

000  
001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053054  
055  
056  
057  
058  
059  
060  
061  
062  
063  
064  
065  
066  
067  
068  
069  
070  
071  
072  
073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107

# IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition

Anonymous CVPR submission

Paper ID 3627

In this supplementary material, we provide the details of the dataset, the evaluation metrics, and the extended experimental results. We will make the dataset and pre-trained models publicly available.

## 1. Evaluation Metrics

We measure the performance of baseline methods under several metrics including *precision*, *recall*, *F-measure*, *G-mean* [4] and  $M_{AUC}$  [2]. The precision and recall are intuitively the ability of the classifier to not label as positive a sample that is negative, and to find all the positive samples corresponding to one specific class, respectively. Assume  $n_{ij}$  is the number of class  $c_i$  samples that are classified as class  $c_j$ . Then the precision  $P_i$  and recall  $R_i$  of class  $c_i$  can be defined as:  $P_i = \frac{n_{ii}}{\sum_{j=1}^k n_{ji}}$  and  $R_i = \frac{n_{ii}}{\sum_{j=1}^k n_{ij}}$ , where  $k$  is the number of classes. The average precision and recall are as follows:

$$\text{Precision} = \frac{1}{k} \sum_{i=1}^k P_i, \text{Recall} = \frac{1}{k} \sum_{i=1}^k R_i, \quad (1)$$

The *F*-measure combines the precision and the recall as a trade-off where the factor  $\beta = 1.0$  (*F1*) indicates that the recall and precision are equally important:

$$\text{F-measure} = \frac{1}{k} \sum_{i=1}^k \frac{(1 + \beta^2) P_i R_i}{\beta^2 P_i + R_i}. \quad (2)$$

The *G-mean* evaluates the average sensitivity of all classes, which is defined as:

$$\text{G-mean} = \left( \prod_{i=1}^k R_i \right)^{\frac{1}{k}}. \quad (3)$$

It is the root of the product of class-wise sensitivity/recall. We simply set the recall which approaching to 0 to 0.001 to avoid the *G-mean* resolves to zero. As for the area under the curve (AUC) metric, we follow the micro average scheme  $M_{AUC}$  of the definition as in [2], similar to the form of *F*-measure and *G-mean*, which integrates the weighted

average of all labels:

$$M_{AUC} = \frac{2M_P M_R}{M_P + M_R}, \quad (4)$$

where micro average precision  $M_P$  and recall  $M_R$  are defined as:  $M_P = \frac{\sum_{i=1}^k n_{ii}}{\sum_{i=1}^k \sum_{j=1}^k n_{ji}}$  and  $M_R = \frac{\sum_{i=1}^k n_{ii}}{\sum_{i=1}^k \sum_{j=1}^k n_{ij}}$ .

## 2. The IP102 Dataset

We show the full list of names for each *sub-class* in Table 1. Fig. 1 and Fig. 2 illustrate the sample images of each sub-class. Note that most of them have bounding box annotations.

## 3. Experimental Results

Fig. 3 shows the performance of ResNet [1] on different hierarchical labels. Each sub-figure shows the performance of each sub-class under the metrics of precision, recall, and F1. Fig. 4 illustrates the classification results of several baseline methods on 102 classes of insect pests. More detection results of Faster R-CNN [3] are shown in Fig. 5 and Fig. 6. We can observe that some cases are hard to be detected due to the small inter-class variance or overlap of multiple targets.

## References

- [1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1
- [2] Q. Kang, L. Shi, M. Zhou, X. Wang, Q. Wu, and Z. Wei. A distance-based weighted undersampling scheme for support vector machines and its application to imbalanced classification. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–14, 2017. 1
- [3] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):1137–1149, 2017. 1
- [4] Y. Sun, M. S. Kamel, and Y. Wang. Boosting for learning multiple classes with imbalanced class distribution. In *International Conference on Data Mining*, 2006. 1

108  
109  
110162  
163  
164

111

165

112

166

113

167

114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215Table 1. Index and name of each insect pest *sub-class* in the IP102 dataset.

Index	Class Name	Index	Class Name	Index	Class Name
1	rice leaf roller	35	wheat sawfly	69	Xylotrechus
2	rice leaf caterpillar	36	cerodonta denticornis	70	Cicadella viridis
3	paddy stem maggot	37	beet fly	71	Miridae
4	asiatic rice borer	38	flea beetle	72	Thripidae
5	yellow rice borer	39	cabbage army worm	73	Erythroneura apicalis
6	rice gall midge	40	beet army worm	74	Tetranychus cinnbarinus
7	Rice Stemfly	41	Beet spot flies	75	Panonchus citri McGregor
8	brown plant hopper	42	meadow moth	76	Phyllocoptes oleiverus ashmead
9	white backed plant hopper	43	beet weevil	77	Icerya purchasi Maskell
10	small brown plant hopper	44	sericaorient alismots chulsky	78	Unaspis yanonensis
11	rice water weevil	45	alfalfa weevil	79	Ceroplastes rubens
12	rice leafhopper	46	flax budworm	80	Chrysomphalus aonidum
13	grain spreader thrips	47	alfalfa plant bug	81	Parlatoria zizyphus Lucas
14	rice shell pest	48	tarnished plant bug	82	Nipaecoccus vastalor
15	grub	49	capsid	83	Aleurocanthus spiniferus
16	mole cricket	50	lytta polita	84	Tetradacus c Bactrocera minax
17	wireworm	51	legume blister beetle	85	Dacus dorsalis(Hendel)
18	white margined moth	52	blister beetle	86	Bactrocera tsuneonis
19	black cutworm	53	therioaphis maculata Buckton	87	Prodenia litura
20	large cutworm	54	odontothrips loti	88	Adristyrannus
21	yellow cutworm	55	Thrips	89	Phylloconitis citrella Stainton
22	red spider	56	alfalfa seed chalcid	90	Toxoptera citricidus
23	corn borer	57	Pieris canidia	91	Toxoptera aurantii
24	army worm	58	Apolygus lucorum	92	Aphis citricola Vander Goot
25	aphids	59	Aphidoidea	93	Scirtothrips dorsalis Hood
26	Potosiabre vitarsis	60	Viteus vitifoliae	94	Dasineura sp
27	peach borer	61	Colomerus vitis	95	Lawana imitata Melichar
28	english grain aphid	62	Brevipoalpus lewisi McGregor	96	Salurnis marginella Guerr
29	green bug	63	oides decempunctata	97	Deporaus marginatus Pascoe
30	bird cherry-oataphid	64	Polyphagotars onemus latus	98	Chlumetia transversa
31	wheat blossom midge	65	Pseudococcus comstocki Kuwana	99	Mango flat beak leafhopper
32	penthealeus major	66	parathrene regalis	100	Rhytidodera bowrinii white
33	longlegged spider mite	67	Ampelophaga	101	Sternochetus frigidus
34	wheat phloethrips	68	Lycorma delicatula	102	Cicadellidae



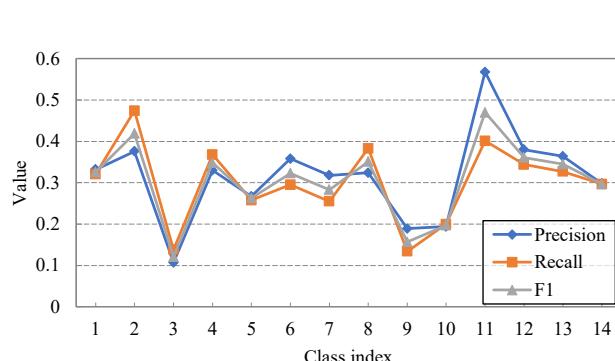
Figure 1. Example images of the IP103 dataset. The number in the gray grid indicates the class index.

CVPR 2019 Submission #3627. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

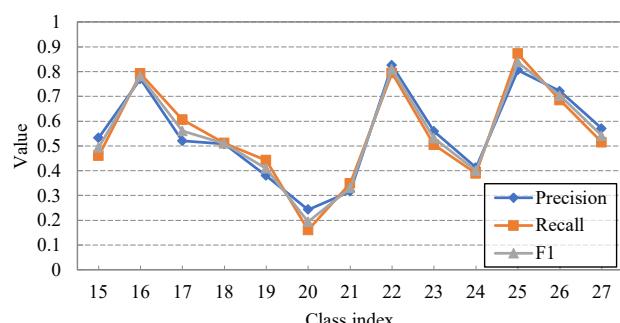


Figure 2. Example images of the IP103 dataset. The number in the gray grid indicates the class index.

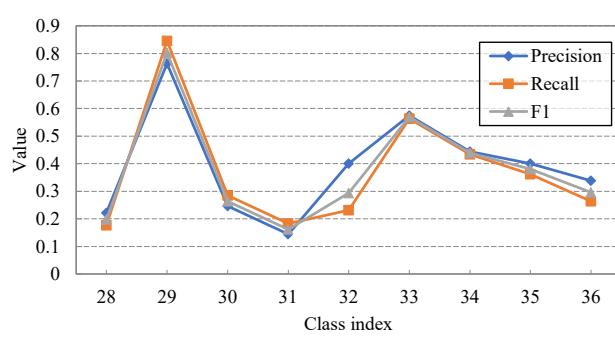
CVPR 2019 Submission #3627. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.



(a)



(b)



(c)



(d)



(e)



(f)



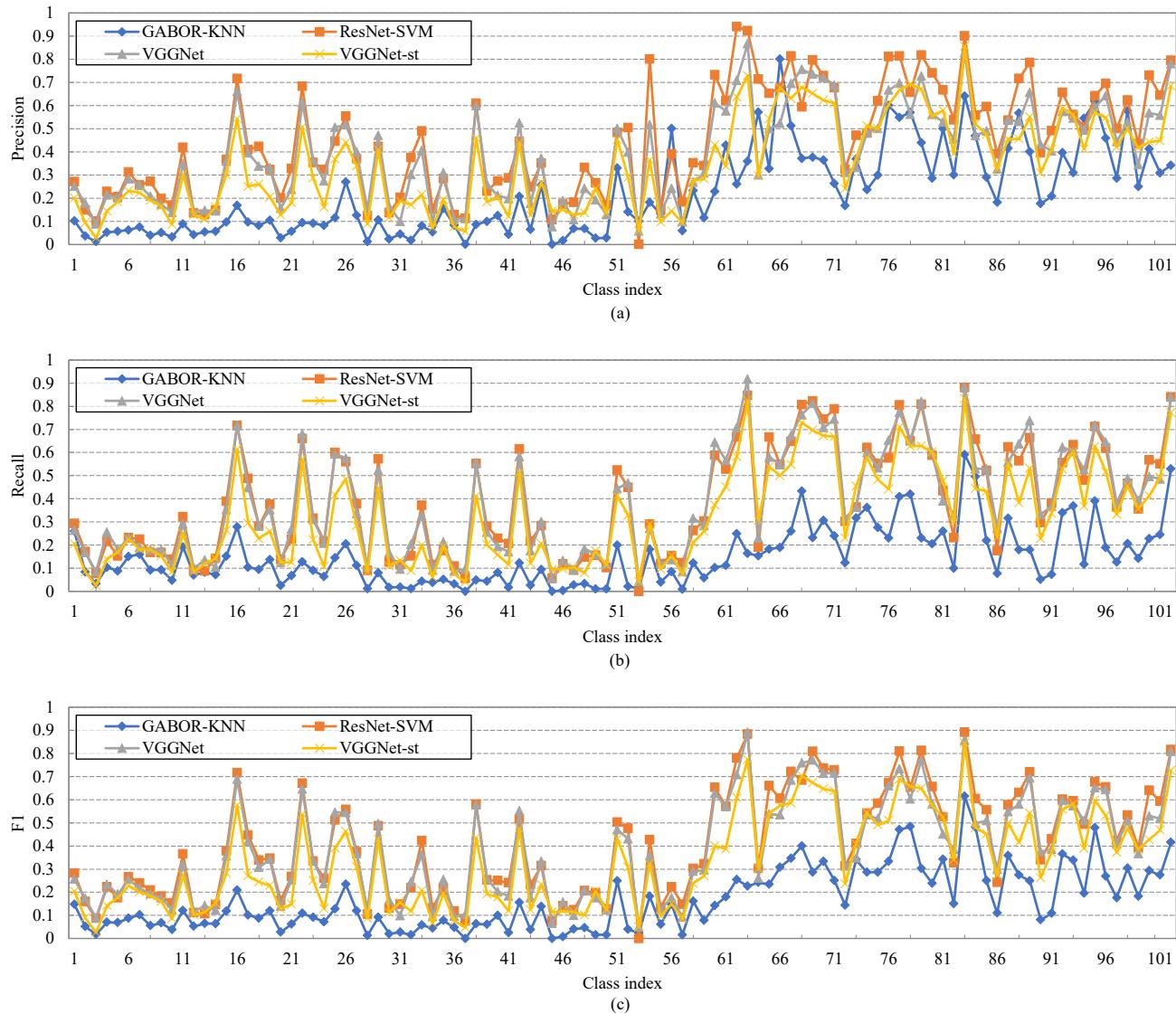
(g)



(h)

482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539

Figure 3. Classification performance of the precision, recall, and F1 metrics with different hierarchical labels. Each sub-figure shows the results of the sub-classes of corresponding crop, i.e., Rice(a), Corn(b), Wheat(c), Beet(d), Alfalfa(e), Vitis(f), Citrus(g), and Mango(h).



584 Figure 4. Classification performance of GABOR with KNN classifier (GABOR-KNN), ResNet with SVM classifier (ResNet-SVM), VG-  
585 GNet, and VGGNet trained from scratch (VGGNet-st) on the metrics of Precision (a), Recall (b), and F1 (c).



Figure 5. Example images of the detection results. The numbers in the images denote the predicted class and probability, respectively. The image with the red frame at the bottom row indicates the failure case.



Figure 6. Example images of the detection results. The numbers in the images denote the predicted class and probability, respectively. The image with the red frame at the bottom row indicates the failure case.