Supplementary Material for Balanced Sharpness-Aware Minimization for Imbalanced Regression

Anonymous ICCV submission

Paper ID 8047

In this supplementary material, we provide comprehensive details to support our main manuscript. Specifically, we present: (1) detailed formulations of our evaluation metrics including MAE, GM, RMSE, and δ_1, which thoroughly assess model performance across different aspects of the prediction distribution; (2) mathematical foundations of our reweighting strategies, including inverse-frequency weighting (INV) and square-root-inverse weighting (SQINV); and (3) the label distributions in the AgeDB-DIR, IMDB-WIKI-DIR and NYUD2-DIR benchmarks, demonstrating the prevalence and characteristics of regression imbalance in real-world vision tasks.

S1. Evaluation Metrics

We employ multiple complementary metrics to evaluate the performance of our proposed method. Let $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^N$ denote the test dataset where:

- $x_i \in \mathcal{X}$ represents the input image;
- $y_i \in \mathcal{Y} \subset \mathbb{R}$ represents the ground truth regression target;
- $\hat{y}_i \in \mathcal{Y}$ represents the predicted value;
- N denotes the total number of samples.

S1.1. Mean Absolute Error (MAE)

MAE measures the average magnitude of errors in prediction without considering their direction:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{1}$$

S1.2. Geometric Mean (GM)

GM provides a measure of the central tendency of the absolute prediction errors by computing their geometric mean:

$$\text{GM} = \left(\prod_{i=1}^{N}(|y_i - \hat{y}_i|)\right)^{\frac{1}{N}}. \tag{2}$$

S1.3. Root Mean Square Error (RMSE)

RMSE emphasizes larger errors due to its quadratic nature:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{3}$$

where $(y_i - \hat{y}_i)^2$ represents the squared difference between the ground truth and predicted value.

S1.4. Threshold Accuracy (δ_1)

δ_1 measures the percentage of predictions within a relative threshold:

$$\delta_1 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}[\max(\frac{y_i}{\hat{y}_i}, \frac{\hat{y}_i}{y_i}) < 1.25], \tag{4}$$

where $\mathbb{1}[\cdot]$ is the indicator function that returns 1 if the condition is true and 0 otherwise. And 1.25 is the threshold for acceptable relative error.

S2. Reweighting Strategies

Let b_k denote the set of samples falling into the k-th interval, and n_k represents the number of samples in the interval b_k.

S2.1. Inverse-frequency Weighting (INV)

INV assigns weights inversely proportional to the frequency of samples:

$$w_{\text{INV}}(k) = \frac{1}{n_k}. \tag{5}$$

S2.2. Square-root-inverse Weighting (SQINV)

SQINV provides a more moderate reweighting scheme:

$$w_{\text{SQINV}}(k) = \sqrt{\frac{1}{n_k}}. \tag{6}$$

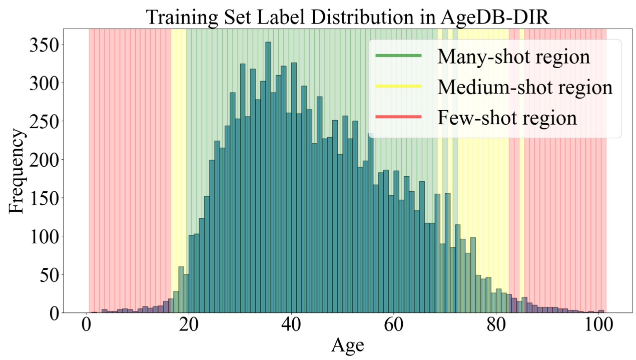

Figure S1. Training set label distribution in AgeDB-DIR.

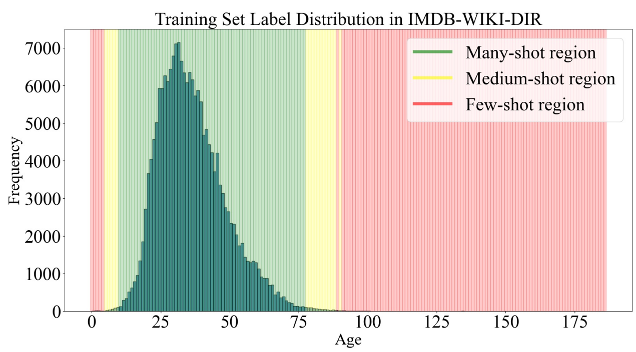

Figure S2. Training set label distribution in IMDB-WIKI-DIR.

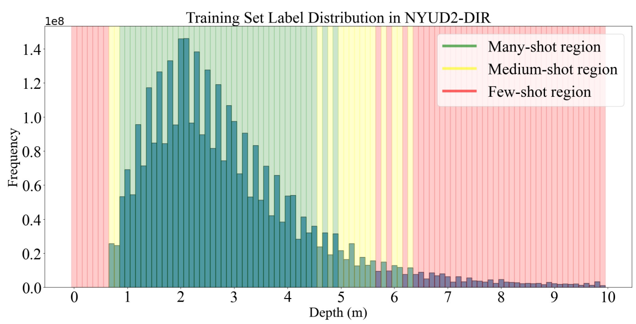

Figure S3. Training set label distribution in NYUD2-DIR.

S3. Dataset Label Distribution Analysis

To comprehensively analyze the imbalanced regression problem in vision tasks, we investigate the label distributions of three representative datasets: AgeDB-DIR (Figure S1), IMDB-WIKI-DIR (Figure S2) and NYUD2-DIR (Figure S3).

S4. More Baselines

We conducted additional experiments on the AgeDB dataset with more approaches, as shown in Tab. S1. The results further demonstrate the superiority of BSAM over stronger baselines.

Table S1. Comparisons with more baselines by MAE metric.

Methods	All	Many	Med.	Few
Vanilla	6.690	5.959	7.740	10.688
TERM [1]	6.518	5.935	7.304	9.848
RRT	6.631	5.957	7.617	10.270
Focal-R [2]	6.565	5.837	7.658	10.427
BalanceMSE (GAI) [3]	6.541	6.036	6.927	10.243
BalanceMSE (BMC) [3]	6.616	5.961	7.313	10.868
BSAM	**6.067**	**5.801**	**6.304**	**7.928**

S5. Limitation

Due to the limitations of existing imbalanced regression benchmarks, we have currently validated our method only on univariate imbalanced regression tasks. Multivariate imbalanced regression should be considered in future work.

References

[1] Tian Li, Ahmad Beirami, Maziar Sanjabi, and Virginia Smith. Tilted empirical risk minimization. *arXiv preprint arXiv:2007.01162*, 2020. 2

[2] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988, 2017. 2

[3] Jiawei Ren, Mingyuan Zhang, Cunjun Yu, and Ziwei Liu. Balanced mse for imbalanced visual regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7926–7935, 2022. 2