

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/281811250>

# Extended Abstract – Handling Data Imbalance in Automatic Facial Action Intensity Estimation

Dataset · September 2015

---

READS

7

2 authors:



[Frerk Saxen](#)

Otto-von-Guericke-Universität Magdeburg

11 PUBLICATIONS 12 CITATIONS

[SEE PROFILE](#)



[Philipp Werner](#)

Otto-von-Guericke-Universität Magdeburg

24 PUBLICATIONS 43 CITATIONS

[SEE PROFILE](#)

# Handling Data Imbalance in Automatic Facial Action Intensity Estimation

Philipp Werner  
Philipp.Werner@ovgu.de

Frerk Saxen  
Frerk.Saxen@ovgu.de

Ayoub Al-Hamadi  
Ayoub.Al-Hamadi@ovgu.de

Otto von Guericke University  
Magdeburg, Germany  
www.iikt.ovgu.de/nit

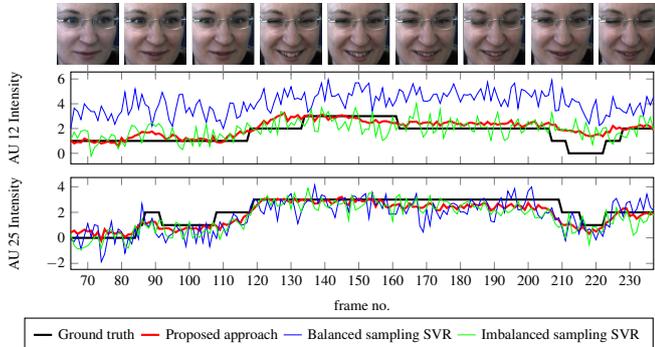


Figure 1: Intensity estimation on test set for AU 12 (lip corner pulling) and 25 (lip parting). Our approach, the MIDRUS SVR ensemble (red), is compared with balanced sampling SVR (green), imbalanced sampling SVR (blue), and ground truth (black).

Automatic Action Unit (AU) intensity estimation is a key problem in facial expression analysis. But limited research attention has been paid to the inherent class imbalance, which usually leads to suboptimal performance. To handle the imbalance, we propose (1) a novel multiclass under-sampling method and (2) its use in an ensemble. We compare our approach with state of the art sampling methods used for AU intensity estimation. Multiple datasets and widely varying performance measures are used in the literature, making direct comparison difficult. To address these shortcomings, we compare different performance measures for AU intensity estimation and evaluate our proposed approach on three publicly available datasets, with a comparison to state of the art methods along with a cross dataset evaluation.

**Sampling Strategy** On the one hand strong imbalance decreases performance on the minority class(es), and on the other hand under-sampling may drop relevant information about the majority class(es). We propose to choose a compromise. Instead of removing the imbalance or ignoring it, we reduce it with a method that we call *Multiclass Imbalance Damping Random Under-Sampling (MIDRUS)*. It is an algorithm with two steps: (1) calculating the number of samples to select from each class, and (2) randomly under-sample the classes without repetition according to the counts calculated in step (1).

Given that we have  $M$  classes  $i = 1, \dots, M$  and  $n_i$  is the absolute frequency of class  $i$  in the dataset, then the number of samples  $n_i^*$  to select from class  $i$  is calculated as follows.

$$n_i^- = \lceil s \cdot (n_i)^{1-\alpha} \rceil, \text{ with } s = \beta \frac{n_{f(k)}}{(n_{f(k)})^{1-\alpha}}, \quad (1)$$

$$n_i^* = \min\{n_i, n_i^-\}. \quad (2)$$

In (1),  $\alpha \in [0, 1]$  is the imbalance damping parameter. It controls to which extend the imbalance is reduced, i.e.  $\alpha = 1$  aims at total balancing of classes,  $\alpha = 0$  keeps the imbalance, and an  $\alpha$  in between reduces it to a certain degree. With  $\alpha > 0$ , the term  $(n_i)^{1-\alpha}$  calculates new and more balanced class ratios. Next, these are scaled by a common factor  $s$ , which controls the total number of samples to be selected. Fig. 2 illustrates a typical use-case of MIDRUS with  $\alpha = 0.5$ ; the imbalance is damped by taking the square root of sample counts and scaling the results in a way that 75% of the second most frequent class’s samples are selected.

**Ensemble** It is state of the art to handle data imbalance by combining ensemble methods with sampling strategies. We propose to combine MIDRUS with an ensemble to further improve predictive performance.

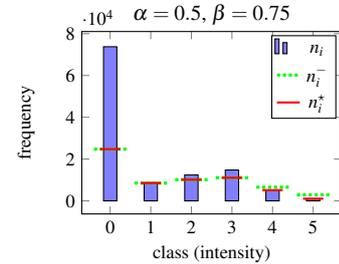


Figure 2: Multiclass Imbalance Damping Random Under-Sampling (MIDRUS) example. MIDRUS damps imbalance, which improves performance (see Fig. 3).  $\beta$  adjusts the number of selected samples.

Dataset Measure	Bosphorus ICC(3,1) <sub>d</sub>	DISFA ICC(3,1) <sub>d</sub>	UNBC PCC <sub>c</sub>
Mavadati [3]		0.235	
Kaltwang [1]			0.306
EasyEnsemble	0.340	0.362	0.301
SVR Ensemble	0.553	0.346	0.286
	0.533	0.412	0.301
	0.603	0.439	0.311

■ imbalanced    ■ ‘balanced’    ■ proposed: MIDRUS

Figure 3: Cross validated performances (mean across AUs).

We use *bagging* and apply MIDRUS  $T$  times to train  $T$  prediction models, each with one of the selected training subsets. For aggregation of the model outputs we train a fusion model by subsampling the training set with MIDRUS again.

Due to the benefit of continuous output we use Support Vector Regression (SVR) models, but the MIDRUS ensemble can also be trained with other models, including classification models.

**Experiments** In Fig. 3 we compare several methods on three databases. On the Bosphorus dataset, EasyEnsemble [2] is clearly outperformed by our proposed SVR Ensemble. MIDRUS improves performance compared to using the originally imbalanced data and to balancing it with the second most frequent class (‘balanced’). On the DISFA database all ensemble methods outperform the person-independent modeling results reported by Mavadati and Mahoor [3]. MIDRUS also performs best on DISFA, but the advantage over EasyEnsemble is lower than for Bosphorus. On the UNBC-McMaster dataset we compare to Kaltwang *et al.* [1] (Relevance Vector Regression on imbalanced data), but do not observe clear benefits of the ensemble methods. In contrast to Kaltwang *et al.*, we use a fully automatic landmark detector and a much simpler alignment, which is less suited for out-of-plane head poses that occur frequently in UNBC-McMaster. A better face alignment would probably improve the results obtained with ensemble methods. Nevertheless, MIDRUS still slightly outperforms the results of Kaltwang.

- [1] S. Kaltwang, O. Rudovic, and M. Pantic. Continuous Pain Intensity Estimation from Facial Expressions. In *Advances in Visual Computing*, LNCS, pages 368–377. Springer Berlin Heidelberg, 2012.
- [2] Xu-Ying Liu, Jianxin Wu, and Zhi-Hua Zhou. Exploratory under-sampling for class-imbalance learning. *IEEE Transactions on System, Man and Cybernetics*, 39(2):539–550, 2009.
- [3] S.M. Mavadati and M.H. Mahoor. Temporal Facial Expression Modeling for Automated Action Unit Intensity Measurement. In *22nd International Conference on Pattern Recognition (ICPR)*, pages 4648–4653, August 2014. doi: 10.1109/ICPR.2014.795.