Tuning of Diffusion Models for Enhancing Synthetic Data Usage Ratio

Novo Nordisk PBL – Team 1 (Track 3)

21 May, 2025

Topics

01 Proposed Timeline

02 Overview

03 Disease Type & Dataset

04 Architecture

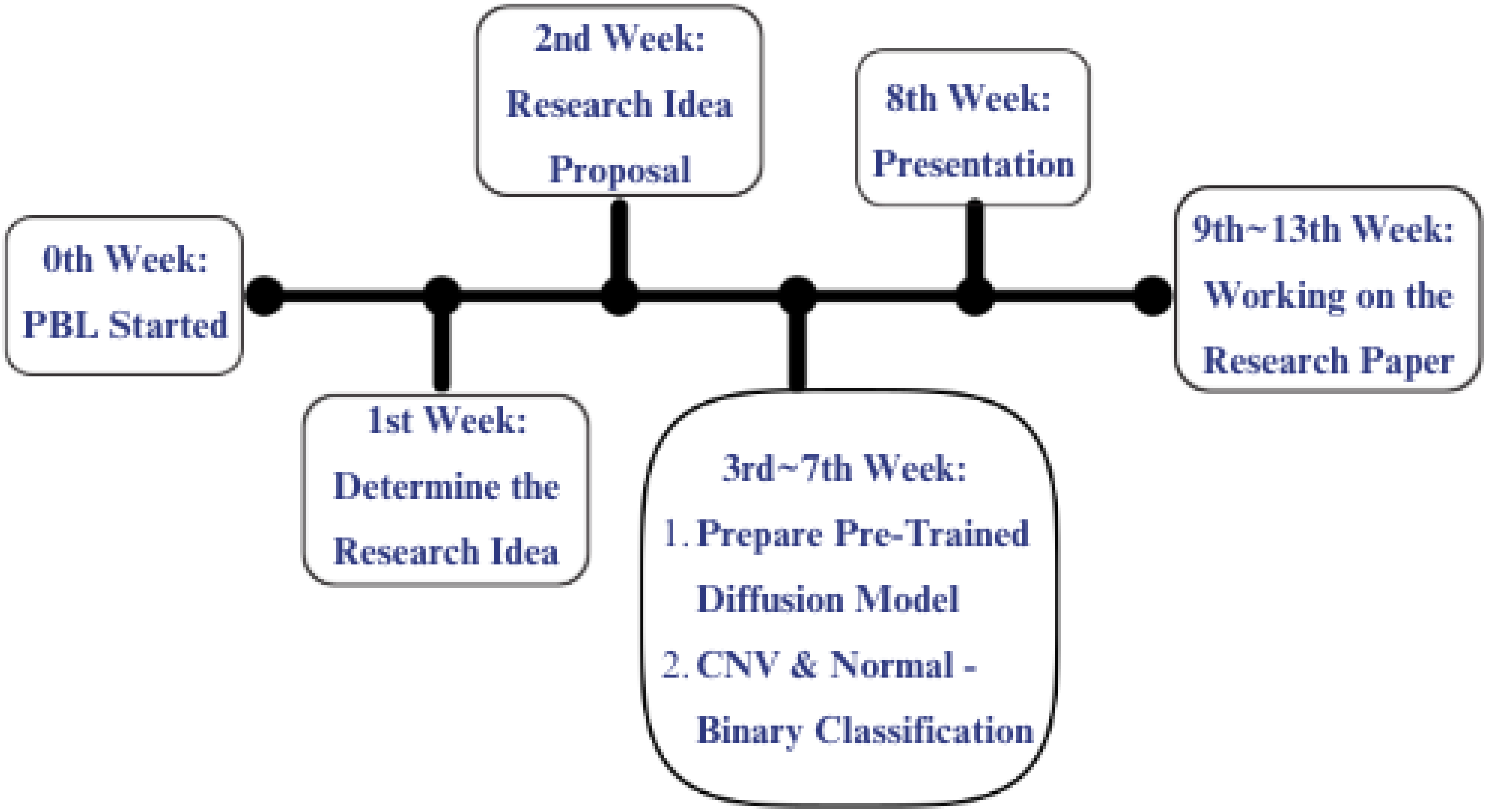
05 Synthetic CNV Images

06 Results

07 Paper Agenda

08 Team 1 – Participants

Proposed Timeline



Overview

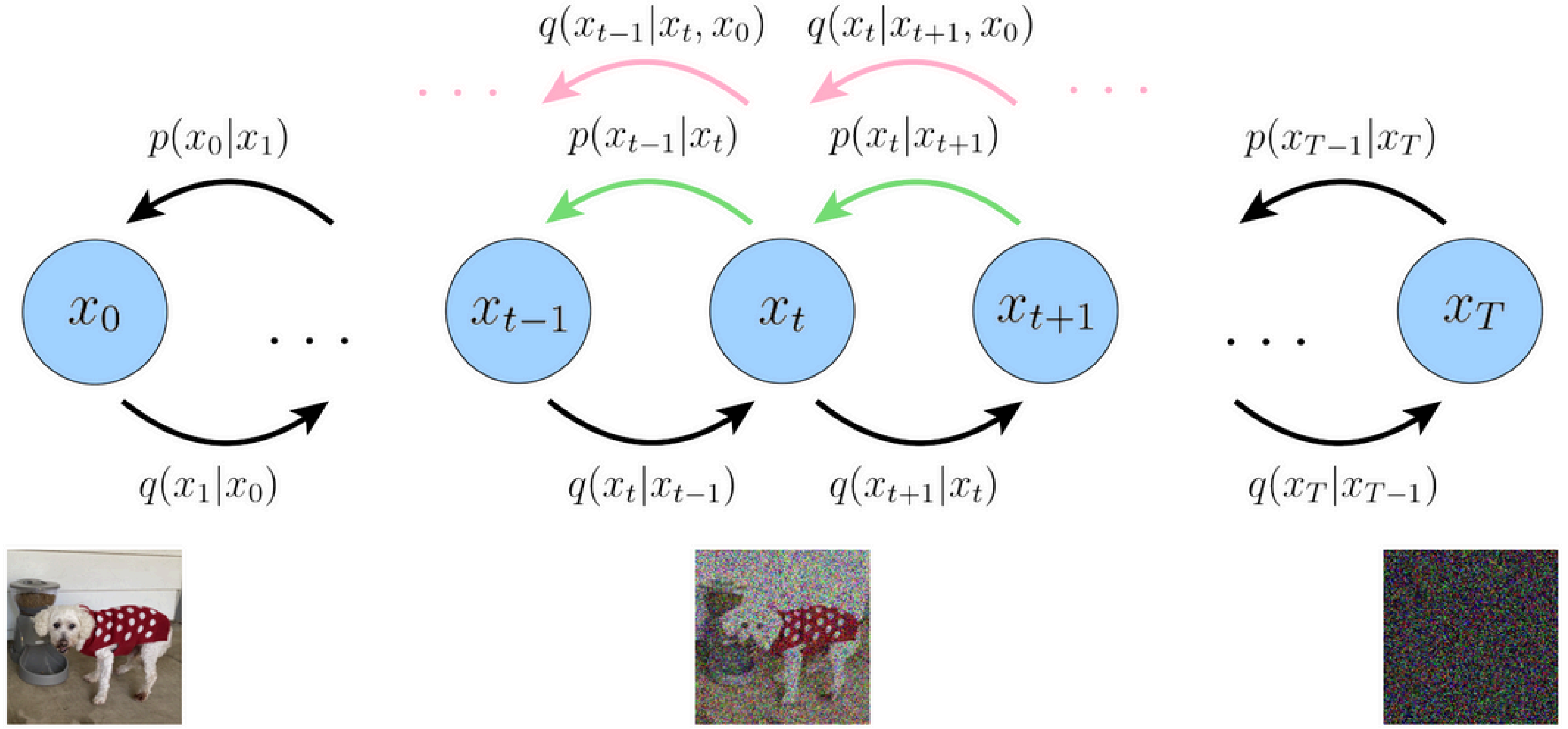
Research Question

How generative models are useful in medical imaging field?

Research Aim

Determine the ratio R *which decides how much the data is augmented using synthetic images*

Synthetic Image Generation with Diffusion Model



(Luo, 2022)

Synthetic Images

Mix Synthetic Images at ratio R

Real Images (Training Set)

Classifier

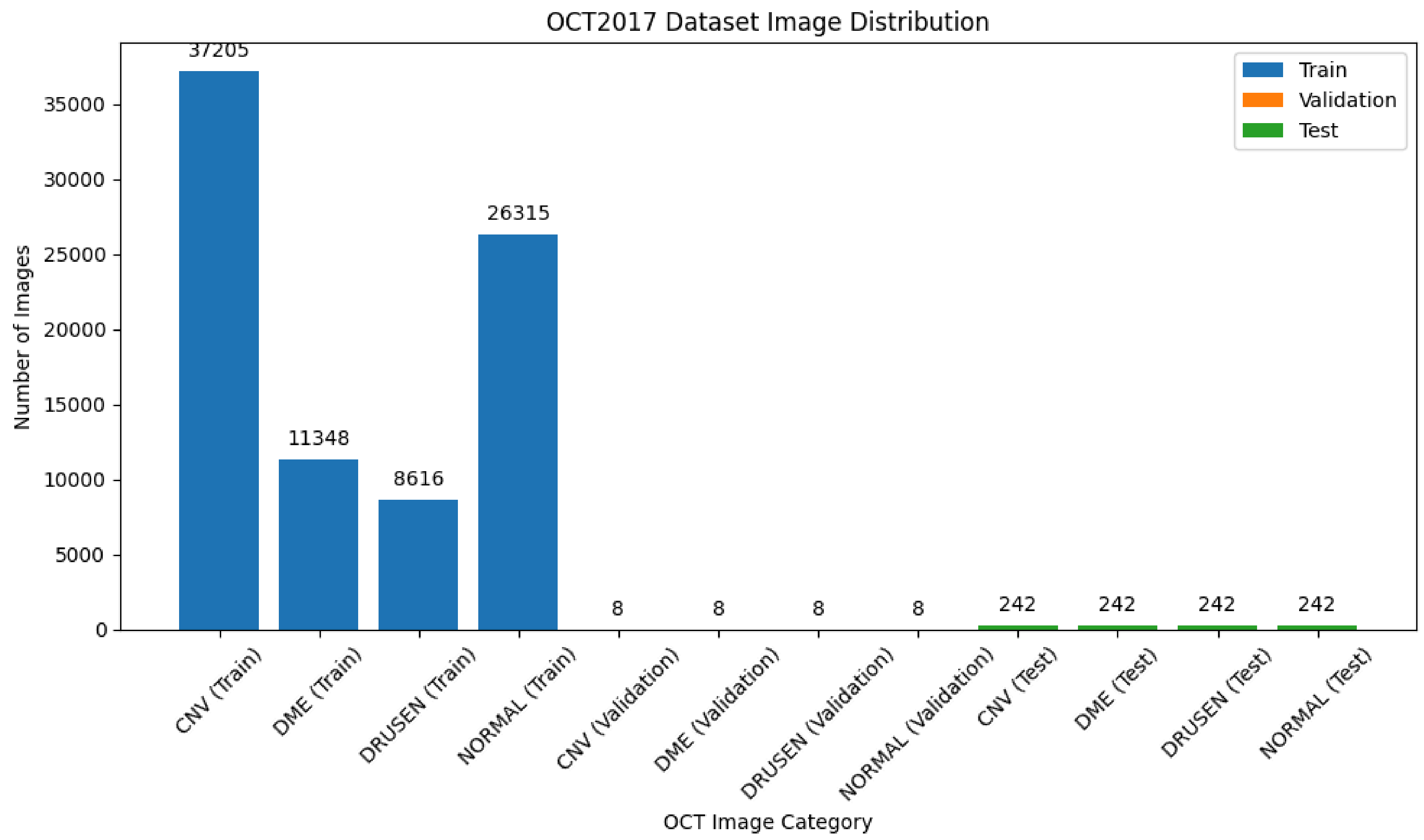
Real Images (Validation Set)

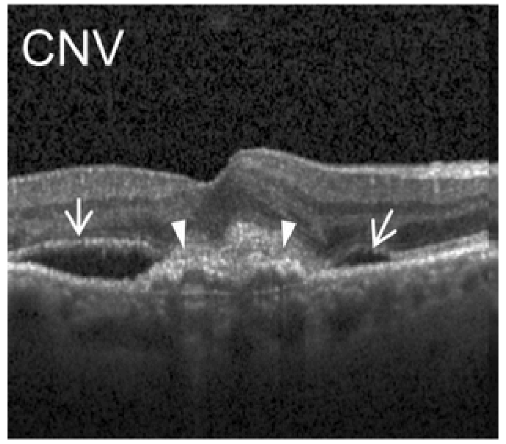
Real Images (Test Set)

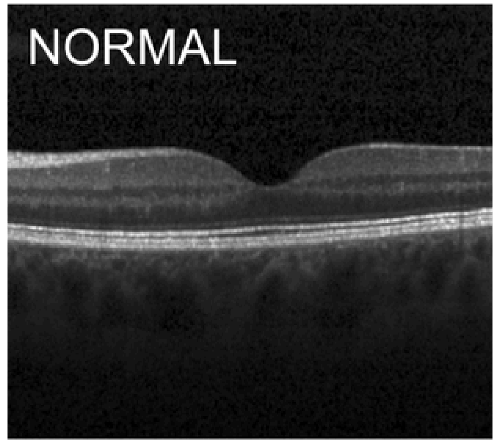
Result

- Accuracy
- F1 Score
- Precision
- Recall

Disease Type & Dataset







Definition: Abnormal new blood vessels growing from the choroid into or under the retina.

Diagnosis: OCT reveals neovascularization and subretinal fluid.

Data distribution

- 83484 images under train set:

CNV: 37205 NORMAL:26315

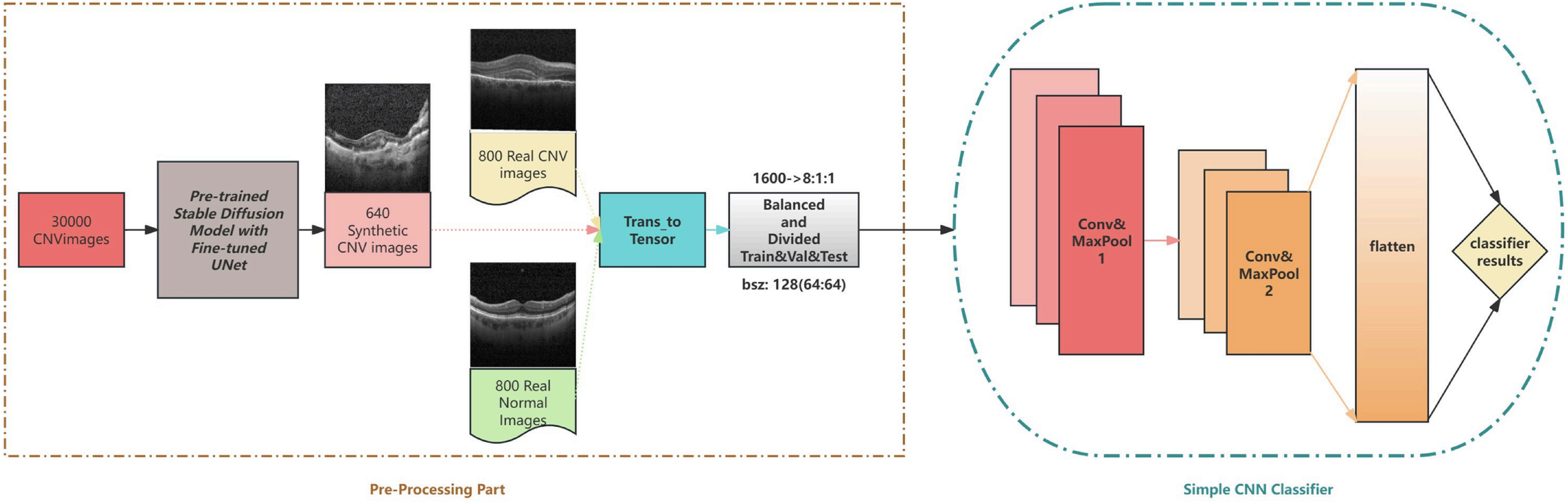
- 32 images under val set:

CNV: 8 NORMAL:8

- 968 images under test set:

CNV: 242 NORMAL:242

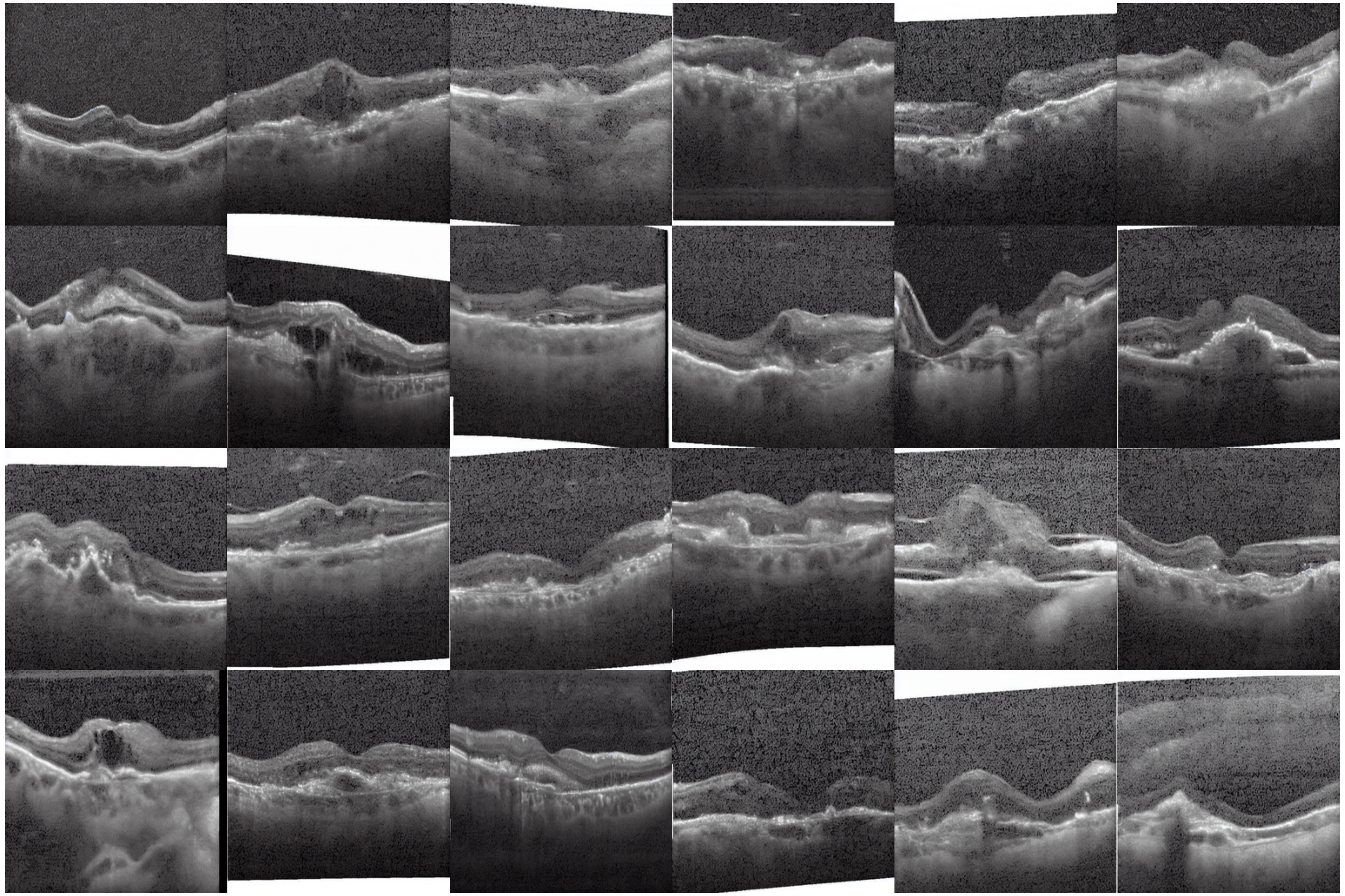
Architecture



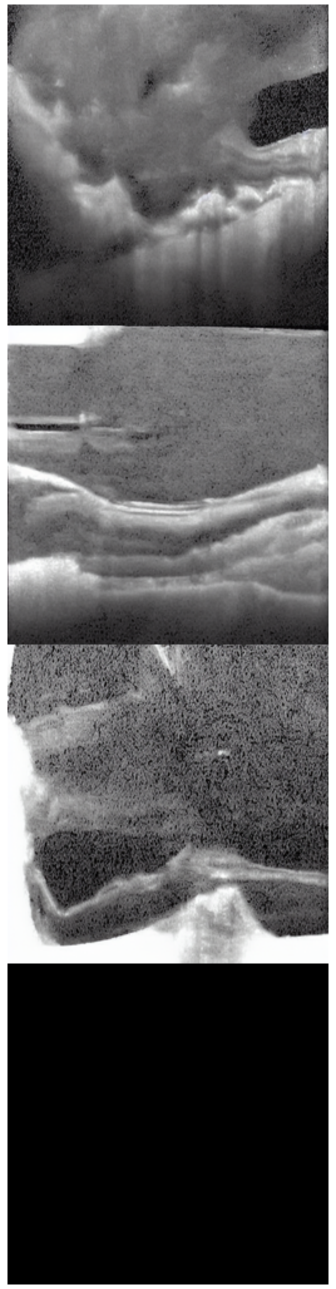
CNV are mixed with ratio R in each training batch

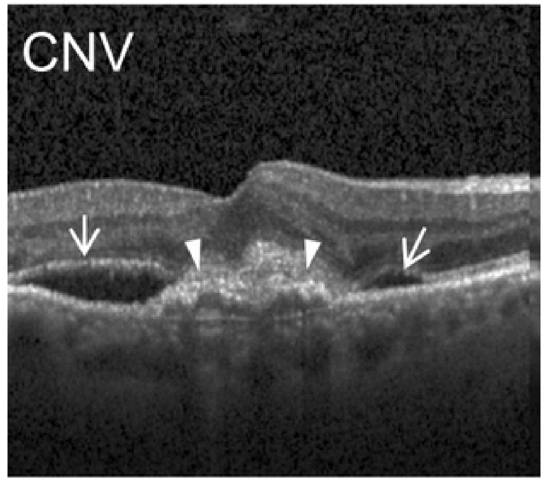
Synthetic CNV Images

Well-Generated



Poorly-Generated





You can see white region in the images, which represents **high-reflectance areas** in the OCT images, as CNV typically appears as bright (hyper-reflective) areas in such scans due to abnormal blood vessel growth, fluid accumulation, or leakage beneath the retina.

CNV, Normal (Binary Classification)

Training Diffusion Model:

- Preprocessed the images (512 resolution, normalised values)
- Used the term of “OCT scan showing CNV” for finetuning and image generation
    - Other methods relating to prompt that are tried but results in poorly generated images:
        - diversify/”randomise” the prompts used for each image
        - textual inversion: <cnv> token to capture the information of the CNV images effectively
- Main idea: Using a freezed pre-trained VAE, tokeniser, we train the UNet to learn how to denoise the images to generate the synthetic images
- Trained on 30000 images, for 10-20 epochs
    - Insight:
        - Number of training images used affects quality of synthetic images more significantly than number of epochs trained
        - Using larger pre-trained model rather than the distilled version with more denoising steps produces better results
    - Methods Tried/ Results:
        - Slower learning rates - better results
        - Decaying learning rates - similar results as above
        - Higher inference steps - effects plateaus/significantly worsens after 300 steps
        - Higher guidance scale - effects plateaus/significantly worsens after 12.5

CNV & Normal Binary Classification Results

Varied Accuracy, F1 Score, Precision, Recall Results by Data Distribution Ratio

```
Results for Ratio 10% to 90%:
Ratio: 10% | Accuracy: 0.8625 | F1 Score: 0.8624 | Precision: 0.8634 | Recall: 0.8625
Ratio: 20% | Accuracy: 0.8562 | F1 Score: 0.8562 | Precision: 0.8563 | Recall: 0.8562
Ratio: 30% | Accuracy: 0.8750 | F1 Score: 0.8750 | Precision: 0.8750 | Recall: 0.8750
Ratio: 40% | Accuracy: 0.8750 | F1 Score: 0.8750 | Precision: 0.8750 | Recall: 0.8750
Ratio: 50% | Accuracy: 0.9000 | F1 Score: 0.8999 | Precision: 0.9010 | Recall: 0.9000
Ratio: 60% | Accuracy: 0.8438 | F1 Score: 0.8436 | Precision: 0.8451 | Recall: 0.8438
Ratio: 70% | Accuracy: 0.8187 | F1 Score: 0.8171 | Precision: 0.8304 | Recall: 0.8187
Ratio: 80% | Accuracy: 0.7750 | F1 Score: 0.7698 | Precision: 0.8022 | Recall: 0.7750
Ratio: 90% | Accuracy: 0.6813 | F1 Score: 0.6632 | Precision: 0.7306 | Recall: 0.6813
Results saved to classifier_results.csv
```

<u>What happens when R=100%?</u>

Accuracy drops significantly as model pick up on the quirks and "fingerprints" of the diffusion generator rather than the true anatomical features present in real OCT scans

<u>Best Results at **Max Ratio R: 50%** (i.e. domain balance)</u>:

- At 50% real / 50% synthetic, our network sees both distributions enough to learn robust, domain-invariant features.
  - Empirical Precedence: In histopathology, Manrai et al. (Nature Med 2024) found optimal classifier performance at a 1:1 real-synthetic ratio; pushing more synthetic data actually hurt generalization.
  - However, the paper does not cover OCT scans (only Histopathology, Chest Radiology, Dermatology where each field shows different optimal ratios)
- Beyond ~60% synthetic, the synthetic features (e.g. overly smooth textures, slight contrast mismatches, repeated micro-patterns) start to dominate the gradient signal, and test-time accuracy on real CNV images collapses.

Paper Agenda #1

How can we modify our Diffusion Model to increase the image quality and ratio R?

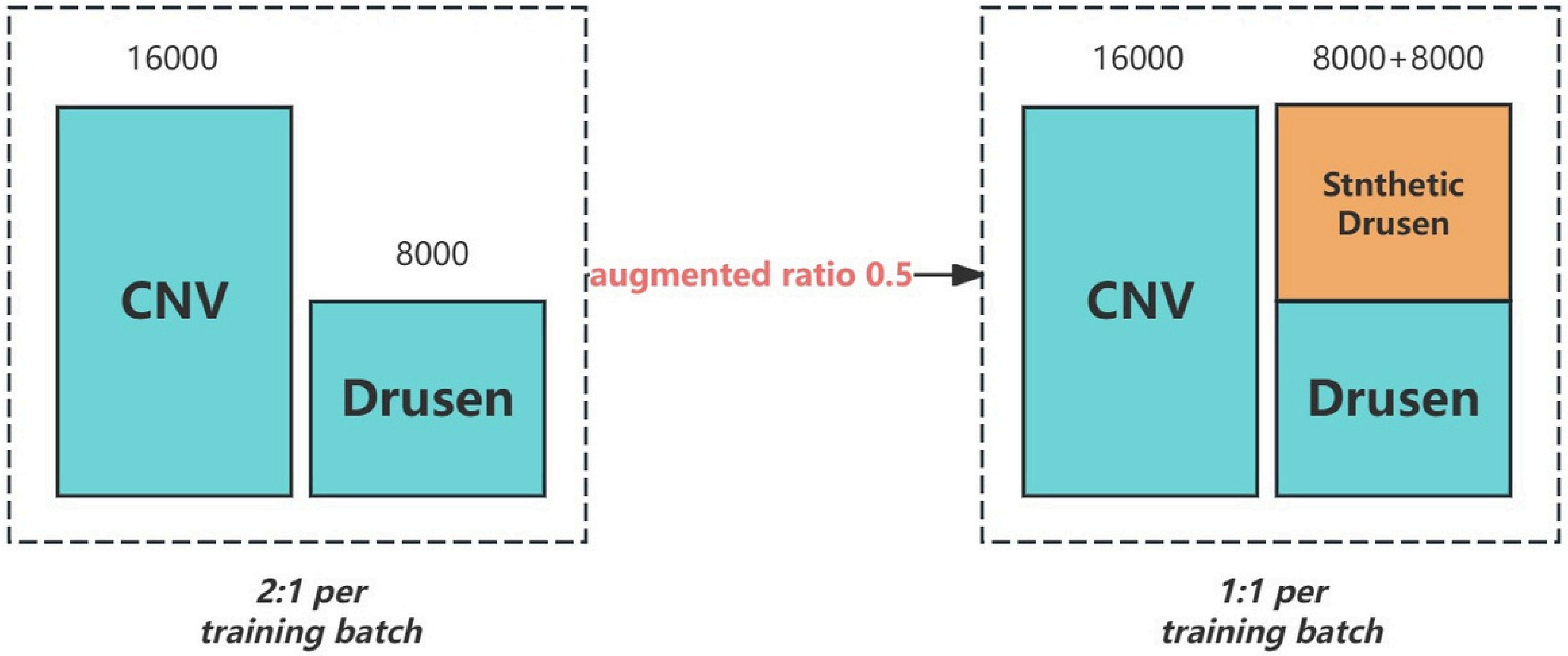
- Must shrink the gap between the real domain and the synthetic domain
- Possible ways:
  - Improve sample fidelity through noise-schedule tuning
    - Adaptive noise schedules: Use a WSNR-equivalent or importance-sampled schedule (e.g. from "Improved Noise Schedule for Diffusion Training" or "Rethinking the Noise Schedule").
    - Cosine or dynamic schedules: Replace the default linear β-schedule with a cosine schedule (Nichol & Dhariwal 2021) or one tailored via a noise-control survey (e.g. Kong et al. "A Comprehensive Review on Noise Control"). Empirically select the schedule that minimises a real-vs-synthetic FID on a held-out validation set.
  - Mimic real-world noise and imaging artifacts
    - Learn the true noise distribution of your OCT device by measuring the noise spectrum in background patches; then add that as a final "noise-injection" step to each synthetic image.
    - Noise-to-noise training (Lehtinen et al.) can recover realistic sensor noise patterns that diffusion alone may smooth out.

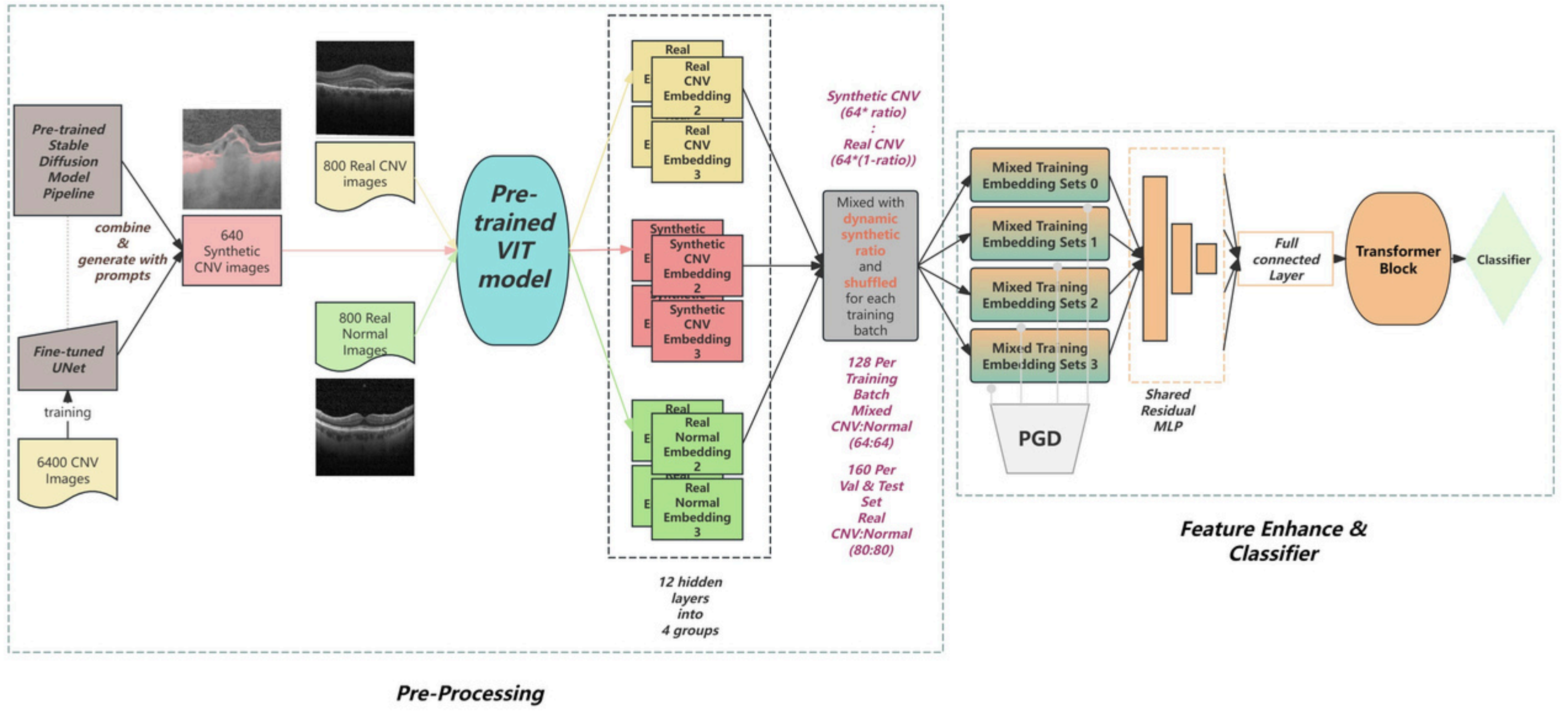
Paper Agenda #2

- Whether should we test the ratio on a harder model?

  Yes! We should go beyond just using simple classifier as a way of validating the quality of synthetic data

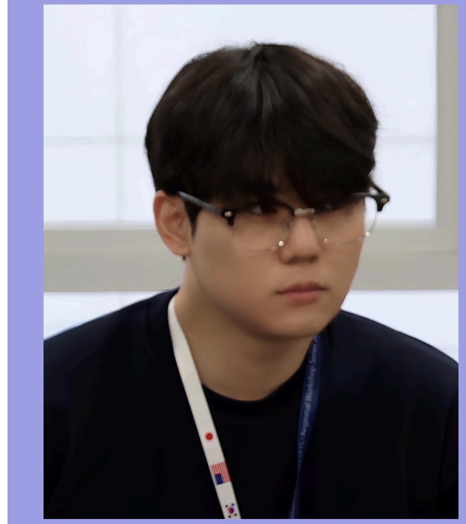
- Whether should we aim to address real-world issue through our research and experiments in the paper? Yes!





The specific experimental direction requires further reflection through practice and reading.

Team 1 - Participants



John Yechan Jo (Sole Leader)

- Worked on combining the initial Fine-Tuned Diffusion Model & Data Distribution & Classifier Codes
- Contributed as a Sole Leader for Team 1 and arranged team meetings

Yiran Qi

- Worked on datasets organizing and pre-processing
- Worked on designing, building and modifying Classifiers

Kazuma Itabashi

- Worked on modifying Lightweight Diffusion Model and Fine-tuning
- Worked on Building a Classifier for Experiment

Joel Leo

- Worked on, fine-tuned and trained initial, lightweight, final diffusion model used in the project.

Thank You

21 May, 2025