Deep Learning (CS60010) Paper Presentation

Mitigating Dataset Imbalance via Joint Generation and Classification

authors:

Aadarsh Sahoo, Abir Das — IIT Kharagpur

Ankit Singh — IIT Madras

Rameswar Panda, Rogerio Feris — MIT-IBM Watson AI Lab

Presenters -
Amatya Sharma (17CS30042)
Dewang Modi (17CS30012)

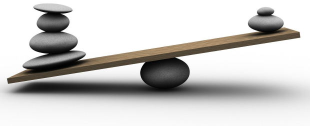
Dataset imbalance??

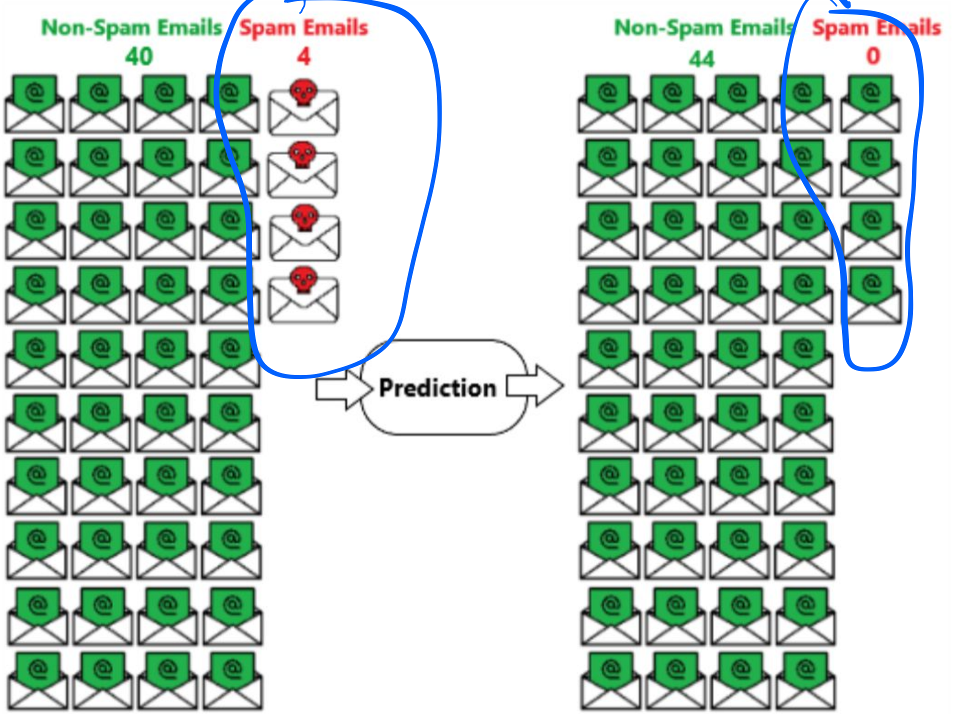
10000 100

- Majority Class >> Minority Class

- Classifier tends to fit to the majority class.

- Under-representation of Minority Class





Prior Work

Largely of two kinds

Dataset Based	Algorithmic
No change at algorithm level	Algorithmic modifications without change in dataset

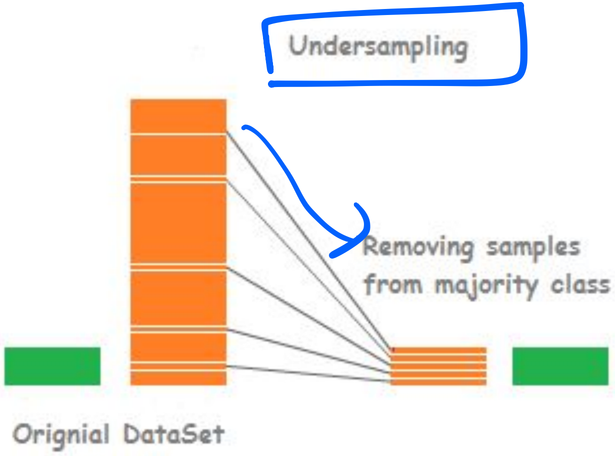
Dataset Based — No change at algorithm level

- ➢ Undersampling
- ➢ Oversampling

Algorithmic — Algorithmic modifications without change in dataset

- ➢ Cost-Sensitive weighing
- ➢ Threshold weighing
- ➢ Prior Probabilities

Dataset based methods

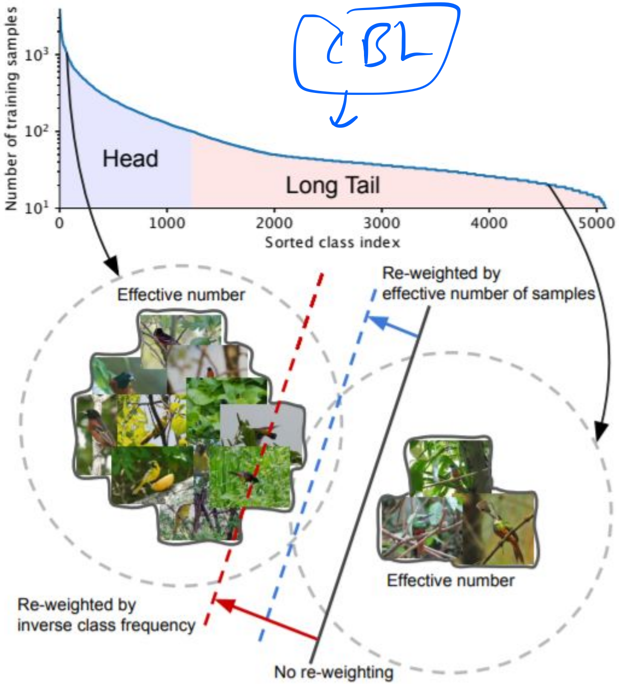


- Random [loss of features]
- Informative [eg.EasyEnsemble, BalancedCascade]

- Random [can lead to overfitting]
- Informative [eg. SMOTE, Borderline SMOTE]
- Class - aware

Algorithmic methods

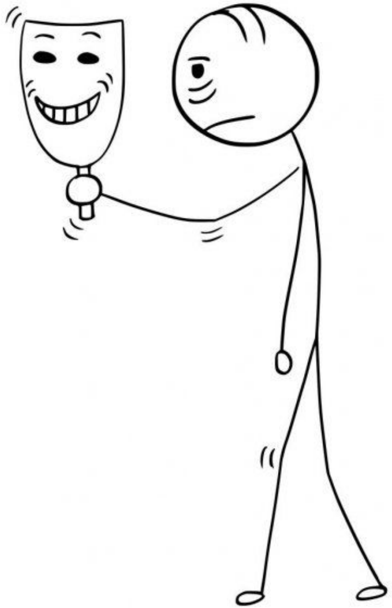
- Cost-Sensitive weighing (CS) $L' + \frac{1}{\gamma} L^{L}$
- Threshold weighing (TS) $0.5 \uparrow \downarrow$
- Prior Probabilities $P(\theta)$



Limitations of current approaches

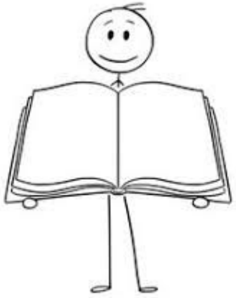
- ➢ Most methods do not generate additional data
- ➢ Existing usages only modify existing images [SMOTE]
- ➢ The dependence of context not solved

Use of GANS is not new
- ➢ Augmentation
- ➢ Combination of existing samples



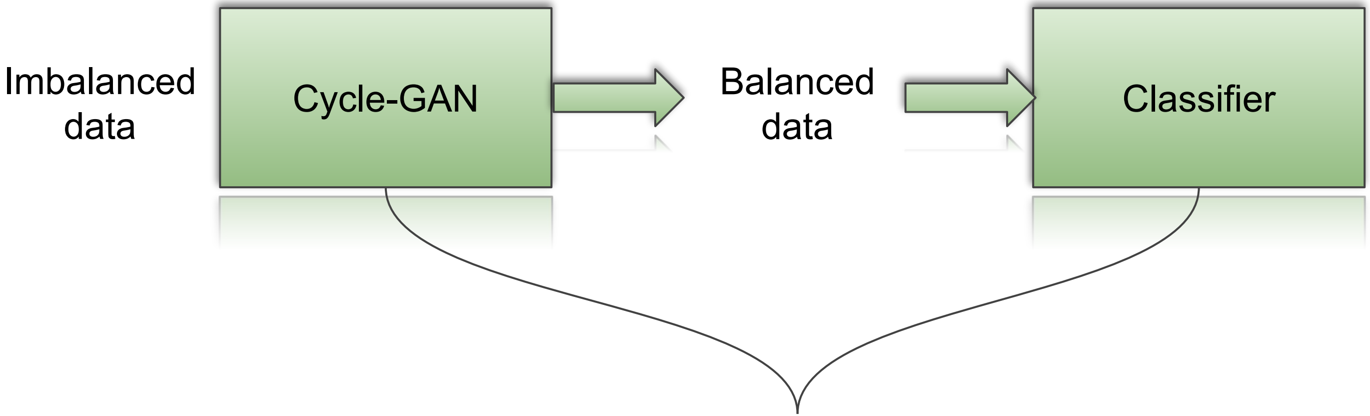
Motivation

- ➢ Prior techniques over-rely on context of data
- ➢ Collection of data of all classes not equally easy
- ➢ Traditional approaches are limited in nature
- ➢ Learning more generalizable features
- ➢ Applications in real world (medical domain) cancer!!!



The Approach...

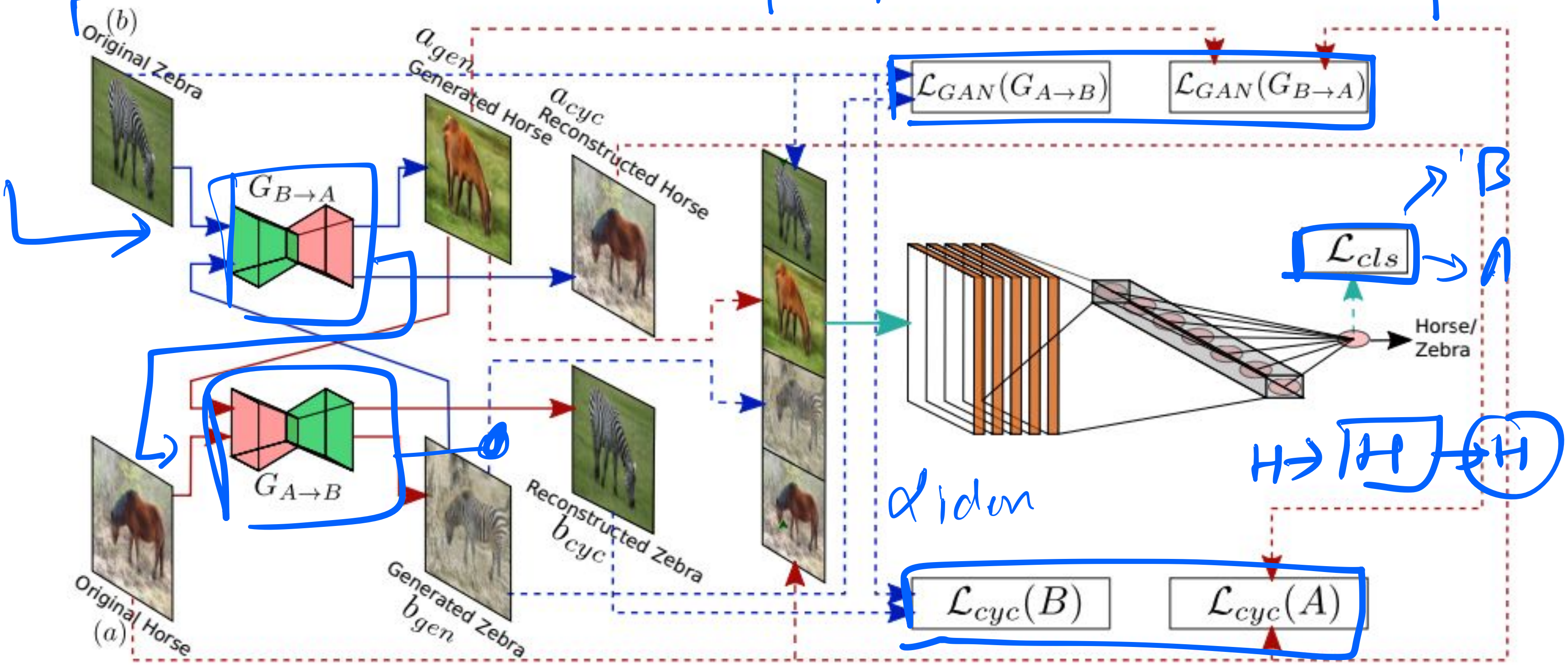
Joint cycle-GAN +classifier



Trained together, not just a part of classification pipeline

CycleGAN m → m ⇒ classifier

The Architecture



stabit

Losses

$$\mathcal{L}_{GAN}(G_{B\to A}, D_A, B, A) = -\mathbb{E}_{a\sim p_A(a)}\left[\log D_A(a)\right] - \mathbb{E}_{b\sim p_B(b)}\left[\log\left(1 - D_A(G_{B\to A}(b))\right)\right]$$

$$\mathcal{L}_{cyc}(B) = \mathbb{E}_{b\sim p_B(b)}\left[||G_{A\to B}\left(G_{B\to A}(b)\right) - b||_1\right]$$

$$\mathcal{L}_{ide}(G_{A\to B}, G_{B\to A}) = \mathbb{E}_{b\sim p_B(b)}\left[||G_{A\to B}(b) - b||_1\right] + \mathbb{E}_{a\sim p_A(a)}\left[||G_{B\to A}(a) - a||_1\right]$$

$$\mathcal{L}_{cls}^{B} = -\mathbb{E}_{b\sim p_B(b)}\left[\log z(b)\right] - \mathbb{E}_{a\sim p_A(a)}\left[\log z(G_{A\to B}(a))\right]$$

$$\mathcal{L}_{cls}^{A} = -\mathbb{E}_{a\sim p_A(a)}\left[\log(1 - z(a))\right] - \mathbb{E}_{b\sim p_B(b)}\left[\log(1 - z(G_{B\to A}(b)))\right]$$

$$\mathcal{L}_{cls} = \mathcal{L}_{cls}^{B} + \frac{1}{\gamma}\mathcal{L}_{cls}^{A}$$

$$\boxed{\begin{aligned}\mathcal{L} = \mathcal{L}_{GAN}(G_{A\to B}, D_B, A, B) + \beta\mathcal{L}_{cyc}(A) + \\ \mathcal{L}_{GAN}(G_{B\to A}, D_A, B, A) + \beta\mathcal{L}_{cyc}(B) + \\ \alpha\mathcal{L}_{ide}(G_{A\to B}, G_{B\to A}) + \mathcal{L}_{cls}^{A} + \mathcal{L}_{cls}^{B}\end{aligned}}$$

Losses

JA + CS

$\mathcal{L}_{GAN}(G_{B\to A}, D_A, B, A)$ … $D_A(G_{B\to A}(b)))]$

$\mathcal{L}_{cyc}(B) = \mathbb{E}_{b\sim p_B(b)}$ …

$\mathcal{L}_{ide}(G_{A\to B}, G_{B\to A})$ … $[||G_{B\to A}(a) - a||_1]$

γ : Cost Sensitive weighing of Loss Technique? (CS)

$\mathcal{L}^B_{cls} = -\mathbb{E}_{b\sim p_B(b)}[\log$ … $)]$ … $)]$

$\mathcal{L}^A_{cls} = -\mathbb{E}_{a\sim p_A(a)}[$ … $(1 - z(a))] - \mathbb{E}_{b\sim p_B(b)}[\log(1 - z(G_{B\to A}(b)))]$

$$\boxed{\mathcal{L}_{cls} = \mathcal{L}^B_{cls} + \frac{1}{\gamma}\mathcal{L}^A_{cls}}$$

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{GAN}(G_{A\to B}, D_B, A, B) + \beta\mathcal{L}_{cyc}(A) +\\ \mathcal{L}_{GAN}(G_{B\to A}, D_A, B, A) + \beta\mathcal{L}_{cyc}(B) +\\ \alpha\mathcal{L}_{ide}(G_{A\to B}, G_{B\to A}) + \mathcal{L}^A_{cls} + \mathcal{L}^B_{cls}\end{aligned}$$

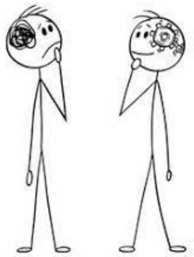
Training Methods

AUG [Augmented mode]

- **Train** cyclic-GAN **once**
- Just Train the classifier
- Only classifier loss minimized

ALT [Alternating mode]

- **Alternatively train** the classifier and cyclic-GAN
- Classifier as a “teacher”
- **BackProp L_{cls}**



$\boxed{G} \leftarrow \boxed{C}$

Training Methods

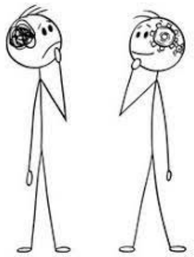
AUG [Augmented mode]

- **Train** cyclic-GAN **once**
- Just Train the classifier
- Only classifier loss minimized

Why not train end-to-end?

ALT [Alternating mode]

- **Alternatively train** the classifier and cyclic-GAN
- Classifier as a “teacher”
- **BackProp L_{cls}**



High Imbalance Data makes end2end training unstable

Evaluation Metrics

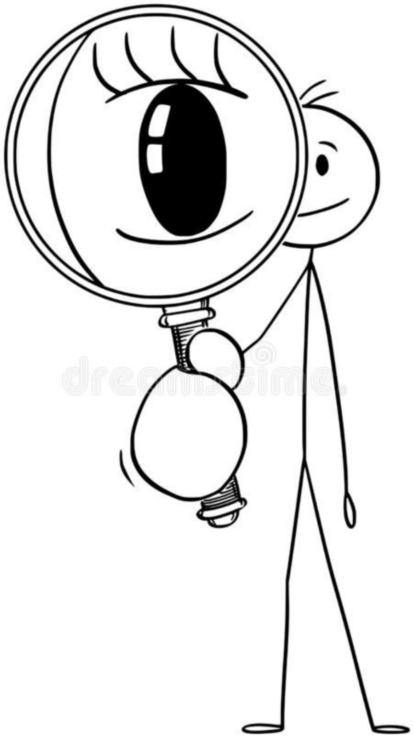
- ➢ F1 score
- ➢ ACSA (Average class Specific Accuracy)

Classifier

Score

- ➢ Inception Accuracy
  - ○ Performance of fine-tuned Inceptionv3

cyclic-GAN



OS | US | CS | OS+CS | US+CS | CBL | SMOTE | TS

Results

10000

Dataset	CelebA						CUB-200-2011					
#Minority examples	50	100	200	300	400	500	12	25	50	75	100	125
Vanilla	0.1500	0.5220	0.7740	0.8460	0.9020	0.9160	0.0240	0.1180	0.2960	0.3700	0.5520	0.6160
TS	0.1560	0.6300	0.7880	0.8420	0.8960	0.9200	0.0240	0.1180	0.2880	0.3640	0.2567	0.6080
CS	0.7825	0.8012	0.8975	0.9137	0.9250	0.9244	0.3674	0.5007	0.5001	0.6384	0.6485	0.7212
US	0.8029	0.8491	0.9036	**0.9176**	0.9179	**0.9307**	0.2952	0.4263	0.6074	0.6893	**0.7169**	0.7080
OS	0.5805	0.7333	0.8749	0.9036	0.9181	0.9188	0.0602	0.1943	0.3910	0.5007	0.6295	0.6322
US + CS	0.8041	0.8463	0.9019	0.9163	0.9191	0.9220	0.5394	0.4760	0.6655	0.6647	0.6995	0.7201
OS + CS	0.7644	0.8249	0.8916	0.9065	0.9223	0.9260	0.3645	0.4225	0.5739	0.6147	0.6246	0.6913
SMOTE [6]	0.6208	0.7685	0.8807	0.8895	0.9167	0.9208	0.0586	0.4533	0.6375	0.5625	0.6708	0.6674
CBL [7] (β=0.9)	0.6736	0.7771	0.8867	0.9061	0.9118	0.9206	0.1342	0.4006	0.5068	0.5624	0.6187	0.6890
CBL [7] (β=0.99)	0.7012	0.8021	0.8938	0.9118	0.9178	0.9226	0.3259	0.5392	0.6001	0.6196	0.6369	0.6600
CBL [7] (β=0.999)	0.7692	0.8250	0.8922	0.9122	0.9179	0.9220	0.3492	0.5256	0.6105	0.5400	0.5937	0.7212
CBL [7] (β=0.9999)	0.7885	0.8099	0.8977	0.9127	0.9241	0.9226	0.3562	0.5950	0.5138	0.5933	0.6547	**0.7344**
(ours) ALT Mode	**0.8240**	0.8520	0.8900	0.8880	0.8520	0.8920	0.5120	0.5120	0.5640	0.6340	0.6940	0.5960
(ours) AUG Mode	0.8060	**0.8740**	**0.9140**	0.9160	**0.9340**	0.9220	**0.5940**	**0.6040**	**0.6680**	**0.7060**	0.7040	0.7180

AUG > ALT

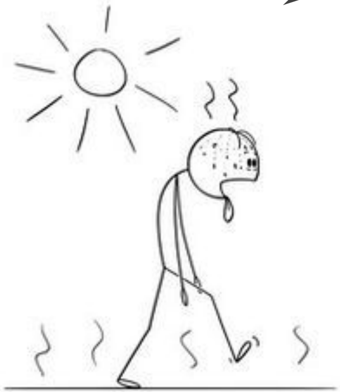
#Zebra	Vanilla	TS	CS	US	OS	US+CS	OS+CS	SMOTE [6]	CBL [7]				Ours	
									$\beta = 0.9$	$\beta = 0.99$	$\beta = 0.999$	$\beta = 0.9999$	ALT	AUG
25	0.0500	0.0340	0.7925	0.1333	0.4750	02865	0.7366	0.2108	0.7613	0.8116	0.8263	0.8120	0.8040	**0.8500**
50	0.5580	0.6260	0.8329	0.7891	0.8012	0.7343	0.8522	0.7298	0.8347	0.8316	0.8222	0.8192	0.8320	**0.8620**
75	0.7180	0.7040	0.8699	0.8457	0.8370	0.8644	0.8680	0.7902	0.8562	0.8587	0.8524	0.8633	0.8020	**0.8780**
100	0.7680	0.7940	0.8684	0.8521	0.8628	**0.8711**	0.8735	0.8254	0.8446	0.8595	0.8683	0.8657	0.8220	0.8520

Heat Maps

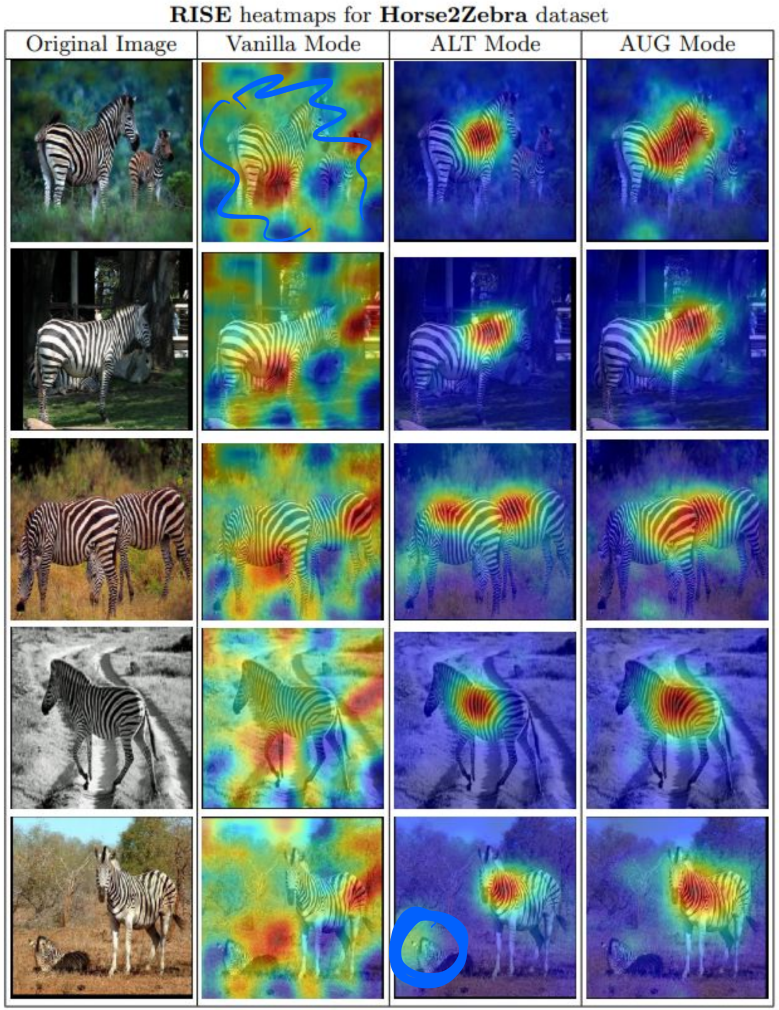
RISE

Grad Cam

- Vanilla is distracted towards the grass (**context**)
- ALT & AUG put more emphasis on Zebra (**content**)



t-SNE Visualizations...



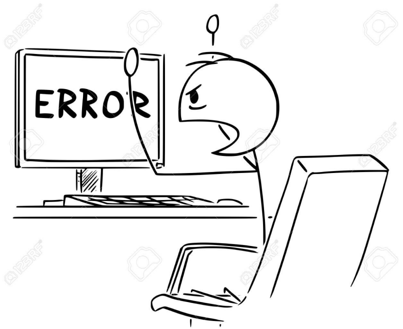
Summing up

- ➢ Data Balancing
- ➢ Retains **context** and changing only **content**
- ➢ **Joint training** of cycle-GAN and classifier (**robustness**)
- ➢ **Two training** method alternatives
- ➢ Merges **cost sensitive load balancing** as well (**γ**) C S

DrawBacks

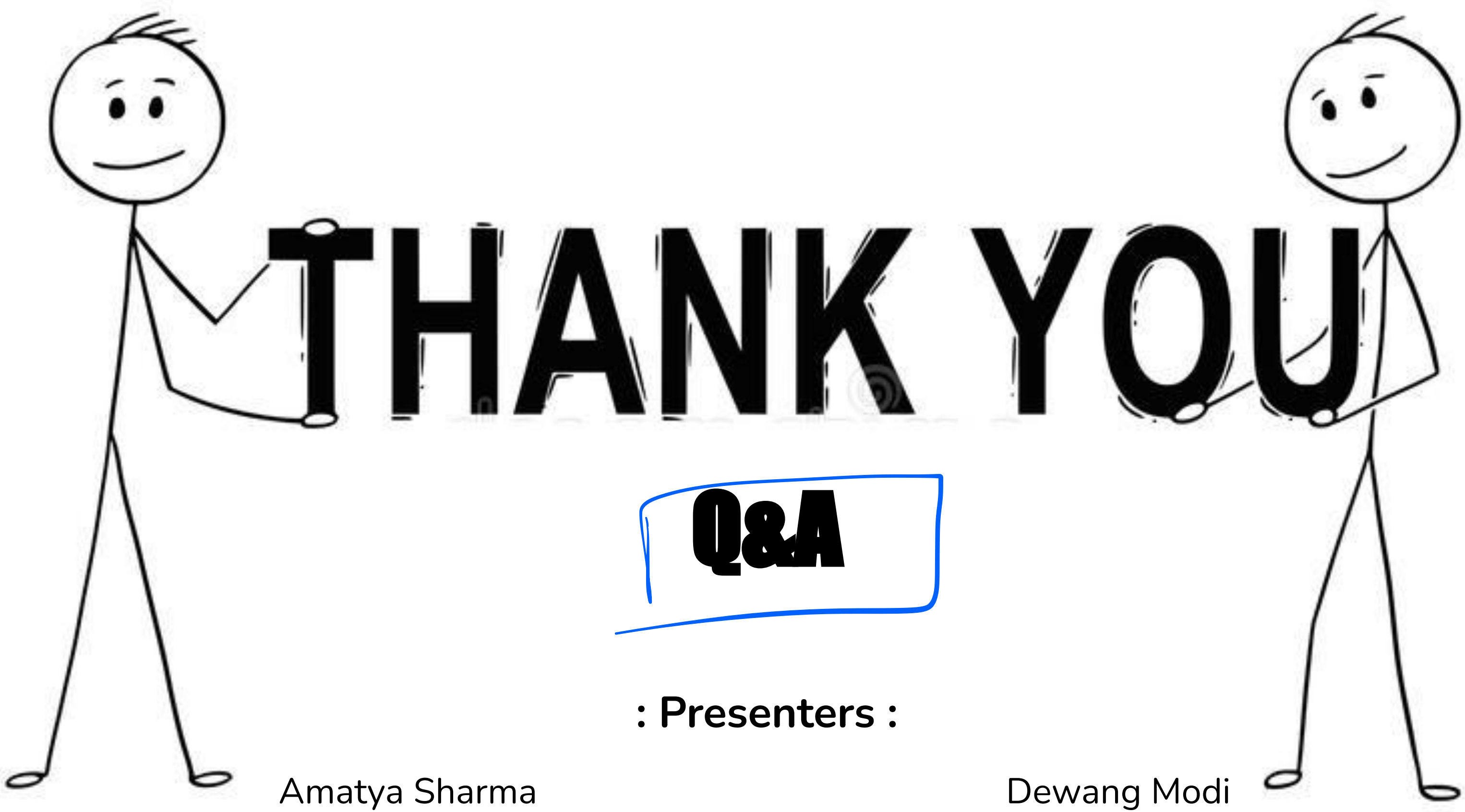
- **Computational cost** of training the GAN

- Difficult to evaluate
    - Tough to compute **Likelihood**
    - **Vanishing gradient** problem for some hyperparam
    - **Boundary distortion**



Future Directions

- Extension from classification task to other CV tasks like **segmentation** (pix2pix)
- **Multiclass** classification ??? Star GAN
- Even applied to **balanced dataset** to
  - Diversify
- Our Model + CBL or learn the weights (γ) to $\mathbf{L}_{\mathbf{cls}}$



Q&A

: Presenters :

Amatya Sharma

Dewang Modi