# Some Papers about Reweighting

Narsil Zhang

Oct. 2019

<□> <@> < E> < E> < E</p>

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	oo		0000	00	000

▲□▶ ▲□▶ ▲ 臣▶ ▲ 臣▶ ― 臣 … のへぐ

# Outline

Foreword

Focal Loss

GHM

Class-balanced loss

Robust Learning via Reweight

Meta-Weight-Net

MentorNet

 Foreword
 Focal Loss
 GHM
 Class-balanced loss
 Robust Learning via Reweight
 Meta-Weight-Net
 MentorNet

 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •
 •

◆ロト ◆御 ト ◆臣 ト ◆臣 ト ○臣 - のへで

Reweight is a method, not our goal. Goal: class imbalance, noisy data(label)

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
O	●0	oo	00	0000	00	000
Probl	ems					

- imbalance of number of data across each class
- the gradient information is dominant by numerous "easy" examples

▲□▶ ▲□▶ ▲ 臣▶ ▲ 臣▶ ― 臣 … のへぐ

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	○●	oo	00		00	000
Focal	Loss [4	4]				



Figure: Focal Loss, t denotes class

▲□▶ ▲□▶ ▲ □▶ ▲ □▶ ▲ □ ● ● ● ●



#### Gradient Harmonized Mechanism [3]

$$L_{GHM} = \sum_{i=1}^{N} \frac{L_{CE}(p_i)}{GD(g_i)}$$

where  $g_i$  is the gradient norm relevant with *i*-th data and GD denotes some "gradient density" estimation

・ロト・日本・日本・日本・日本・日本

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	○●	00	0000	00	000
Effect	S					



GHM can suppress those extremely hard examples.

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	00	•0	0000	00	000

### Effective number of samples

Denote the effective number of samples  $E_n$  is the expected volume of n samples.

Proposition

$$E_n = \frac{1-\beta^n}{1-\beta}, \quad \beta = \frac{n-1}{n}$$

▲ロト ▲御 ト ▲ 臣 ト ▲ 臣 ト 二 臣 … のへで

Foreword Fo	cal Loss G	GHM (	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0 00	) (	00	•0	0000	00	000

### Effective number of samples

Denote the effective number of samples  $E_n$  is the expected volume of n samples.

Proposition

$$E_n = \frac{1-\beta^n}{1-\beta}, \quad \beta = \frac{n-1}{n}$$

#### Proof.

Induction. Denote  $p = \frac{E_{n-1}}{N}$ , then

$$E_n = pE_{n-1} + (1-p)(E_{n-1}+1) = 1 + \frac{N-1}{N}E_{n-1}$$

▲□▶ ▲□▶ ▲ 臣▶ ▲ 臣▶ ― 臣 … のへで

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	oo	○●	0000	00	000
Class	-balanc	ed Ic	ss[1]			

Suppose class y has  $n_y$  training samples, the class-balanced (CB) softmax cross-entropy loss is:

$$CB_{\text{softmax}}\left(\mathbf{z}, y\right) = -\frac{1-\beta}{1-\beta^{n_y}} \log \left(\frac{\exp\left(z_y\right)}{\sum_{j=1}^{C} \exp\left(z_j\right)}\right)$$

where class y has  $n_y$  samples; In practice  $\beta$  is hyperparamenter.

▲□▶ ▲□▶ ▲ 臣▶ ▲ 臣▶ ― 臣 … のへぐ

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	00	○●	0000	00	000
Class-	-balanc	ed Io	ss[1]			

Suppose class y has  $n_y$  training samples, the class-balanced (CB) softmax cross-entropy loss is:

$$CB_{\text{softmax}}\left(\mathbf{z}, y\right) = -\frac{1-\beta}{1-\beta^{n_y}} \log \left(\frac{\exp\left(z_y\right)}{\sum_{j=1}^{C} \exp\left(z_j\right)}\right)$$

where class y has  $n_y$  samples; In practice  $\beta$  is hyperparamenter.

\*ロ \* \* @ \* \* ミ \* ミ \* ・ ミ \* の < ?

Class-balanced loss can be combined with focal loss.

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	00	00	0000	00	000

# **Bilevel Optimization**

$$egin{aligned} & heta^*(w) = rg\min_{ heta} \sum_{i=1}^N w_i f_i( heta) \ & w^* = rg\min_{w,w \geq 0} rac{1}{M} \sum_{i=1}^M f_i^v\left( heta^*(w)
ight) \end{aligned}$$

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	oo		○●○○	00	000

#### Influence analysis

$$f_{i,\epsilon}(\theta) = \epsilon_i f_i(\theta)$$
$$\hat{\theta}_{t+1}(\epsilon) = \theta_t - \alpha \nabla \sum_{i=1}^n f_{i,\epsilon}(\theta) \bigg|_{\theta = \theta_t}$$
$$\epsilon_t^* = \arg \min_{\epsilon} \frac{1}{M} \sum_{i=1}^M f_i^{\nu}(\theta_{t+1}(\epsilon))$$

▲□▶ ▲□▶ ▲ □▶ ▲ □▶ ▲ □ ● ● ● ●

v denotes validation loss.

Foreword Focal Loss GHM Class-balanced loss Robust Learning via Reweight Meta-Weight-Net Ooo Learning to Reweight Examples for Robust Deep Learning[5]

$$u_{i,t} = -\eta \frac{\partial}{\partial \epsilon_{i,t}} \frac{1}{m} \sum_{j=1}^{m} f_{j}^{\nu} \left(\theta_{t+1}(\epsilon)\right) \bigg|_{\epsilon_{i,t}=0}$$
$$w_{i,t} = \frac{|u_{i,t}|}{\sum_{j} |u_{j,t}|}$$

◆ロト ◆御 ト ◆臣 ト ◆臣 ト ○臣 - のへで

Foreword 0 Robust Learning via Reweight 000● Meta-Weight-Net 00 MentorNet 000



*Figure 1.* Computation graph of our algorithm in a deep neural network, which can be efficiently implemented using second order automatic differentiation.

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	oo	00	0000	●0	000
Meta	-NN					

Use a meta-NN  $\mathcal{V}(L; \Theta)$  to give the reweight coefficient for loss term L.

$$\mathbf{w}^{*}(\Theta) = \underset{\mathbf{w}}{\arg\min} \mathcal{L}^{\mathsf{train}}(\mathbf{w}; \Theta) \triangleq \frac{1}{N} \sum_{i=1}^{N} \mathcal{V}\left(L_{i}^{\mathsf{train}}(\mathbf{w}); \Theta\right) \cdot L_{i}^{\mathsf{train}}(\mathbf{w})$$

▲□▶ ▲□▶ ▲ 臣▶ ▲ 臣▶ ― 臣 … のへぐ

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	00		0000	●0	000
Meta	-NN					

Use a meta-NN  $\mathcal{V}(L; \Theta)$  to give the reweight coefficient for loss term *L*.

$$\mathbf{w}^{*}(\Theta) = \underset{\mathbf{w}}{\operatorname{arg\,min}} \mathcal{L}^{\operatorname{train}}(\mathbf{w}; \Theta) \triangleq \frac{1}{N} \sum_{i=1}^{N} \mathcal{V}\left(L_{i}^{\operatorname{train}}(\mathbf{w}); \Theta\right) \cdot L_{i}^{\operatorname{train}}(\mathbf{w})$$

Alternatively update  $\boldsymbol{w}$  and  $\boldsymbol{\Theta}$  to minimize loss

$$\Theta^* = \operatorname*{arg\,min}_{\Theta} \mathcal{L}^{\mathsf{meta}}\left(\mathbf{w}^*(\Theta)
ight) riangleq rac{1}{M} \sum_{i=1}^M \mathcal{L}^{\mathsf{meta}}_i\left(\mathbf{w}^*(\Theta)
ight)$$

▲□▶ ▲□▶ ▲臣▶ ▲臣▶ 三臣 - のへぐ

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	00	00	0000	○●	000

## Meta-weight-net[6]



Figure 2: Main flowchart of the proposed MW-Net Learning algorithm (steps 5-7 in Algorithm 1).

・ロト ・ 同ト ・ ヨト ・ ヨト

Э

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
O	00	00	00	0000	00	●00
Currio	culum					

$$\min_{\mathbf{w},\mathbf{v}} \mathbb{F}(\mathbf{w},\mathbf{v}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{v}_{i}^{T} \mathbf{L} \left( \mathbf{y}_{i}, g_{s} \left( \mathbf{x}_{i}, \mathbf{w} \right) \right) + G(\mathbf{v}; \lambda) + \theta \|\mathbf{w}\|_{2}^{2}$$

<□▶ < □▶ < 三▶ < 三▶ = 三 のへぐ

Each G gives a curriculum.

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	oo		0000	00	0●0
Ment	orNet[2	2]				

Use an nn (MentorNet  $g(; \Theta)$ ) to learn data-driven curriculum :

$$g_m(\mathsf{z}_i;\Theta^*) = rgmin_{v_i\in[0,1]}\mathbb{F}(\mathsf{w},\mathsf{v}), orall i\in[1,n]$$

or

$$\Theta^{*} = \arg\min_{\Theta} \sum_{(\mathbf{x}_{i}, y_{i}) \in \mathcal{D}} g_{m}(\mathbf{z}_{i}; \Theta) \ell_{i} + G(g_{m}(\mathbf{z}_{i}; \Theta); \lambda)$$

▲□▶▲圖▶▲≣▶▲≣▶ ≣ のへで

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
0	00	oo		0000	00	00●
Refer	ences I					

- Y. CUI, M. JIA, T.-Y. LIN, Y. SONG, AND S. BELONGIE, *Class-balanced loss based on effective number of samples*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 9268–9277.
- L. JIANG, Z. ZHOU, T. LEUNG, L. LI, AND L. FEI-FEI, Mentornet: Regularizing very deep neural networks on corrupted labels, CoRR, abs/1712.05055 (2017).
- B. LI, Y. LIU, AND X. WANG, *Gradient harmonized* single-stage detector, CoRR, abs/1811.05181 (2018).
- T. LIN, P. GOYAL, R. B. GIRSHICK, K. HE, AND P. DOLLÁR, *Focal loss for dense object detection*, CoRR, abs/1708.02002 (2017).

Foreword	Focal Loss	GHM	Class-balanced loss	Robust Learning via Reweight	Meta-Weight-Net	MentorNet
O	00	oo	00	0000	00	○○●
Refer	ences I	I				

- M. REN, W. ZENG, B. YANG, AND R. URTASUN, *Learning to reweight examples for robust deep learning*, arXiv preprint arXiv:1803.09050, (2018).
- J. SHU, Q. XIE, L. YI, Q. ZHAO, S. ZHOU, Z. XU, AND D. MENG, *Meta-weight-net: Learning an explicit mapping for sample weighting*, arXiv preprint arXiv:1902.07379, (2019).

▲ロ ▶ ▲ □ ▶ ▲ □ ▶ ▲ □ ▶ ▲ □ ▶ ● ○ ○ ○