Pneumonia Symptom Classification with Diffusion Models and CLIP

Group 2: Xiaomeng Xu; Wenfei Mao; Yingzhen Wang; Shuoyuan Gao

Introduction

Background

Pneumonia: A Global Health Challenge

- A severe lung infection causing alveolar inflammation.
- **Leading cause of mortality** in children under 5 years old.
- **Deaths per year:** Higher than HIV, malaria, or tuberculosis.
- Early and accurate diagnosis is critical to reducing mortality.

Traditional Diagnosis:

- **Symptoms-based:** Fever, cough, difficulty breathing.
- **Tools:** Physical exams and chest X-rays.
- **Challenges:** Relies on radiologist expertise; limited access in low-resource settings.

Data Source

Dataset Overview

- **Source:** Kaggle Chest X-ray dataset.
- **Images:** 5,856 verified chest X-rays.
- **Division:** Training and testing sets.
- **Origin:** Guangzhou Women and Children’s Medical Center.

Advantages of Dataset:

- Focus on pediatric cases.
- High-quality, verified images for robust model training.

Disadvantages of Dataset

- Uneven distribution across pneumonia categories: **Cancerous pneumonia** may have more images compared to normal or viral pneumonia.

Motivation and Objective

Research Motivation

1. Data Imbalance Challenges

- Uneven distribution of pneumonia categories (e.g., bacteria, normal, viral).
- Rare categories lack sufficient data, leading to poor model performance.

2. Complexity in Image Classification

- Chest X-rays exhibit overlapping visual features across pneumonia types.
- Traditional methods fail to utilize additional contextual information like text descriptions.

3. Demand for Automated Diagnosis

- Limited medical resources, especially in low-resource settings.
- Need for efficient and accurate pneumonia classification systems to support clinicians.

Research Objective:

Build an efficient and accurate model system to achieve: Generate high-quality medical images, and accurately classify pneumonia cases

Methods

Overview

- Pneumonia data is **unbalanced** across three categories. There are 2538 images of bacterial pneumonia, 1349 of normal pneumonia, and 1345 of viral pneumonia.
- Deploy **Stable diffusion Version 2** locally and fine-tuned the Stable Diffusion model with **LoRA**.
- Use fine-tuned Stable diffusion model to generate **1000** images each for viral and normal pneumonia images.
- The **CLIP** was fine-tuned with LoRA utilizing **mixed-precision training**, and the fine-tuned model was subsequently employed to classify the three categories.

LoRA

- Low-Rank Adaptation of Large Language Models, was proposed in 2021.
- Freeze the pre-trained model weights and only fine-tuned LoRA A and LoRA B.
- LoRA can reduce the number of trainable parameters by 10,000 times.
- Only fine-tuned attention layer and projection layer of **U-Net**.

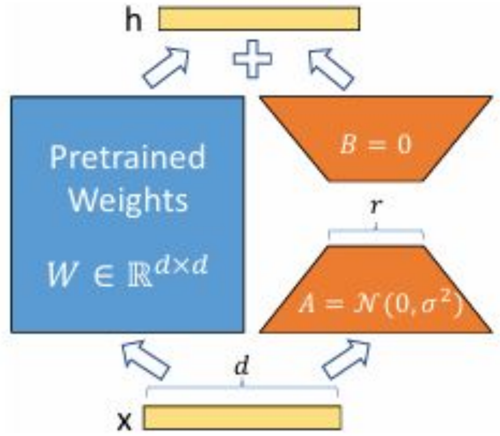
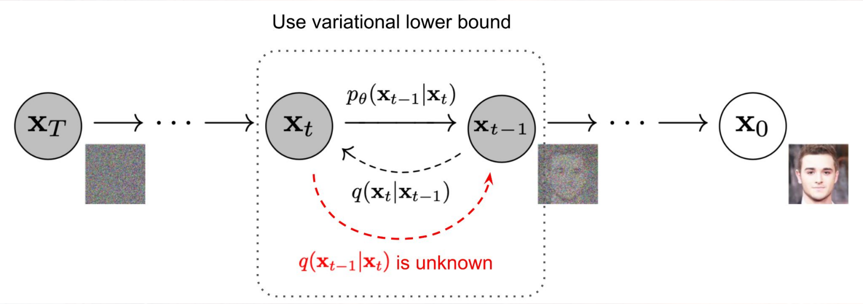


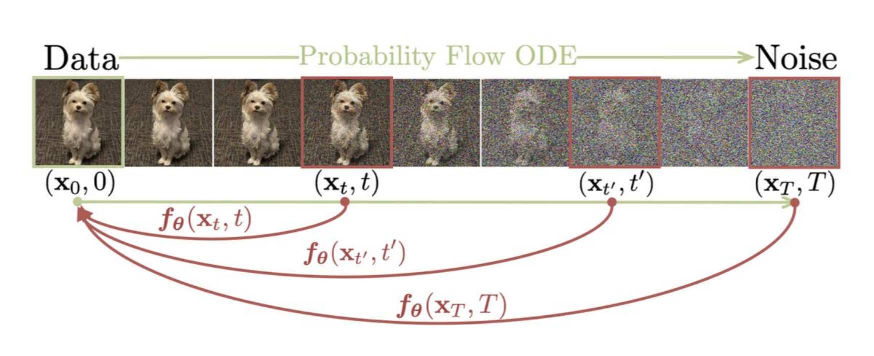
Figure 1: Our reparametrization. We only train A and B.

```
trainable params: 5,406,720 || all params: 433,023,233 || trainable%: 1.2486
```

Diffusion Model

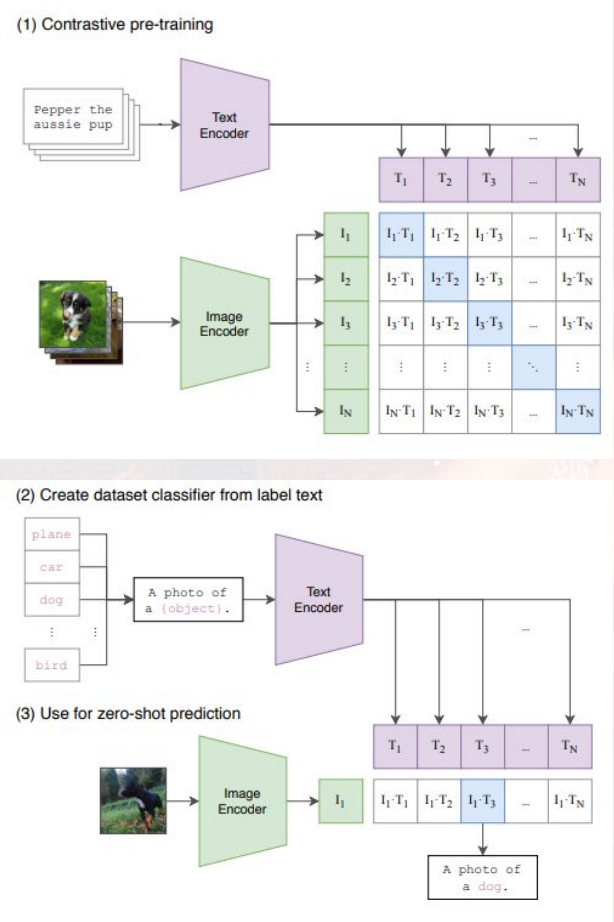
- Diffusion models became more and more popular since 2020.
- Use **865M U-Net** as image generator and use **OpenCLIP ViT-H/14** as image-text encoder. It could generate 768×768px outputs.
- **Consistency models**(ICML 2023)is the most promising method to generate images.





CLIP

- Contrastive Language-Image Pre-training was proposed by OpenAI in 2021.
- CLIP use ResNet or ViT as image encoder and use CBOW or Text Transformer as text encoder.
- Deploy **CLIP-ViT-large-patch14**. The model use **ViT-L/14 Transformer** as image encoder and use **masked self-attention Transformer** as text encoder



Experiments

Experiment Setup

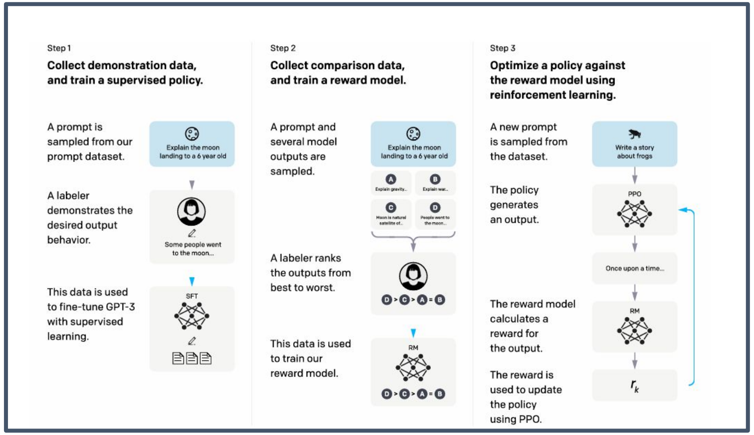
- NVIDIA RTX 3090 24GB Memory
- CUDA 12.2 Toolkit
- PyTorch Version 2.5.1
- Use Flash Attention = False
- Optimizer: AdamW; Batch size = 4; Learning rate = 1e-4
- Number of epochs for fine-tune: 1 epoch
- LoRA configuration: LoRA alpha = 16; LoRA dropout = 0.1

UNIVERSITY OF
MICHIGAN

Discussion

Limitations & Future Work

- Stable Diffusion uses **DDPM** for training, while **consistency models** enable faster image generation with significantly reduced time and computational resources
- Adopted **reinforcement learning** methods (e.g., RLHF) to train a reward model, which improved the model's performance
- Use **Dreambooth** to generate more customized images instead of using LoRA



Thank You for Listening

Please feel free to ask us if you have any questions