Computer Vision Laboratory
cvl.cs.nott.ac.uk

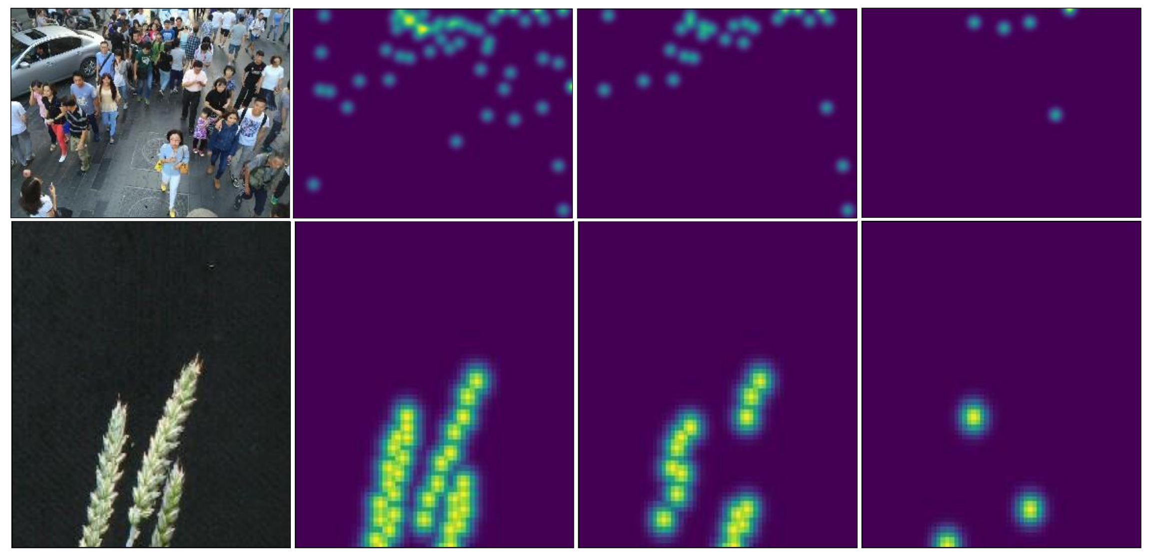
The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Learning to Localise and Count with Incomplete Dot-annotation

Feng Chen[1], Michael P. Pound[1], Andrew P. French[1, 2]
[1]School of Computer Science, University of Nottingham, UK
[2]School of Biosciences, University of Nottingham, UK

Introduction

- Annotating images with many objects (e.g. >100) is tedious and humans can make mistakes (e.g. missing annotations).
- We present an approach which uses **incomplete annotations** (e.g. Fig 1. (c) and (d)) in every training image with an **Asymmetric Mean Squared Error (AMSE)** loss function to tackle the mentioned problems.



(a) Input (b) Original heatmap (c) 50% reduced (d) 90% reduced

Fig 1. Examples of incomplete annotations

AMSE V.S. MSE

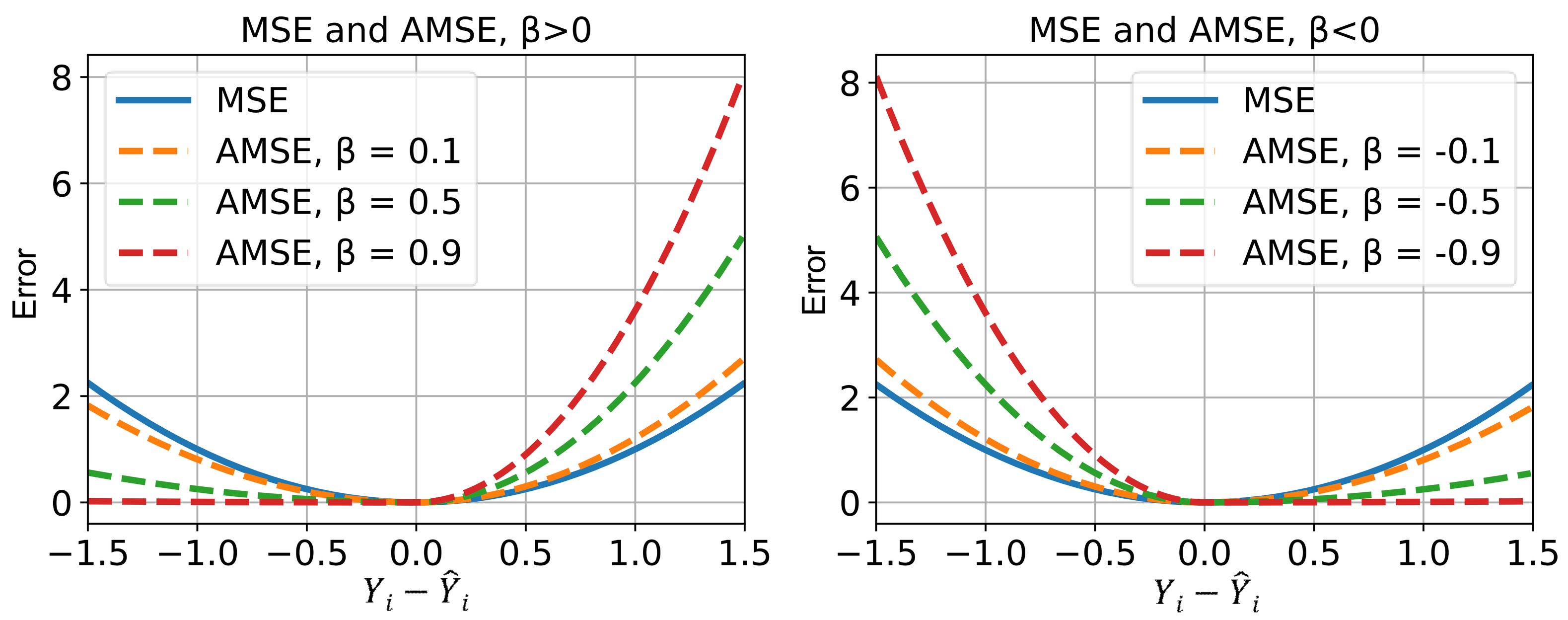


Fig 2. Curves of AMSE (on different β) and MSE

$$AMSE = \frac{1}{N}\sum_{i=1}^{N}\{[\beta + sign(Y_i - \hat{Y}_i)] * (Y_i - \hat{Y}_i)\}^2 \qquad (1)$$

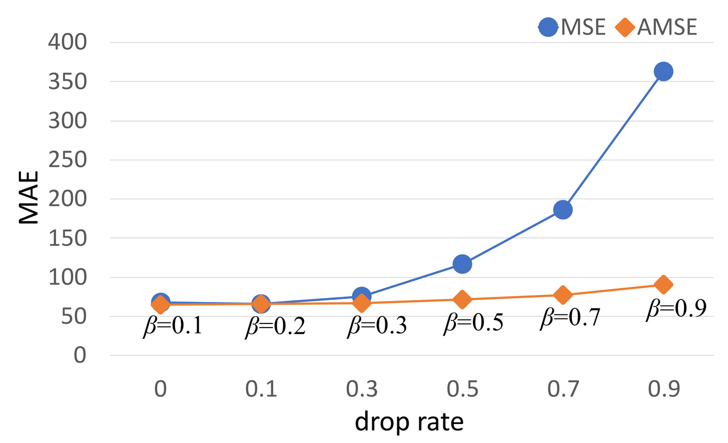
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \qquad (2)$$

Y_i is the ground truth, $\hat{Y}_i$ is the prediction, $\beta \in [-1,1]$ is the only tuning parameter of AMSE, and

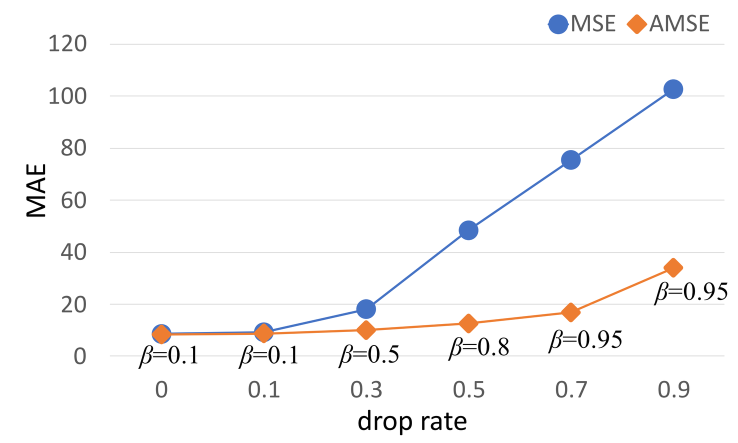
$$sign(Y_i - \hat{Y}_i) = \begin{cases} -1, & Y_i - \hat{Y}_i < 0 \\ 0, & Y_i - \hat{Y}_i = 0 \\ 1, & Y_i - \hat{Y}_i > 0 \end{cases}$$

$\boldsymbol{\beta > 0}$ is adopted because we expect the model to be punished more in the false-negative areas (i.e. objects exist in these areas but they are not annotated).

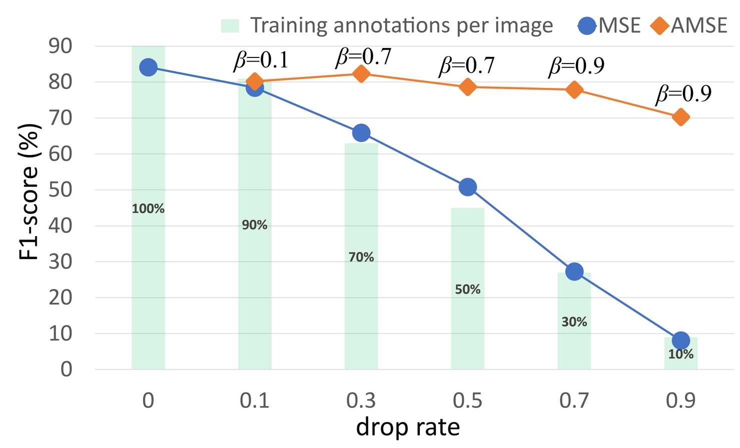
Experiments and Results



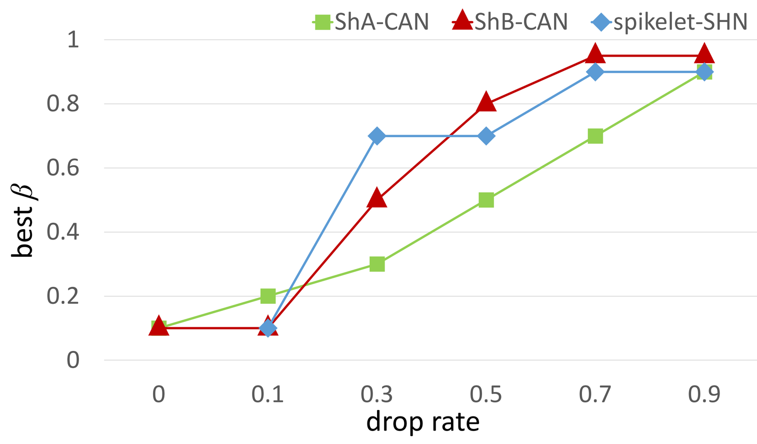
(a) ShA-CAN (MAE v.s. drop rate)



(b) ShB-CAN (MAE v.s. drop rate)



(c) Spikelet-SHN (F1-score v.s. drop rate)

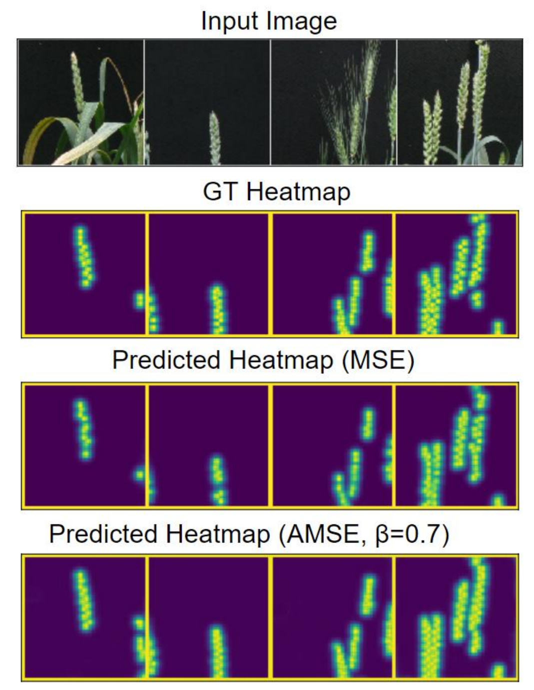


(d) Optimal β for different drop rates

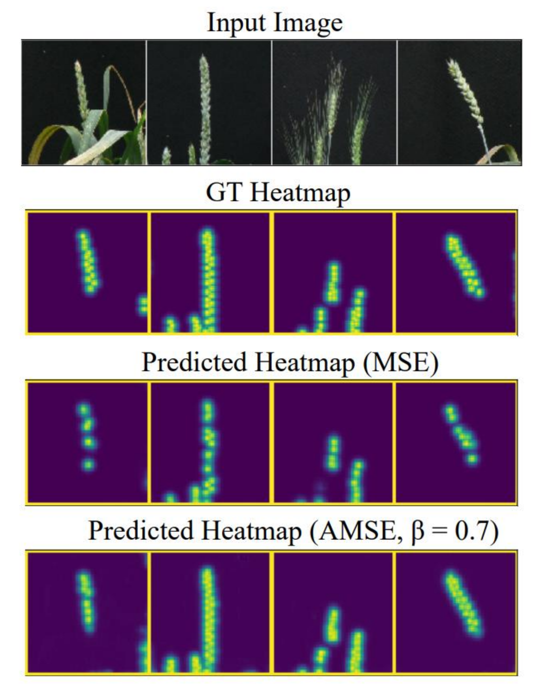
Fig 3. (a – c) Test results on ShA/ShB-CAN (crowd counting) and spikelet-SHN (spikelet localisation) respectively; (d) optimal β for various drop rates on different models/dataset

- We ran experiments on a crowd counting task (ShA/ShB-CAN, Fig 1 first row) and a wheat spikelet localization task (spikelet-SHN, Fig 2 second row)
- We randomly drop annotations from **every** image; the dropped portion is determined by drop rate (dr), where $dr \in [0,1]$
- We compare the performance between MSE and AMSE under different dr; the results are shown in Fig 3.

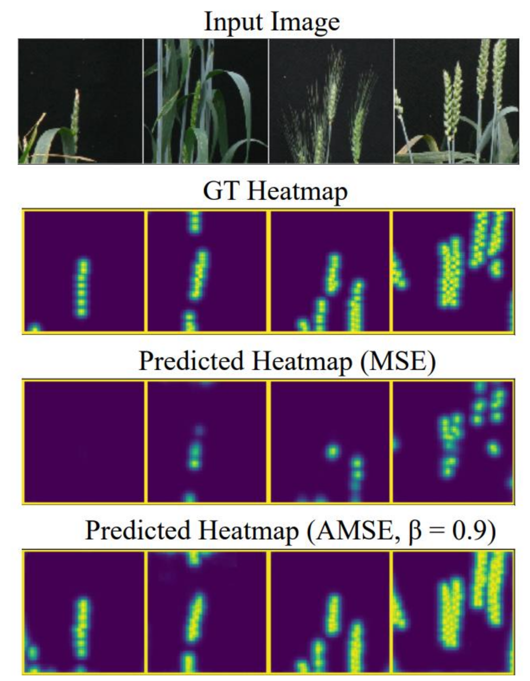
- We also present some visual results of the wheat spikelet localization task in Fig 4.



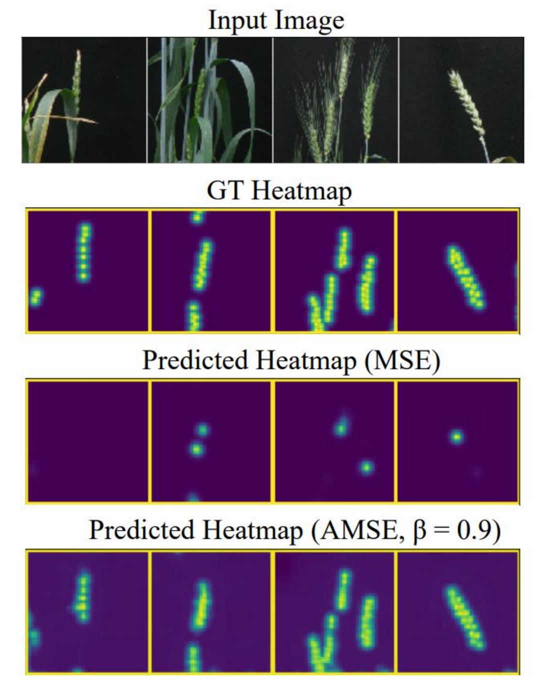
(a) dr = 0.3, optimal $\boldsymbol{\beta}$ = 0.7



(b) dr = 0.5, optimal $\boldsymbol{\beta}$ = 0.7



(c) dr = 0.7, optimal $\boldsymbol{\beta}$ = 0.9



(d) dr = 0.9, optimal $\boldsymbol{\beta}$ = 0.9

Fig 4. Visual results of spikelet localization task on different drop rates

Conclusion

- AMSE (with tuned β) significantly improves the performance of counting/localizing models trained with incomplete annotations per image, even on extreme drop rate (e.g. 0.7 and 0.9).
- For lower drop rates (i.e. <0.5) using AMSE achieves comparable performance as trained on fully-annotated dataset.
- The optimal $\boldsymbol{\beta}$ **is positively related to the drop rate**, though the exact value varies depending on dataset and models.