - Rq2: What innovative approach can be developed to enhance user experience by allowing undoing and redoing of edits?

Theoretical Background of Diffusion - [1]

- Diffusion is the process derived from thermodynamics.
- Diffusion is the movement of molecules from a region of higher concentration to a region of lower concentration
- In diffusion models, this concept is extended to add noise randomly to the regular image and get noise.
- The model will learn the probability of the added noise to remove it.

Theoretical Background of Diffusion – [2]
(Forward Diffusion)



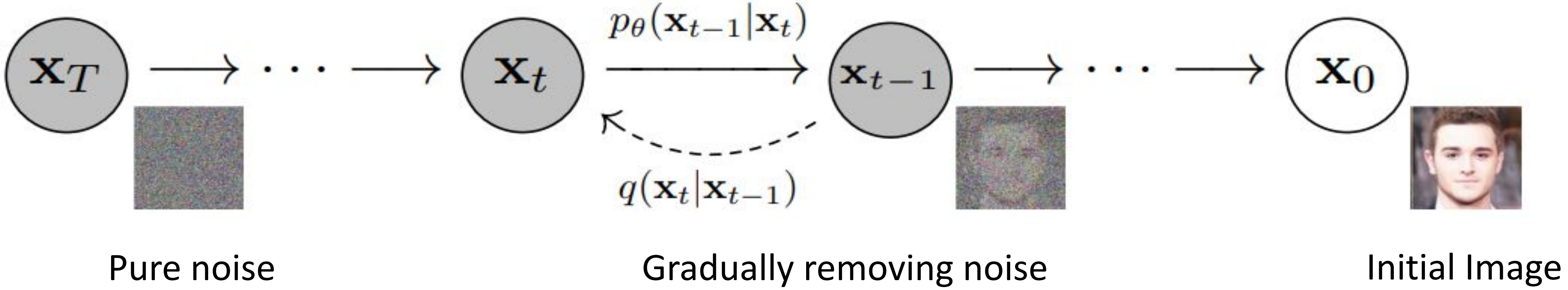
Initial image

Gradually adding noise to image

Pure noise

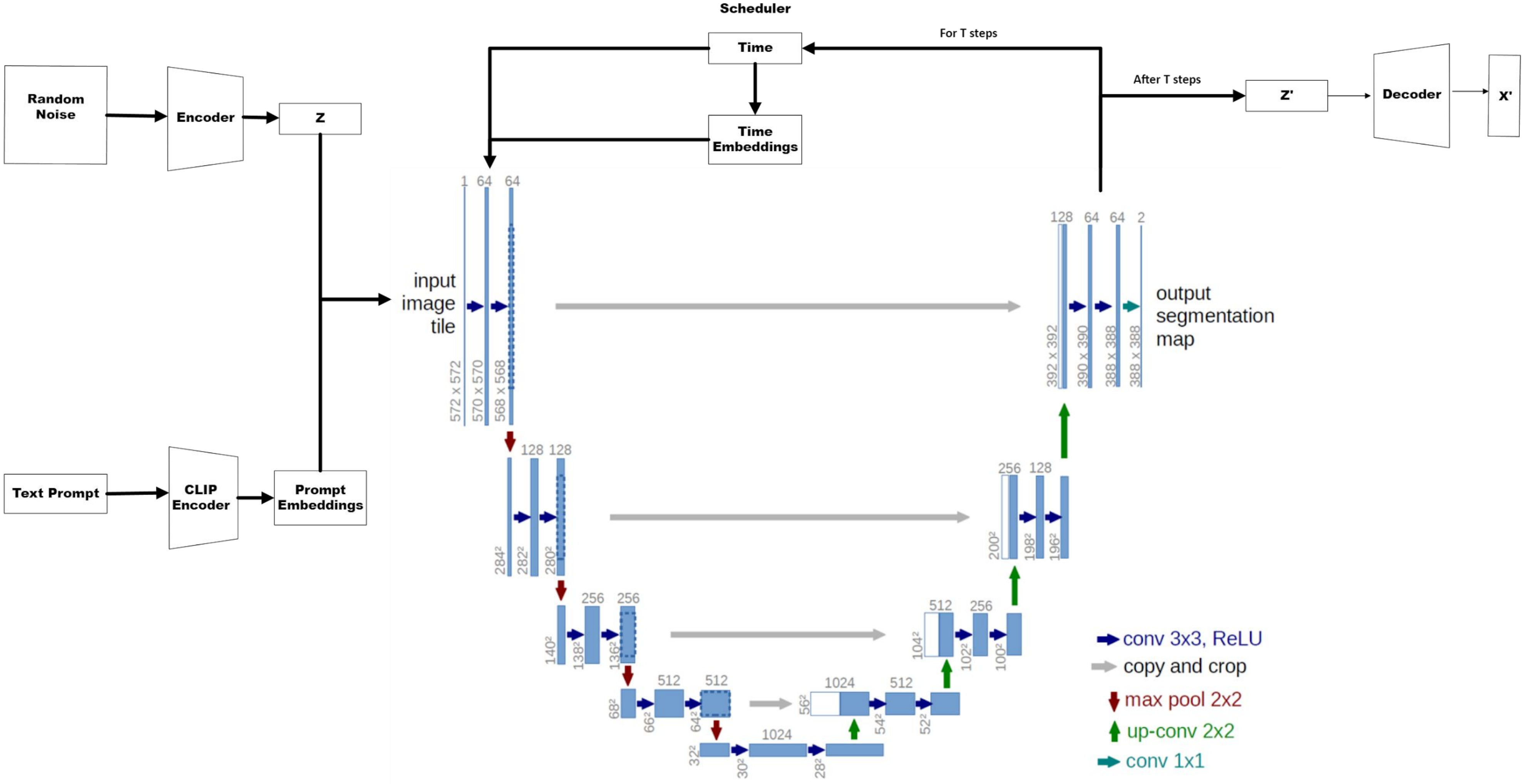
- The forward diffusion process involves gradually corrupting an image with noise.
- This process is done in stages, over a series of steps called "timesteps"
- At each timestep, a small amount of Gaussian (random) noise is added to the image.
- The original image is transformed into pure, unstructured noise.
- This transformation is done in a controlled way so that the model can learn each step of the process.

Theoretical Background of Diffusion – [3] (Backward Diffusion)



- The primary goal of the diffusion model is to learn how to reverse the forward diffusion process.
- The model starts with noise and gradually removes it, step by step,
- The model reconstructs the original image or generate new ones.
- The model is trained to predict the noise that was added at each step of the forward process and then to reverse this by subtracting this noise.

Guidance for Diffusion Process - [1] (Architecture)



Guidance for Diffusion Process - [2]

- Text Prompt is used to guide diffusion process
- Text is converted to embedding using Contrastive Language–Image Pre-training Encoder
- Diffusion uses Classifier Free Guidance
  - Classifier-Free Guidance is a training technique where a generative model is steered by text input alone, without a separate classifier network
  - During Reconstruction step, one image is generated using noise only
  - Another image is generated using same noise as well as the text embedding
  - The two images are combined to generate final image

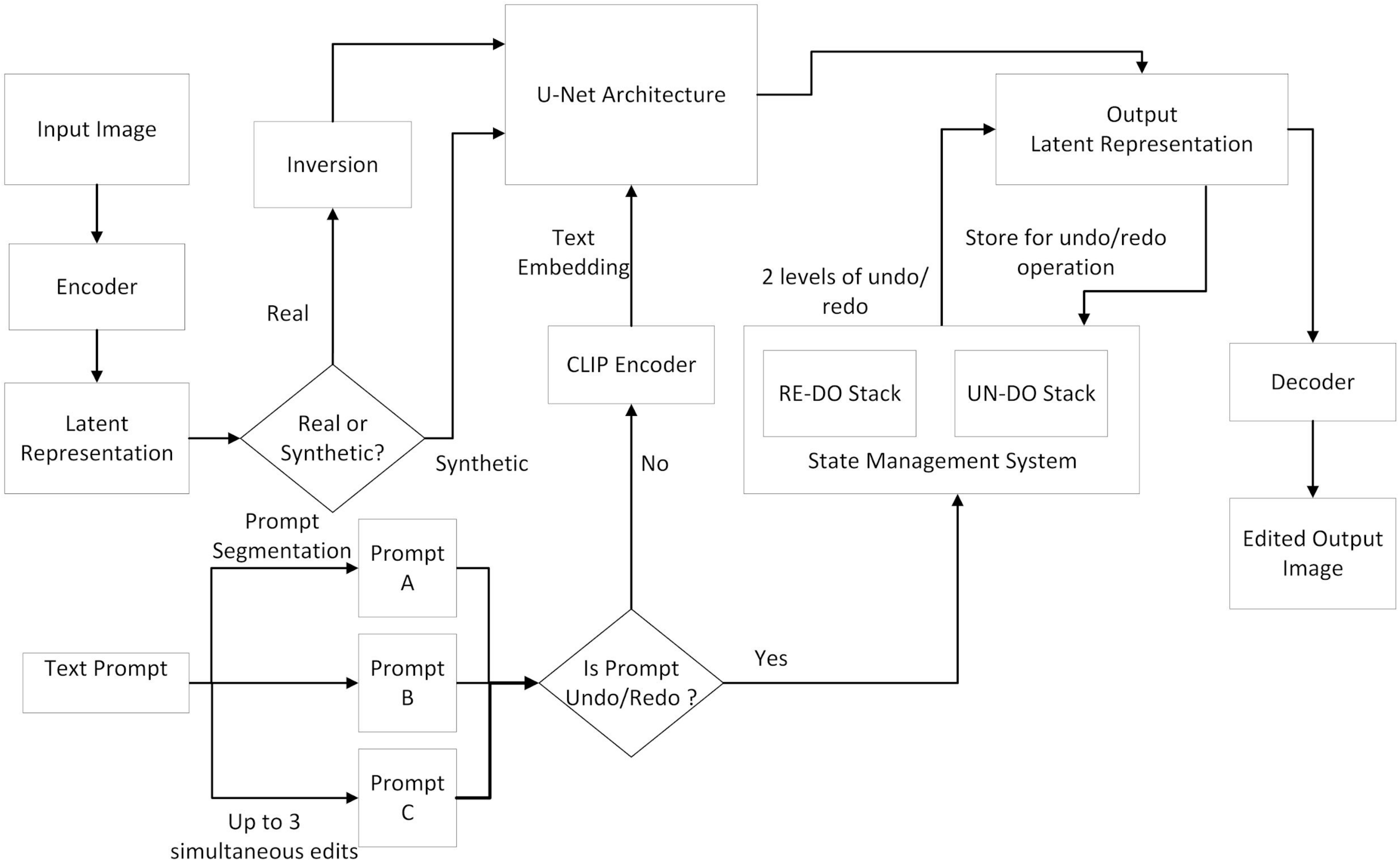
Mathematical Formulation - [1]

- Forward Diffusion
  - $q(x_t|x_{t-1}) = N\left(x_t; \mu_t = \sqrt{1-\beta_t}x_{t-1}, \Sigma_t = \beta_t I\right)$
  - $x_t = data\ at\ time\ t,\ q(x_t|x_{t-1}) = distribution\ model\ of\ data,\ N = Normal\ distribution$
  - $\mu = mean,\ \Sigma = vairance,\ I = identitymatrix,\ \beta_t = standard\ deviation,\ t = timestep$
  - $q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$
  - $x_t \sim q(x_t|x_0) = N(x_t; \sqrt{\bar{\alpha}_t}\, x_{0,}(1-\bar{\alpha}_t)I)$
  - $\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \Pi_{s=0}^{t}\alpha_s$
- Backward Diffusion
  - $p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t t), \Sigma_\theta(x_t, t))$
  - $p_\theta(x_{0:T}) = p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)$
  - $p_\theta\ (x_{0:T}|C) = p_\theta\ (x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t, C)$
  - $p_\theta(x_{t-1}|x_t) = parameterized\ distribution\ model$
  - $p_\theta\ (x_{0:T}|C) = parameterized\ conditional\ distribution\ model,$
  - $C = conditioning\ information,\ \mu_\theta(x_t t) = predicted\ mean,\ \Sigma_\theta(x_t, t) = predicted\ variance$

Mathematical Formulation - [2]

- Classifier Free Guidance
  - $\tilde{\epsilon}_\theta(x_t, t, C, \emptyset) = \ (1 - w)\epsilon_\theta(z_t, t, \emptyset) + w\epsilon_\theta(z_t, t, C)$
  - $w = classifier\ free\ guidance$
  - $\epsilon_\theta(z_t, t, \emptyset) = predicted\ noise\ without\ prompt$
  - $\epsilon_\theta(z_t, t, C) = predicted\ noise\ with\ prompt,$
  - $\tilde{\epsilon}_\theta(x_t, t, C, \emptyset) = final\ predicted\ noise$
  - $C = conditional\ text\ embedding$
  - $\emptyset = null\ text\ embedding$

System Block Diagram



Hardware and Software Requirements

Hardware Required	Reason
Nvidia A100 Tensor Core GPU	Accelerated parallel processing, matrix multiplications, efficient deep learning tasks due to dedicated Tensor Cores
1-5 TB Storage	Storing and managing data generated during iterative image editing tasks
50+ GB Memory	Accurate handling of stochastic process and accommodating high-dimensional data manipulation

Software Required	Reason
Pytorch, TersorFlow	TensorFlow excels in scalability and deployment, ideal for production, whereas PyTorch offers a dynamic, user-friendly interface, favored for research.
Natural Language Translation Toolkit, Python	Processing and understanding textual prompts
Django,React	Creating a user-friendly web interface for image editing

Hardware Used in Papers

Papers	Hardware Used
Iterative Multi-granular Image Editing using Diffusion Models	Single NVIDIA A100 GPU
Negative-prompt Inversion: Fast Image Inversion for Editing with Text-guided Diffusion Models	NVIDIA RTX A6000 GPU and an AMD EPYC 7343 CPU (16 cores, 3.2 GHz clock speed)
Prompt Tuning Inversion for Text-Driven Image Editing Using Diffusion Models	Single Tesla V100 GPU
Null-text Inversion for Editing Real Images using Guided Diffusion Models	Single A100 GPU

Available Hardware - [1]
(Specifications of available hardware)

- Hardware at Advanced Materials Research Laboratory

GPU	GPU Memory	CUDA Core count	Tensor core count	RAM
NVIDIA A100 Tensor Core GPU	160 GB	27,648	1,728	DDR4, 512 GB
NVIDIA RTX 6000	24 GB	4,608	576	DDR6, 24 GB

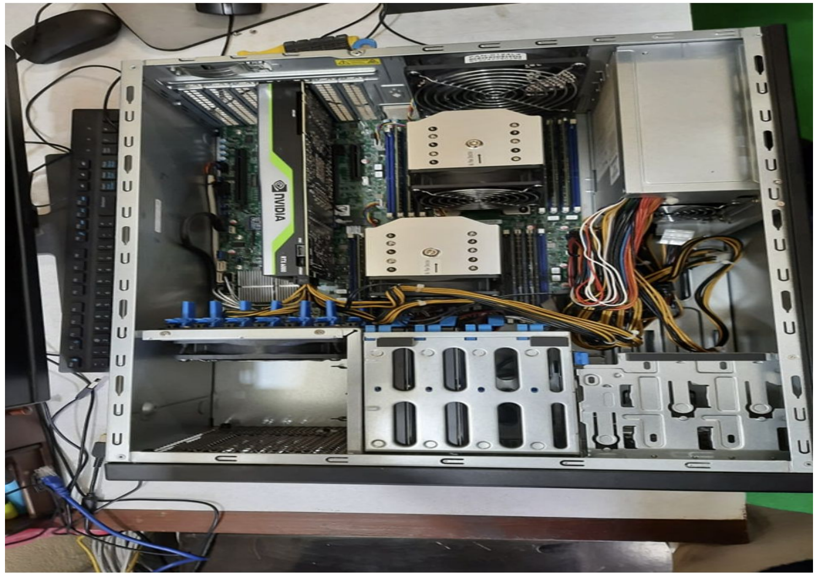
- Hardware at Thapathali Campus

Processor	Memory	Storage	Cores	Threads	Graphics	Operating System
9th Generation Intel Core i5 Processor	8 GB	1 TB	4 cores	8 threads	Intel UHD Graphics 630	Windows 10 Pro
9th Generation Intel Core i3 Processor	4 GB	500 GB	4 cores	4 threads	Intel UHD Graphics 630	Windows 10 Pro

Available Hardware - [2]
(Pictures of resources at AMRL)

Workstation at AMRL, CDP, TU

Stacks of servers at AMRL, CDP

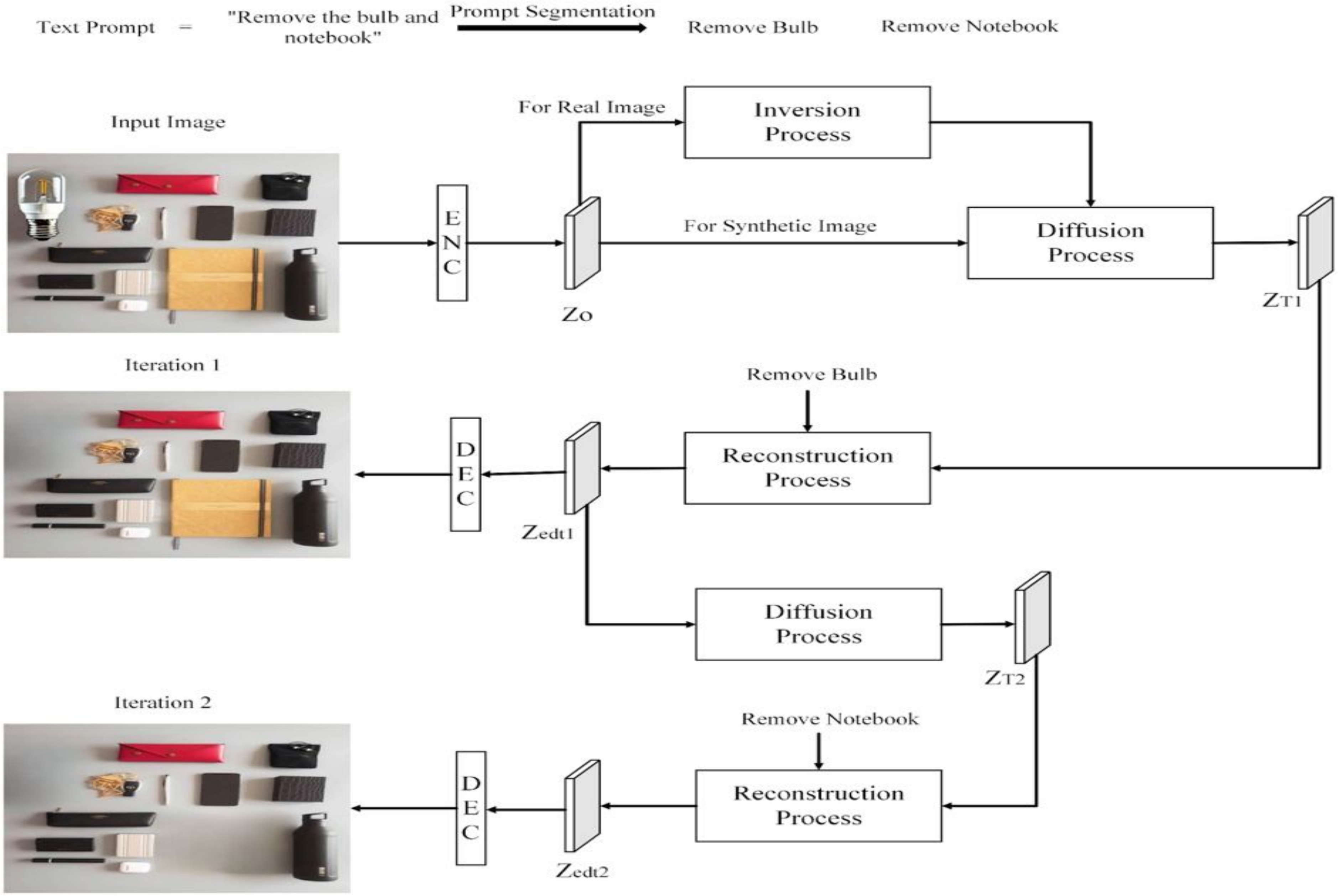


Workstation with NVIDIA RTX 6000 at AMRL, CDP

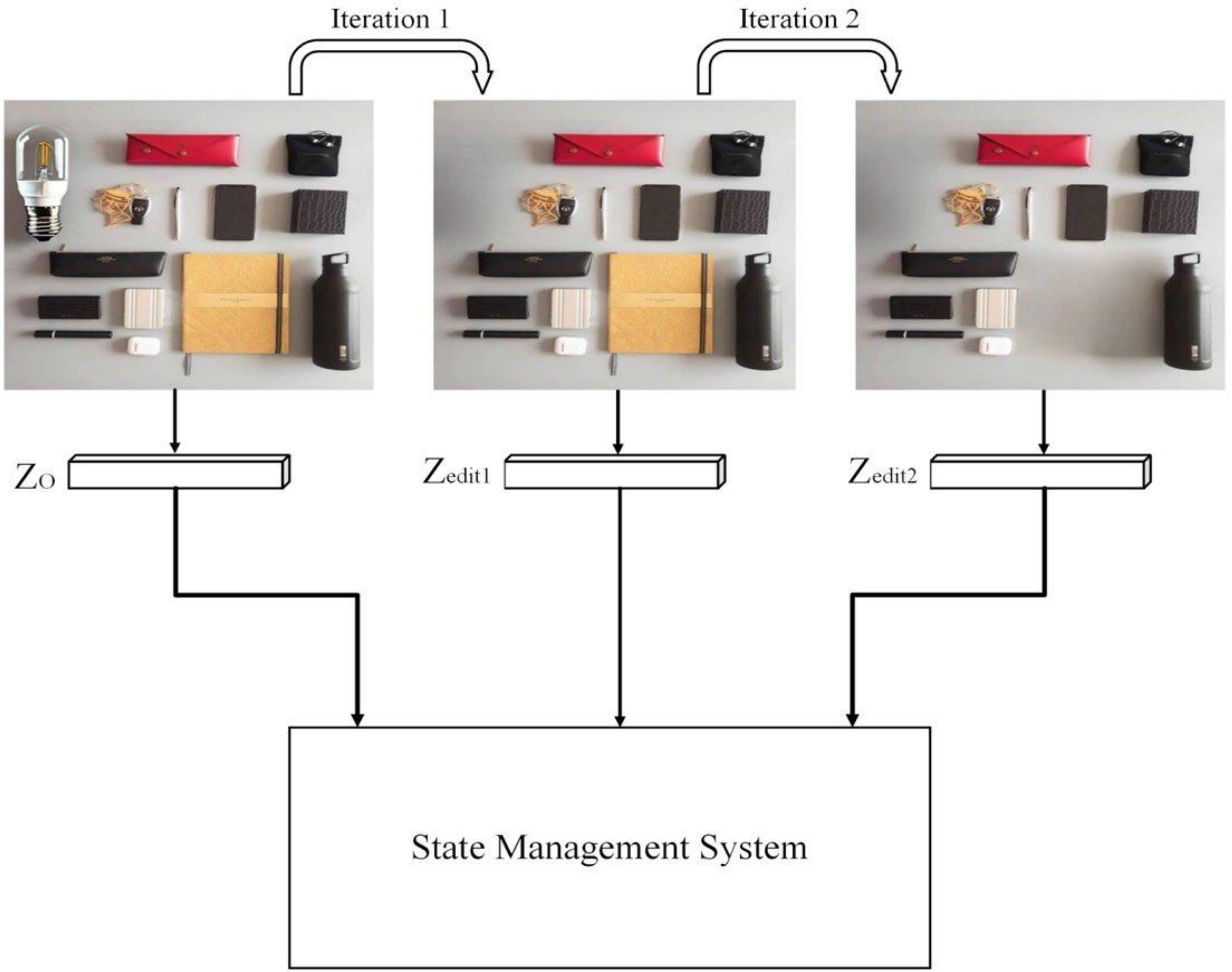
Available Hardware - [3]
(Picture of resources at Thapathali Campus)

Computer Lab at Thapathali Campus, IOE, TU

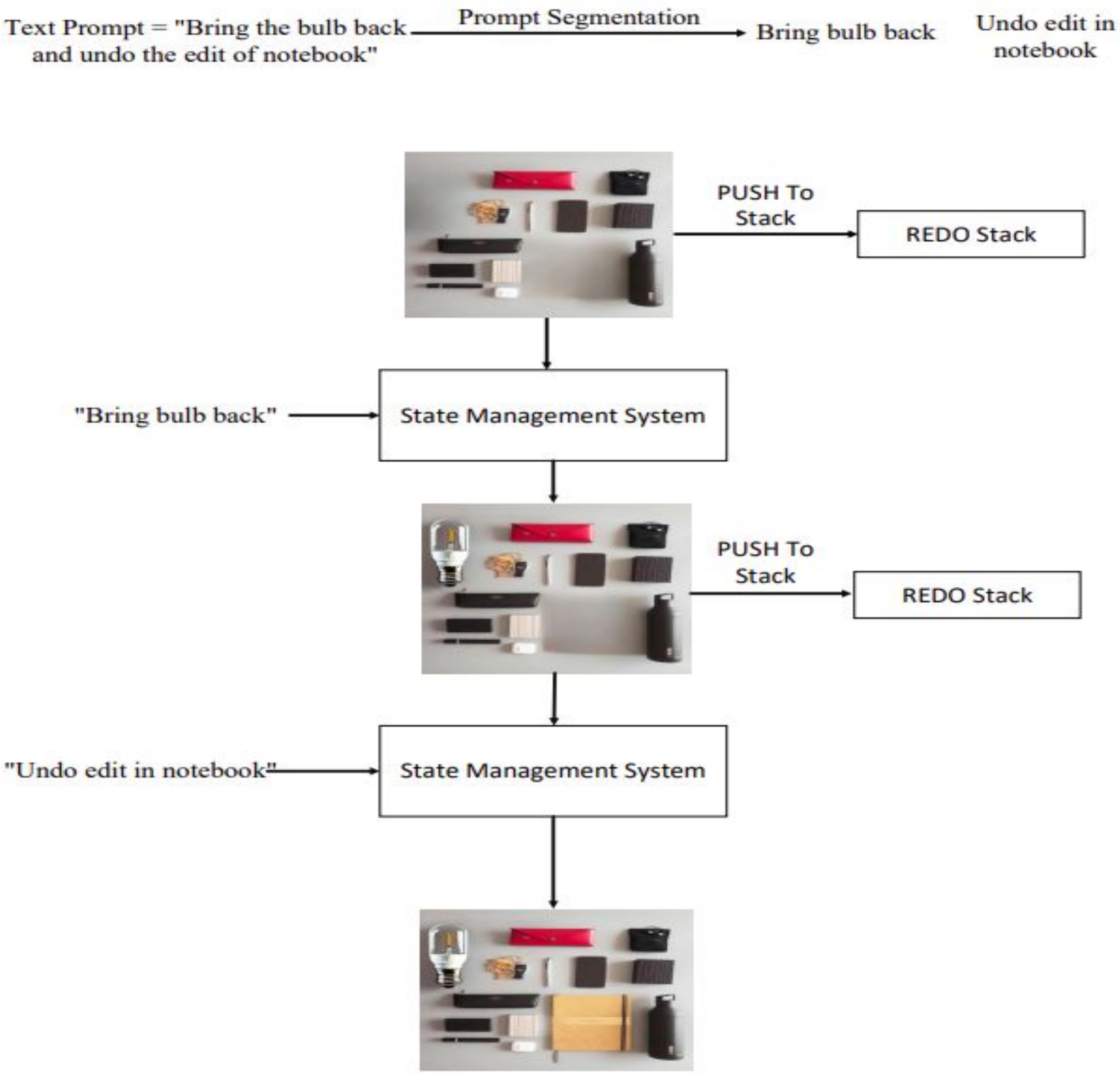
Concept Figure - [1]
(Editing of Multiple Object)



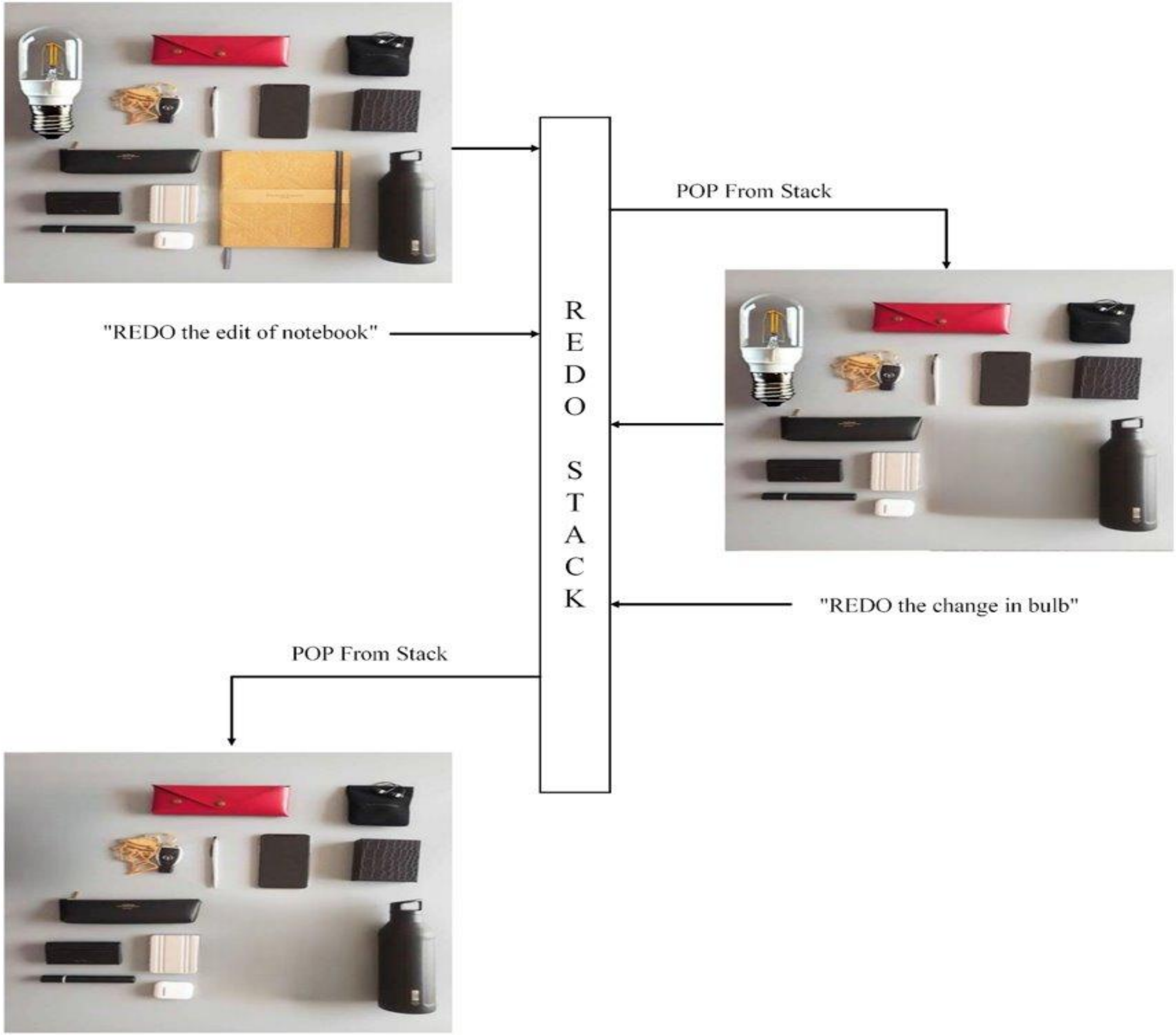
Concept Figure - [2]
(State Management System)



Concept Figure - [3] (Undoing Edits)



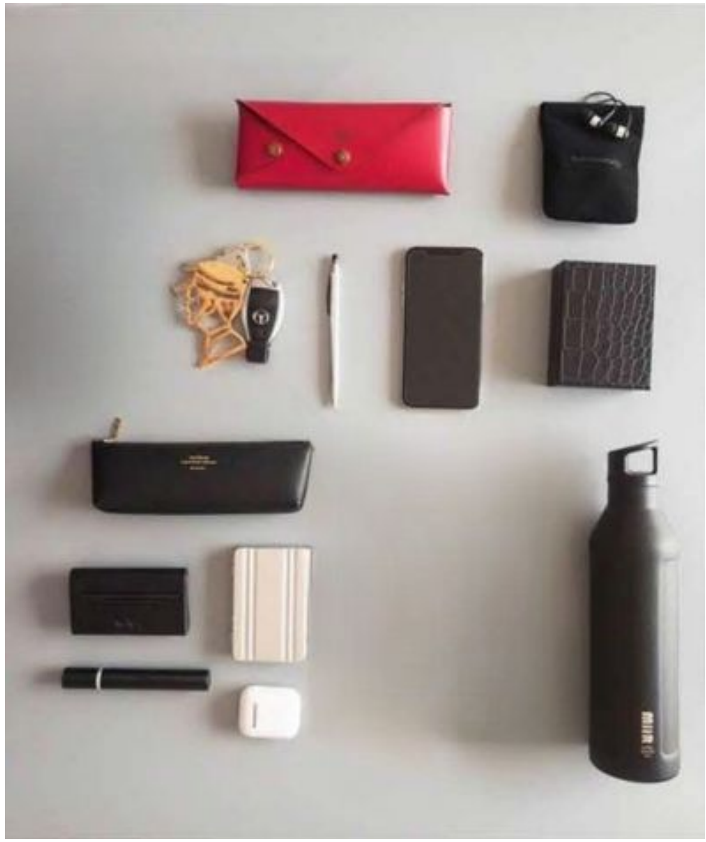
Concept Figure - [4] (Redoing Edits)



Expected Outputs - [1]
(Removing Objects)

Input Image

Text Prompt: Remove Bulb and Notebook



Output Image

Expected Outputs - [2]
(Changing Colors)

Input Image

Text Prompt: Change the color of purse to blue and notebook to green

Output Imge

Expected Outputs - [3]
(Changing Spatial Position)

Input Image

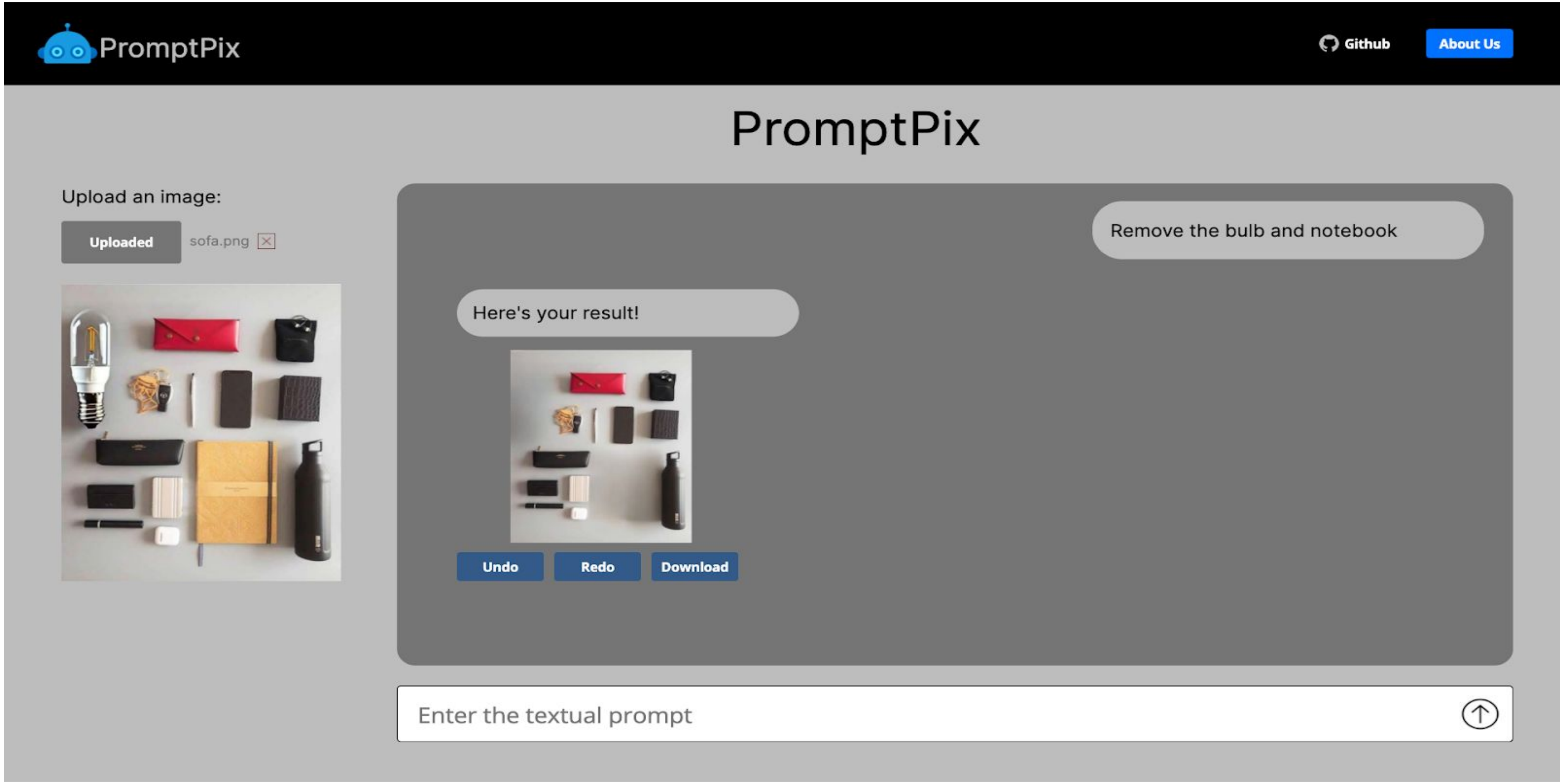
Text Prompt: Move the purse to bottom left

Output Image

Deliverables - [1]

- Web interface
  - An intuitive web interface for editing the input image using edit prompt.
- Journals and Conferences
  - Submit paper to NJSE for national exposure
  - Attend various conferences to present our paper

Deliverables - [2]
(Web Interface)



Journals and Conference

- Submit paper to NJSE for foundational national recognition;
- Attend IOE and Thapathali graduate conferences.
- Post-NJSE, enhance research for submission to top international AI and Computer Science journals.
- Target journals like IJPRAI and Procedia Computer Science, focusing on open access publication.
- Utilize Procedia's free policy and high CiteScore for global dissemination and impactful research presentation.

Responsibilities - [1]
(Sub-Objective 1)

Name of Researcher	Education Level	Responsibilities
Abhinav Chalise	Undergraduate	Efficient Implementation and Integration
Nimesh Gopal Pradhan		
Nishan Khanal		Performance Optimization and Robustness Testing
Prashant Raj Bista		Mathematical Modeling For simultaneous edits.

Responsibilities - [2] (Sub-Objective 2)

Name of Researcher	Education Level	Responsibilities
Abhinav Chalise	Undergraduates	Implementing in the program of Undo and Redo Functionalities
Nishan Khanal		Implementation of Iterative Feedback system
Nimesh Gopal Pradhan		Mathematical Modeling for simultaneous edits.
Prashant Raj Bista		

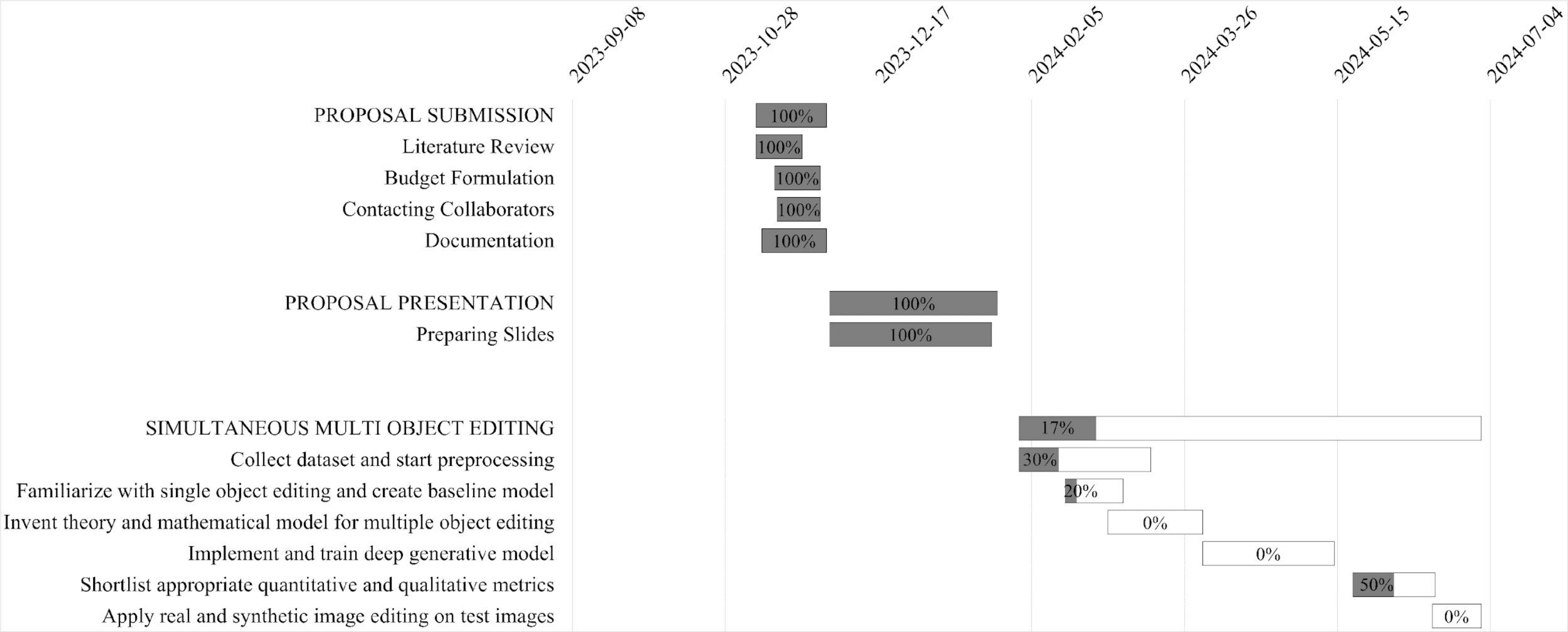
Responsibilities

Name	Responsibilities
Dr. Madhav Ghimire	Providing resources of ARML lab, CDP, TU.
Bishartha Manandhar	Helping with Mathematical derivations and formulas.

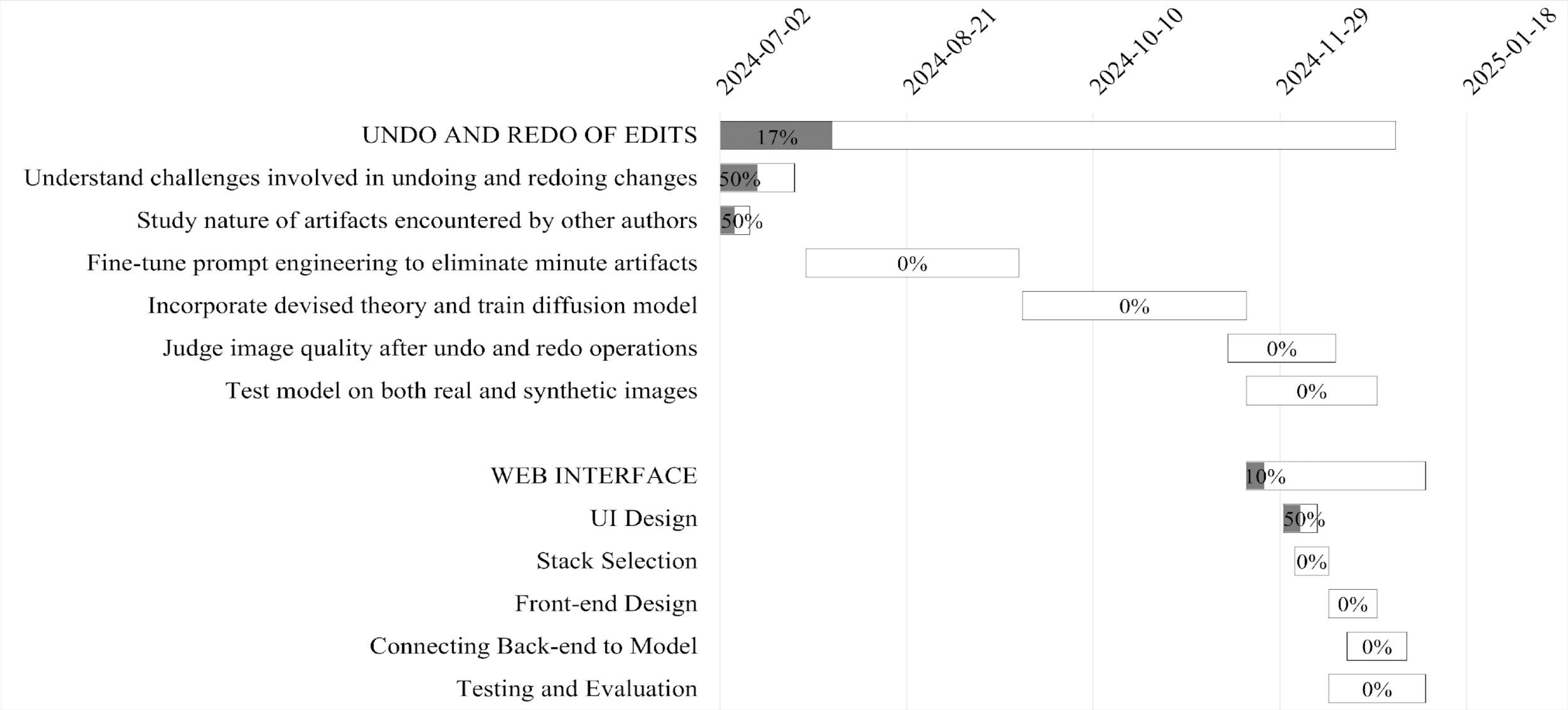
Budget Breakdown

Budget Plan	Percentage	Allocated Amount
Personnel salaries and stipends	30%	Rs. 60,000
Research materials and equipment	30%	Rs. 60,000
Travel and accommodation for conferences and collaborations	20%	Rs. 40,000
Data collection and analysis and publication costs	10%	Rs. 20,000
Other justifiable research expenses	10%	Rs. 20,000
Total	100%	**Rs. 2,00,000**

Gantt Chart - [1]



Gantt Chart - [2]



Appendix - [1] (Letter Of Intent to Collaborate From AMRL, CDP, TU)

TRIBHUVAN UNIVERSITY
CENTRAL DEPARTMENT OF PHYSICS
Kirtipur, Kathmandu, Nepal

☎ 4331054
www.tucdp.edu.np

Ref. No.: (F.No ....) CDP ....

Date: 29-11-2023

Tribhuvan University Central Department of Physics Kirtipur

To Whom It May Concern

This is to confirm that the Advanced Materials Research Laboratory at the Central Department of Physics, Tribhuvan University, is collaborating with the Electronics and Computer Community Amidst Students (ECAST) research unit at the Institute of Engineering, Thapathali Campus in the upcoming research project entitled **"Deep Generative Modeling for Automated Image Manipulation by Interpreting Text-Guided Prompts with Natural Language Instructions."**

After fruitful discussions aimed at fostering cooperation in research endeavors, both parties have identified areas of mutual interest. The purpose of this collaboration is to explore opportunities that may include but are not limited to, (a) the provision and sharing of high-performance computational resources, specifically the NVIDIA A100 Tensor Core GPU, and (b) Joint participation in research activities.

The Advanced Materials Research Laboratory at the Central Department of Physics acknowledges and agrees to provide the necessary hardware resources for the successful execution of the collaborative research initiatives.

This letter of intent is valid for the duration from 01/01/2024 to 14/07/2025 and is subject to the approval and endorsement of the authorities within the Central Department of Physics.

Madhav Prasad Ghimire, PhD
Associate professor &
Head
Advanced Materials Research Laboratory
email: madhav.ghimire@cdp.tu.edu.np

Appendix - [2]
(Proof of Affiliation for Faculty)

त्रिभुवन विश्वविद्यालय
Tribhuvan University
इन्जिनियरिङ्ग अध्ययन संस्थान
Institute of Engineering
थापाथली क्याम्पस
THAPATHALI CAMPUS

GPO Box-280, Thapathali, Kathmandu
Tel: 4-246465, 4218300, Fax: 977-1-4247340
E-mail: info@tcioe.edu.np
Website: www.tcioe.edu.np
गोश्वारा पो.ब.नं. २८०, थापाथली, काठमाडौं
फोन-४२४६४६५, ४२४६३०७
फ्याक्स: ९७७-१-४२४७३४०

Date:- 28/11/2023

To Whom It May Concern

This is to confirm that Mr. Dinesh Baniya Kshatri is currently affiliated with the Institute of Engineering, Thapathali Campus in Kathmandu, Nepal.

Mr. Dinesh Baniya Kshatri is affiliated as an Assistant Professor in the Department of Electronics and Computer Engineering.

We fully support Mr. Kshatri in his application for the present research grant call of the National College of Engineering, Tribhuvan University, Nepal.

Kind regards,

Asst. Prof. Dr. Khem Gyanwali
Campus Chief

Appendix - [3] (Proof of Affiliation for Undergraduate Students)

त्रिभुवन विश्वविद्यालय
Tribhuvan University
इन्जिनियरिङ्ग अध्ययन संस्थान
Institute of Engineering
थापाथली क्याम्पस
THAPATHALI CAMPUS

GPO Box-280, Thapathali, Kathmandu
Tel: 4-246465, 4218300, Fax: 977-1-4247340
E-mail: info@tcioe.edu.np
Website: www.tcioe.edu.np
गोश्वारा पो.ब.नं. २८०, थापाथली, काठमाडौं
फोन-४२४६४६५, ४२१८३००
फ्याक्स: ९७७-१-४२४७३४०

Date:- 28/11/2023

To Whom It May Concern

This is to confirm that Mr. Abhinav Chalise, Mr. Nimesh Gopal Pradhan, Mr. Nishan Khanal, and Mr. Prashant Raj Bista are currently affiliated with the Institute of Engineering, Thapathali Campus in Kathmandu, Nepal.

Mr. Abhinav Chalise, Mr. Nimesh Gopal Pradhan, Mr. Nishan Khanal, and Mr. Prashant Raj Bista are affiliated as Undergraduate Students in the Department of Electronics and Computer Engineering.

We fully support Mr. Abhinav Chalise, Mr. Nimesh Gopal Pradhan, Mr. Nishan Khanal, and Mr. Prashant Raj Bista in their application for the present research grant call of the National College of Engineering, Tribhuvan University, Nepal.

Kind regards,

Asst. Prof. Dr. Khem Gyanwali
Campus Chief

Appendix - [5]
(Equivalency of certificate belonging to Principal Investigator)

त्रिभुवन विश्वविद्यालय
Tribhuvan University
इन्जिनियरिङ अध्ययन संस्थान
Institute of Engineering
डीनको कार्यालय
OFFICE OF THE DEAN
GPO box- 1915, Pulchowk, Lalitpur
Tel: 977-5-521531, Fax: 977-5-525830
dean@ioe.edu.np, www.ioe.edu.np
गोश्वारा पो.ब. नं- १९१५, पुल्चोक, ललितपुर
फोन- ५५२१५३१, फ्याक्स- ५५२५८३०

पत्र संख्या: डी.का.यो.फा.नं.(　　)च.नं.। १०४१ ।०७२/०७३　　मिति :२०७२।१०।६

जो जसलाई सम्वन्ध छ ।

उपरोक्त बारे Stanford University, USA बाट M.Sc. Electrical Engineering मा उपाधि प्राप्त गर्ने **दिनेश वानिया क्षेत्री**को उक्त उपाधि इ.अ.सं. स्तर निर्धारण राय सुझाव समितिको मिति २०७२।१०।६ गते बसेको बैठकको निर्णयानुसार M.Sc., Electronics and Communication Engineering संग सम्वन्धित भएको व्यहोरा प्रमाणित गरिन्छ ।

२०७२।१०।२०

(नगेन्द्रबहादुर अमात्य)
स.डीन (शैक्षिक प्रशासन)

त्रिभुवन विश्वविद्यालय
Tribhuvan University
इन्जिनियरिङ अध्ययन संस्थान
Institute of Engineering
डीनको कार्यालय
OFFICE OF THE DEAN
GPO box- 1915, Pulchowk, Lalitpur
Tel: 977-5-521531, Fax: 977-5-525830
dean@ioe.edu.np, www.ioe.edu.np
गोश्वारा पो.ब. नं- १९१५, पुल्चोक, ललितपुर
फोन- ५५२१५३१, फ्याक्स- ५५२५८३०

पत्र संख्या: डी.का.यो.फा.नं.(　　)च.नं.। ८०४　।०७३/०७४　　मिति :२०७३।१०।१४

जो जसलाई सम्वन्ध छ ।

Stanford University, USA बाट श्री दिनेश वानिया क्षेत्रीले प्राप्त गर्नु भएको M.Sc. in Electrical Engineering उपाधि M.Sc. in Information and Communication Engineering संग सम्वन्धित भएको प्रमाणित गराई पाउँ भनी दिनु भएको निवेदन सम्वन्धमा इ.अ.सं. स्तर निर्धारण तथा राय सुझाव समितिको मिति २०७३।१०।१४ गते वसेको बैठकले निजको उक्त उपाधि Information and Communication Engineering संग सम्वन्धित भएको व्यहोरा प्रमाणित गरिन्छ ।

(नगेन्द्रबहादुर अमात्य)
स.डीन (शैक्षिक प्रशासन)

References

- A. Hertz, R. Mokady, J. Tenenbaum, K. Aberman, Y. Pritch, and D. Cohen-Or, “Prompt-to-Prompt Image Editing with Cross Attention Control,” arXiv.org, Aug. 02, 2022. https://arxiv.org/abs/2208.01626 (accessed Nov. 19, 2023).
- W. Dong, S. Xue, X. Duan, and S. Han, “Prompt Tuning Inversion for Text-Driven Image Editing Using Diffusion Models,” arXiv.org, May 08, 2023. https://arxiv.org/abs/2305.04441 (accessed Nov. 19, 2023).
- R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-Resolution image synthesis with latent diffusion models,” in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2022. Accessed: Nov. 22, 2023. [Online]. Available: http://dx.doi.org/10.1109/cvpr52688.2022.01042
- R. Mokady, A. Hertz, K. Aberman, Y. Pritch, and D. Cohen-Or, “Null-text Inversion for Editing Real Images using Guided Diffusion Models,” arXiv.org, Nov. 17, 2022. https://arxiv.org/abs/2211.09794 (accessed Nov. 19, 2023).
- D. Miyake, A. Iohara, Y. Saito, and T. Tanaka, “Negative-prompt Inversion: Fast Image Inversion for Editing with Text-guided Diffusion Models,” arXiv.org, May 26, 2023. https://arxiv.org/abs/2305.16807 (accessed Nov. 19, 2023).
- K. J. Joseph et al., “Iterative Multi-granular Image Editing using Diffusion Models,” arXiv.org, Sep. 01, 2023. https://arxiv.org/abs/2309.00613 (accessed Nov. 19, 2023).