Determining the optimal light conditions and camera parameters for effective weed detection in digital images using artificial neural networks

Chengsong Hu[1]*, Bishwa Sapkota[1], Alex Thomasson[2], and Muthukumar Bagavathiannan[1]

[1]Department of Soil & Crop Sciences, Texas A&M University, College Station, TX; [2]Department of Agricultural & Biological Engineering, Mississippi State University, Starkville, MS. *Contact: huchengsong@tamu.edu

Introduction

- Recent computer vision techniques based on convolutional neural networks (CNNs) are considered as state-of-the-art tools in weed mapping.
- However, their performance has been shown to be sensitive to quality degradation and light inconsistency.
- To determine the minimum quality that should be met for image dataset collection, a camera system can be designed to obtain the images, followed by training data annotation and model performance analysis, but this is a very expensive process.
- A robust alternative is to simulate the image formation pipeline for different qualities (**Farrell et al. 2003**).

Objective

- We focus on determining the influence of image quality and light consistency on the performance of CNNs in weed mapping by simulating the image formation pipeline.

Materials & Methods

Image collection

- We used a 100-megapixel FUJIFILM GFX 100 RGB camera mounted to a Hylio AG-110 drone for image collection.
- Image collection took place at the Texas A&M University research farm in June 2020 from a cotton and a nearby soybean field roughly one month after planting.
- Three light conditions were targeted: sunny-around noon (June 5), sunny-close to sunset (June 4), and fully cloudy (June 5). We denote the collected images as DS_{noon}, DS_{sunset}, and DS_{cloudy}.

Quality simulation

- We consider the raw images we collected as "ground truth" on which all the simulations were conducted.
- Five image degradations were simulated on DS_{noon}: resolution (scale ratio s at 0.707, 0.5, 0.353, and 0.25), exposure (exposure value EV at 2, 2.5, 3 and 3.5), Gaussian blur (standard deviation σ_B at 1.25, 2.5, 5, and 10), motion blur (kernel length l at 3, 5, 7, and 9 pixels), and noise (standard deviation σ_N at 80, 160, 320, and 640).

Neural network training and evaluation

- Mask R-CNN (**He et al. 2017**) is used as a CNN example for object detection and instance segmentation while semantic segmentation is represented by Deeplab-v3 (**Chen et al. 2017**).
- For object detection and instance segmentation, we report the results following the COCO-style average precision (AP) and mean average precision (mAP).
- For semantic segmentation, we report intersection over union (IoU) for each category as well as the mean IoU (mIoU).

Results

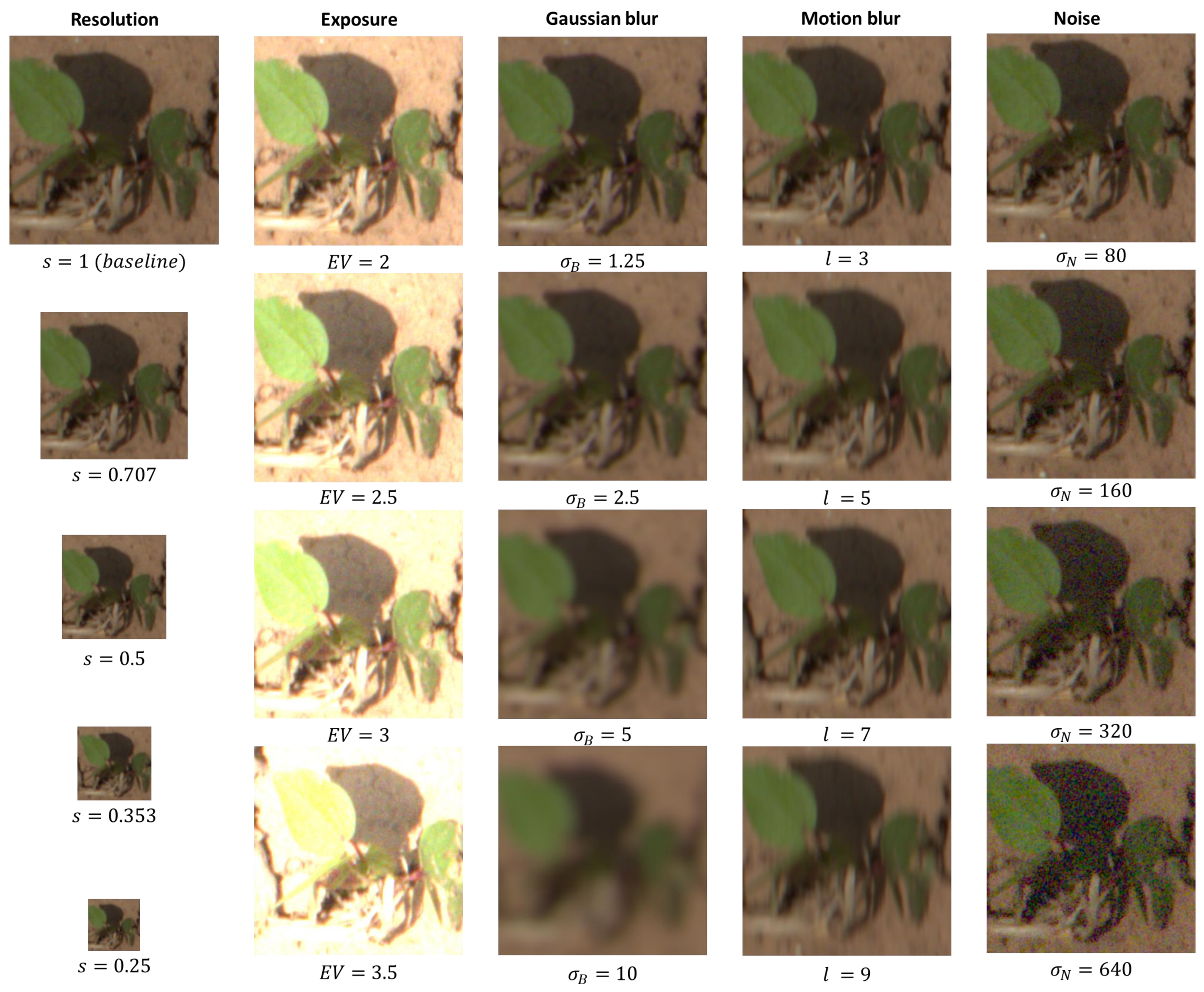

Figure 1. Zoomed-in view of simulation results for reduced resolution, overexposure, Gaussian blur, motion blur, and noise.

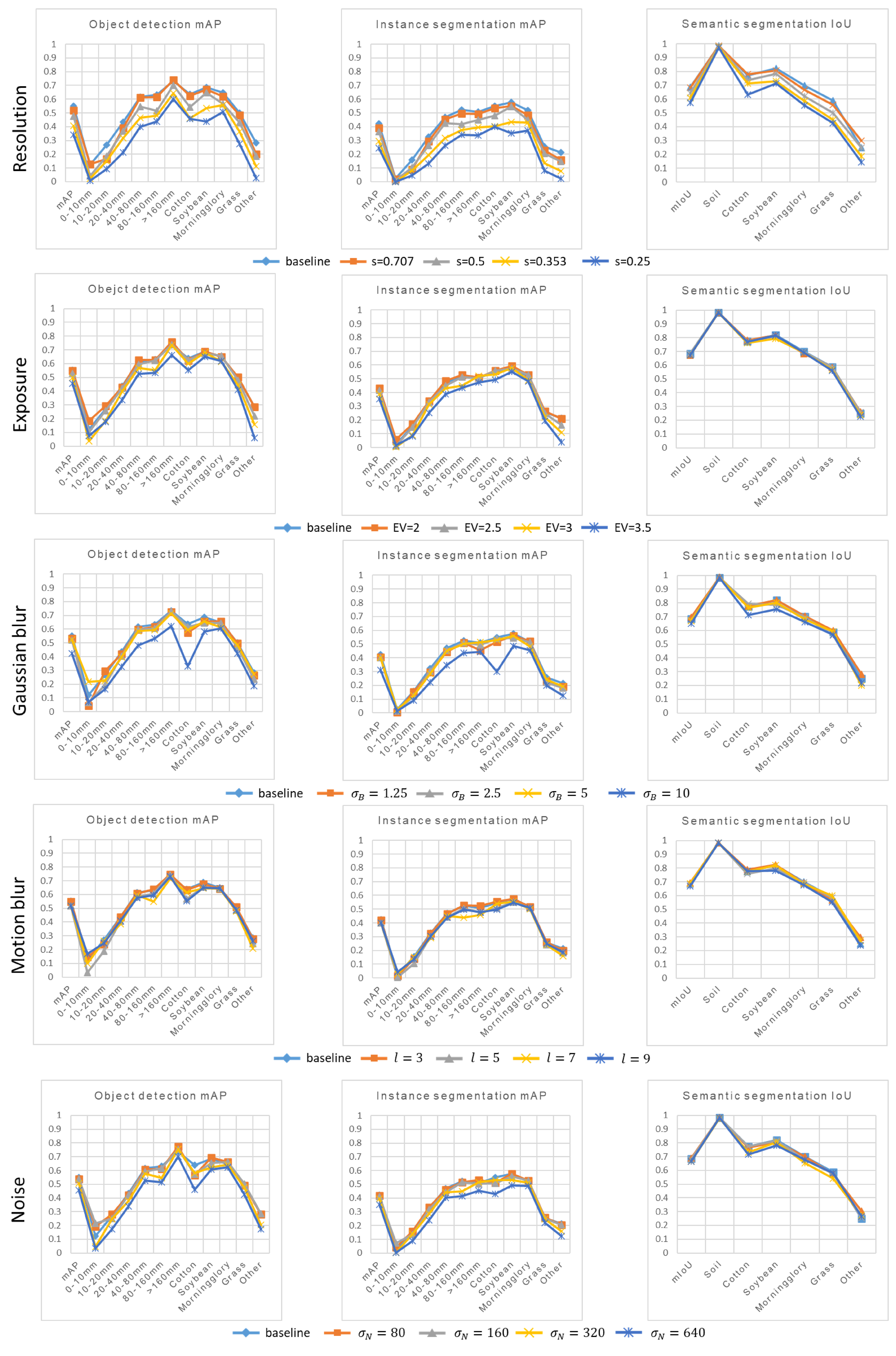

Figure 2. The influence of reduced resolution, overexposure, Gaussian blur, motion blur, and noise on the performance of Mask R-CNN and Deeplab-v3 for object detection, semantic segmentation, and instance segmentation

Table 1. Influence of light inconsistency on the performance of CNN models. Training image sets pertaining to DS_{noon}, DS_{sunset}, DS_{cloudy} were used to train the CNN models, which were then applied on 500 DS_{noon} images.

No. training images in DS_{noon}	No. training images in DS_{sunset}	No. training images in DS_{cloudy}	Bounding box mAP (%)	Semantic Segmentation mIoU (%)	Instance segmentation mAP (%)
500	0	0	**50.0**	**68.4**	**39.4**
334	0	0	47.6	67.9	37.4
334	166	0	48.0	**68.4**	37.8
334	0	166	48.4	67.4	38.3
166	0	0	42.1	67.1	33.4
166	166	166	47.1	67.9	36.8
166	334	0	46.7	67.2	36.7
166	0	334	45.8	67.6	36.3
0	166	334	41.4	63.3	32.6
0	334	166	40.6	63.0	31.9
0	500	0	38.7	41.2	30.9
0	0	500	41.5	66.3	32.8

Results (Continued)

- CNN performance is most impacted by resolution, regardless of plant size (**Figure 2**).
- Mask R-CNN is tolerant to low levels of overexposure, Gaussian blur, motion blur, and noise for object detection and instance segmentation tasks (**Figure 2**).
- Deeplab-v3 tolerates overexposure, motion blur, and noise at all tested levels for semantic segmentation (**Figure 2**).
- Light inconsistency reduces CNN performance. Increasing the diversity of light conditions in the training images may alleviate this reduction (**Table 1**)

Discussion & Conclusions

- The results provide insights into the impact of image quality and light consistency on the CNN performance.
- The quality threshold established in this study can be used to guide the selection of camera parameters and light conditions in future weed mapping applications.

Future Research

- Our simulation of image degradation was based on the raw images which inevitably contain noise and blur. The CNN performance achievable on "perfect" images is still unknown. Images with higher quality are needed in future research.
- We only tested weed mapping on young crops and weeds. How CNNs perform in detecting and segmenting mature plants still needs to be determined.

Acknowledgements

References

- Farrell, Joyce E., Feng Xiao, Peter Bert Catrysse, and Brian A. Wandell. "A simulation tool for evaluating digital camera image quality." In Image Quality and System Performance, vol. 5294, pp. 124-131. International Society for Optics and Photonics, 2003.
- He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn." In Proceedings of the IEEE international conference on computer vision, pp. 2961-2969. 2017.
- Chen, Liang-Chieh, George Papandreou, Florian Schroff, and Hartwig Adam. "Rethinking atrous convolution for semantic image segmentation." arXiv preprint arXiv:1706.05587 (2017).