Air Pollution Estimation from Sky Photos (UPenn Project)

Overview
This project estimates ground-level sulfur dioxide (SO2) and nitrogen dioxide (NO2) from ordinary sky photos. The emphasis is on a dependable end-to-end workflow: cleaning noisy images, training under a consistent setup, and evaluating results fairly.

What Is In This Repository
- UPenn Research Paper.pdf
  Full write-up of the problem, approach, experiments, and results. The dataset source is listed in the citations of the paper.
- Air_Pollution_Models.ipynb
  Notebook used for preprocessing, training, and evaluation.

Dataset
I used the Air Pollution Image Dataset referenced in the citations of the research paper. The dataset is not included in this repository. I have included the link to the dataset here:
https://www.kaggle.com/datasets/adarshrouniyar/air-pollution-image-dataset-from-india-and-nepal

Results
Using a consistent split and training setup, the best models reached:
- SO2: 99.25 percent test accuracy
- NO2: 94.72 percent test accuracy

How To Review Quickly
1) Read the research paper for context and conclusions
2) Skim the notebook for preprocessing, model comparison, and evaluation

How To Run
Open Air_Pollution_Models.ipynb and update the dataset path in the first data loading cell to match your local folder. Run the notebook top to bottom to reproduce preprocessing, training, and evaluation.

Notes
This repository is a portfolio snapshot of the work. It does not include the dataset or trained weights. Accuracy numbers depend on the same split and preprocessing described in the paper and notebook.