Vision-Language Models are Strong Noisy Label Detectors (Code reproduction results)

#1) Team Member Details

1. Name: Anvit Gupta
   Enrolment Number: 22114009
   Email ID: anvit_g@cs.iitr.ac.in
   Phone Number: +91 9462011044

2. Name: Sarvasva Gupta
   Enrolment Number: 22114086
   Email ID: sarvasva_g@cs.iitr.ac.in
   Phone Number: +91 7989904811

#2) Changes made in the code

- Comment out torch and torchvision from requirements.txt and installed using the following command:
  - pip3 install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cu118
- from clip import clip -> from .clip import clip (File: ./model/load_clip.py)
- Manually downloaded CIFAR-100.
- File: main_phase2.py, Line: 127, add the argument, weights_only=False (Reason: PyTorch2.0 used weights_only=False as default and PyTorch2.6 and beyond uses weights_only=True as default).
- Download CIFAR-100_human.pt from below and place it in data/cifar-100/
  - cifar-10-100n/data/CIFAR-100_human.pt at main · UCSC-REAL/cifar-10-100n
- dataloader/dataloader_cifar.py (Line: 70), add the argument weights_only=False
- Main_real_phase2.py, Line: 120, add the argument, weights_only=False

#3) Phase 1 results (Synthetic Dataset: CIFAR-100)

Method	Sym. 0.2		Sym. 0.4		Sym. 0.6		Ins. 0.2		Ins. 0.3		Ins. 0.4	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
DeFT (ours)	99.51	97.77	98.75	97.91	97.04	97.27	98.47	97.88	96.32	97.63	94.08	95.28
Reproduced Results	99.57	97.90	98.77	97.65	96.91	97.49	98.48	98.05	96.45	97.82	89.07	97.61

Method	Sym. 0.2		Sym. 0.4		Sym. 0.6		Ins. 0.2		Ins. 0.3		Ins. 0.4	
	Best Acc.	Last Acc.	Best Acc.	Last Acc.	Best Acc.	Last Acc.	Best Acc.	Last Acc.	Best Acc.	Last Acc.	Best Acc.	Last Acc.
DeFT	87.50	87.50	86.69	86.69	84.00	83.88	87.58	87.58	85.68	85.64	83.73	83.71
Reproduced Results	86.63	86.63	85.94	85.94	84.85	84.85	86.32	86.28	85.22	85.22	82.59	82.26

#4) Phase 2 results (Synthetic Dataset: CIFAR-100)

Method	Sym. 0.2		Sym. 0.4		Sym. 0.6		Ins. 0.2		Ins. 0.3		Ins. 0.4	
	Best Acc.	Last Acc.	Best Acc.	Last Acc.	Best Acc.	Last Acc.	Best Acc.	Last Acc.	Best Acc.	Last Acc.	Best Acc.	Last Acc.
DeFT	89.38	89.35	88.17	88.11	85.81	85.72	89.38	89.35	88.68	88.68	85.75	85.74
Reproduced Results	89.39	89.39	88.12	88.12	85.64	85.61	89.20	89.20	88.43	88.43	85.44	85.33

#5) Results on Real dataset (CIFAR-100N)

Method	Accuracy (%)
DeFT	79.04
Reproduced Result	74.10

#6) References

Research Paper - https://openreview.net/forum?id=haUnEiXgQ7