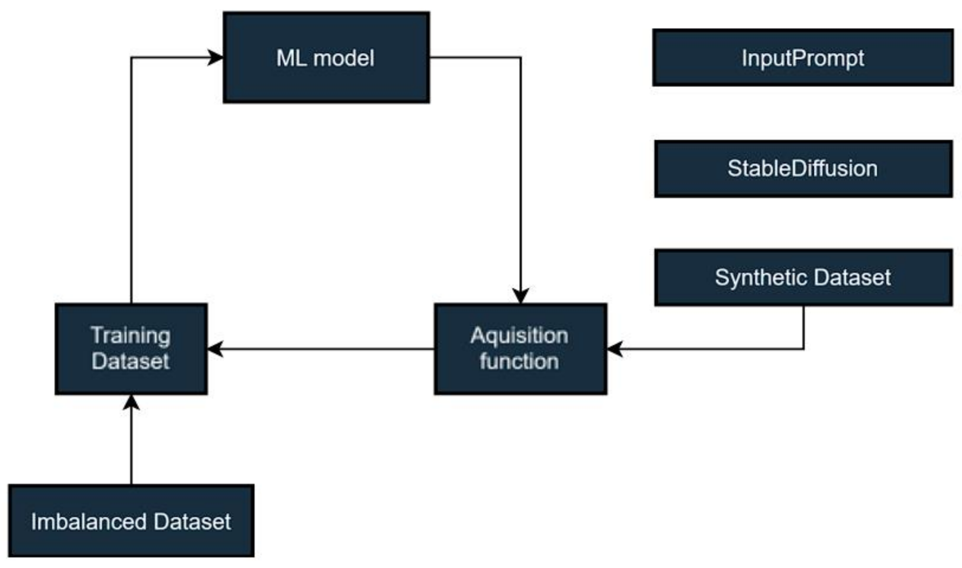
Data shortage? Let Diffusion fill the gaps!

ImbalanceSD

Project overview

This project aims to investigate possible improvements that synthetic data generation can bring to the unbalanced classification problem. For instance, I'll be downsampling CIFAR10 classes to 1%, and replacing remaining 99% with generated images



CIFAR10 Image generation

- **CIFAR10** images have **32x32 resolution**, which is problematic for large pre-trained diffusion models trained to produce 512x512 images
- Different diffusion models were used to generate data:, while only 2 generated approximately good results:
  - **StableDiffusion 3.5 Large Turbo**
  - **StableDiffusion XL**
- All the **models struggle with lower resolutions**, so I had to generate samples in 512x512 and then perform bicubic down sampling to 32x32
- Different prompts were used, often the models didn't follow the prompt. **The simpler the better!**
- Simple LoRA/Dreambooth don't help
- For final prompt (below) I had 1k different combinations of modifiers

<QUALITY_MODIFIER> photo of a <CLASS_DESCRIPTOR> <CLASS_NAME> <EXTRA_CONTEXT>"

Example images

Cat Airplane Truck

CIFAR10

SDXL

Experiments

All results are obtained using non-pretrained ResNet18

Experiment	Test accuracy	Cat test accuracy
Full CIFAR10	0.868	0.748
Cat downsampled to 1%	0.764	0.036
Cat downsampled to 1% + ADASYN	0.803	0.01
Cat downsampled to 1% + FLUX.1-Redux augmentation	0.815	0.073
Cat downsampled to 1% + SD3.5L-Turbo	0.817	0.094
Cat downsampled to 1% + SDXL	0.811	0.061
Cat downsampled to 1% + SDXL + similarity filter	0.815	0.075
Cat downsampled to 1% + SDXL + LoRA	0.819	0.117
All classes downsampled to 1% + Full Synthetic	0.493	**0.429**
Full CIFAR10 + Full Synthetic	**0.871**	0.741

Conclusions

- When introducing synthetic data to only one imbalanced class the model overfits to synthetic data type
- Future work may include augmentation techniques to mitigate domain gap between synthetic and real data

Karol Cyganik
Under the supervision of Dr Wojciech Pałubicki

RESEARCH UNIVERSITY
EXCELLENCE INITIATIVE

UAM
Uniwersytet im. Adama Mickiewicza w Poznaniu