# Ear density estimation in wheat crops from high resolution RGB images and deep learning methods

Simon Madec
INRA UMR EMMAH
Avignon
Simon.madec@inra.fr

Simon Madec
INRA UMR EMMAH
Avignon
Simon.madec@inra.fr

Frederic Baret
INRA UMR EMMAH
Avignon
frederic.baret@inra.fr

Benoit De Solan
Arvalis
UMR EMMAH Avignon
b.desolan@arvalisinstitutduvegetal.fr

## Abstract

*Ear density estimation in wheat crops is an appealing trait for plant breeders as it is highly correlated with the total aerial biomass [1]. It is also one of the main yield components and can provide a better understanding of its variability.*
*Current methods based on manual counting on a limited sample size of a plot are tedious and prone to human errors and lack of spatial representativeness. Computer vision offers an alternative and desirable high throughput approach that has been evaluated over others species than wheat [2]. However due to the large number of wheat ears per square meter and the partial occlusion observed between ears, wheat ear counting from high resolution RGB images is still a challenging task. High spatial resolution sensors and a careful acquisition of the images are then required.*
*The recent advances in hardware, particularly using GPU processing capacity, along with the availability of huge collection of examples from which computers can learn foster important developments in the field of computer vision and convolutional neural networks (CNN). Nowadays CNN are achieving impressive results for image classification (He et al., 2015). Because the amount of labeling images needs to be very important in order to train CNN models, CNN models pre-trained on popular benchmark computer vision database like Alexnet, Vgg16, ResNet and more recently Googlenet are often used as a starting point. These models are then fine-tuned [5] with a smaller training dataset to be applied to another specific task [6]–[8]. For the purpose of object detection and counting the algorithm also needs to identify each ear in an image. As a starting point, an object proposal method is then required. The basic strategy is to use an exhaustive search but the potential regions can vary in location, size and aspect ratio, providing an enormous number of potential candidates. Recently Faster-RCNN [9] achieved excellent performances for general object detection by using the convolutional features of the full-image network to propose regions. The region proposal and the classification steps are also sharing weight, making the pipeline much faster and accurate than the current state of the art methods.*
*In this study, an automatic method for wheat ear counting with high spatial resolution RGB – images and the faster-RCNN algorithm is developed and evaluated. The availability of the region proposal step implementation along with the pre-processing and the acquisition of the images are described and discussed. In addition, the influence of the used CNN pre-trained model for the classification step is evaluated. The ear density estimated by these techniques is compared to independent assessments for quantifying the associated uncertainties. Finally, discussion of the potential of the deep learning approach for ear counting in wheat crops is discussed along with the possible vectors (UAVs, UGVs) carrying the RGB camera. The resulting CNN-model and the training/validation datasets are also shared for future benchmarking.*

References

[1] J. M. Whaley, D. L. Sparkes, M. J. Foulkes, J. H. Spink, T. Semere, and R. K. Scott, “The physiological response of winter wheat to reductions in plant density,” Ann. Appl. Biol., vol. 137, no. 2, pp. 165–177, 2000.
[2] S. W. Chen et al., “Counting Apples and Oranges With Deep Learning: A Data-Driven Approach,” IEEE Robot. Autom. Lett., vol. 2, no. 2, pp. 781–788, 2017.
[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in

Advances in neural information processing systems, 2012, pp. 1097–1105.
[4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ArXiv14091556 Cs, Sep. 2014.
[5] G. M. Y. Dauphin et al., "Unsupervised and Transfer Learning Challenge: a Deep Learning Approach," in PMLR, 2012, pp. 97–110.
[6] [6] J. Donahue et al., "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition.," in Icml, 2014, vol. 32, pp. 647–655.
[7] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 806–813.
[8] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in Advances in neural information processing systems, 2014, pp. 3320–3328.
[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," ArXiv150601497 Cs, Jun. 2015.