# SUPPLEMENTARY MATERIAL: DETECTION OF SMALL BIRDS IN LARGE IMAGES BY COMBINING A DEEP DETECTOR WITH SEMANTIC SEGMENTATION

*Akito Takeki, Tu Trinh Tuan, Ryota Yoshihashi, Rei Kawakami, Makoto Iida and Takeshi Naemura*

The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan

## 1. RESULTS

In this supplementary document, we show more examples of detection results on the bird image dataset that could not be included in the main paper. Figs. 1–4 include images of misdetections and typical birds that were not detected. The results of individual methods (CNNs, FCNs and SP, when they are solely used) are summarized in Table 1 with the categorization of input image size of birds.

Specifically, Fig. 1 shows the advantage of using CNNs, which can handle normal-size birds as well as tiny ones. Fig. 2 shows difficult birds for detection with any methods, typically those that overlap with background objects. Fig. 3 shows an example of failure cases of FCNs, which usually have difficulty in detecting thin birds. Fig. 4 shows typical regions prone to be misdetected by the methods. All the figures also show that SP is more able to segment the shape of birds correctly. Table 1 summarizes the advantage of individual methods.

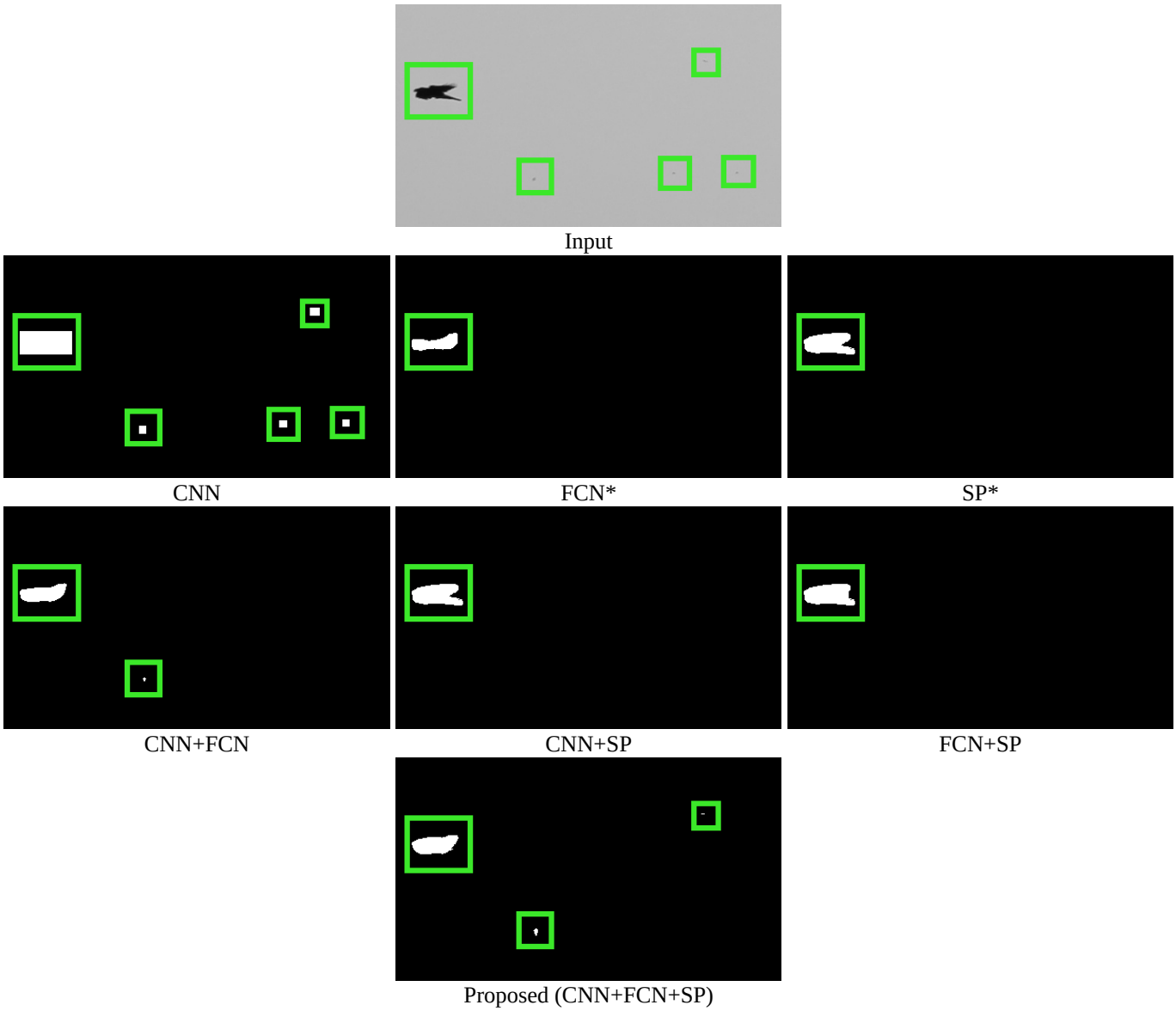


**Fig. 1**. Examples of tiny bird detection results. The green squares mean true positives. It shows that CNNs can detect normal-size as well as tiny-size birds. SP can segment the shape of birds well, as also shown in other figures.

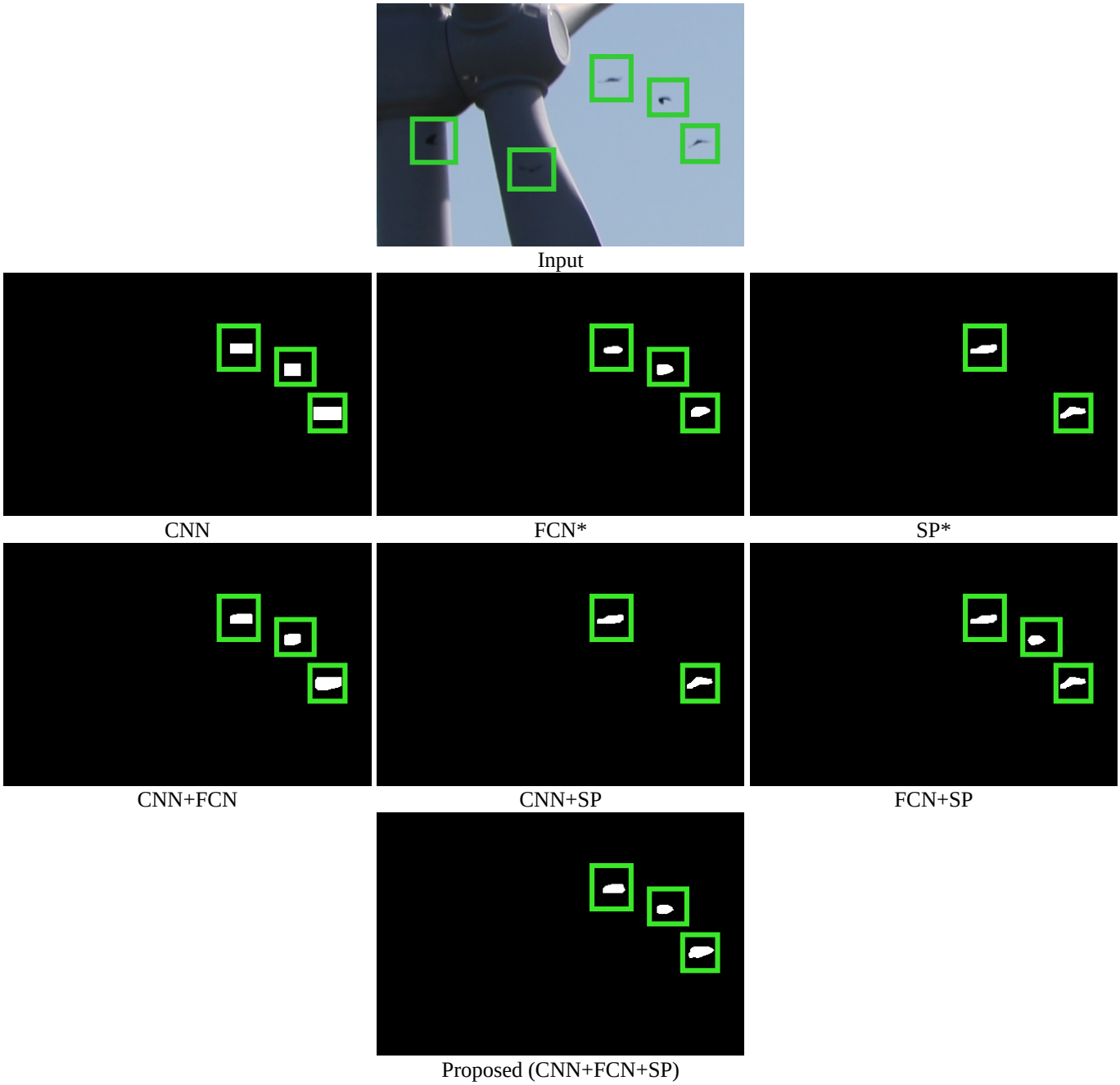


**Fig. 2**. Examples of birds that overlap with background objects. The green squares mean true positives. All of the methods have difficulty in detecting birds in front of the wind turbine, because of the low contrast.

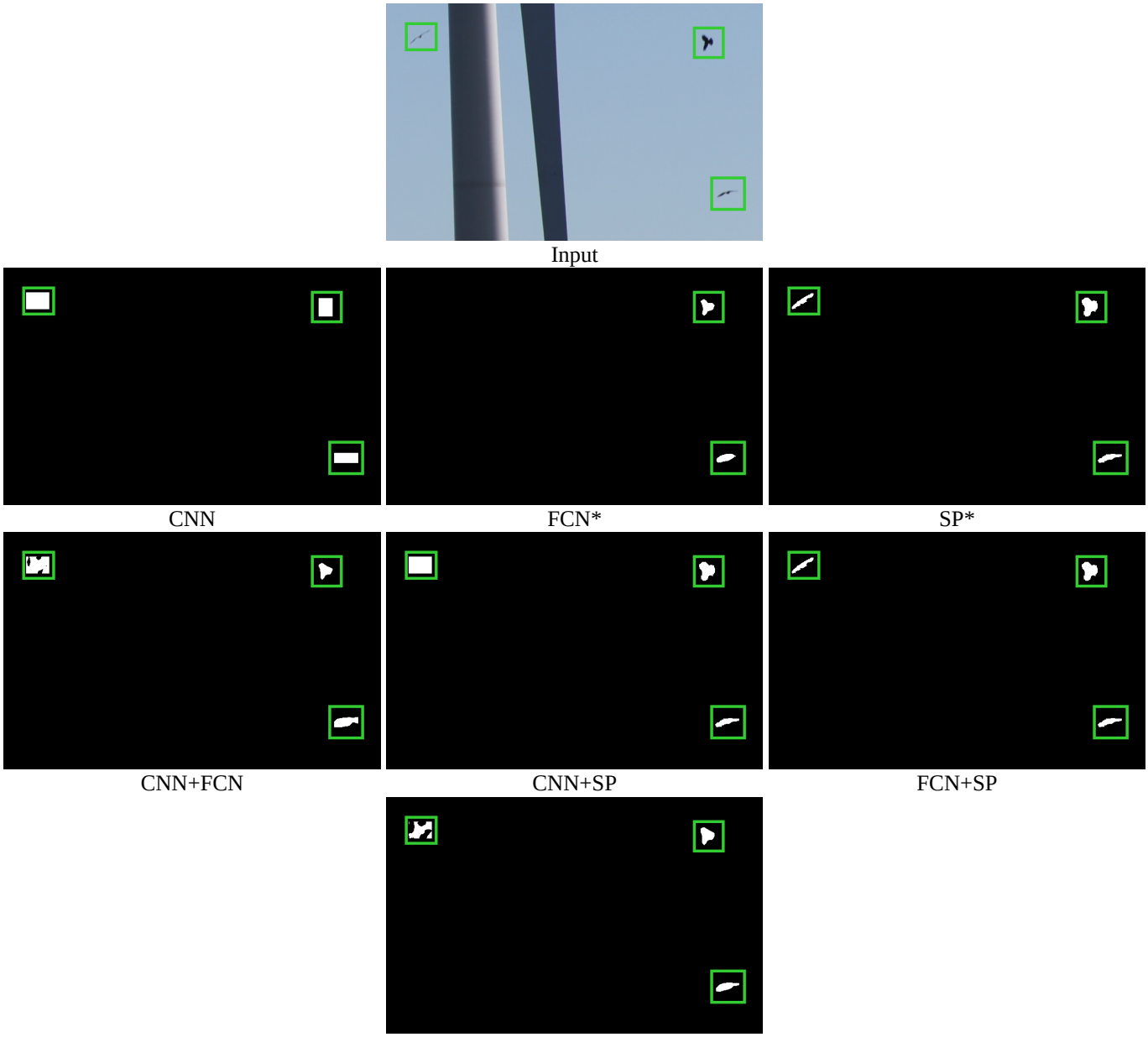


**Fig. 3**. Examples of difficult birds for FCNs. The green squares mean true positives. FCN* does not see the thin bird, while it does when combined with CNNs or SP.

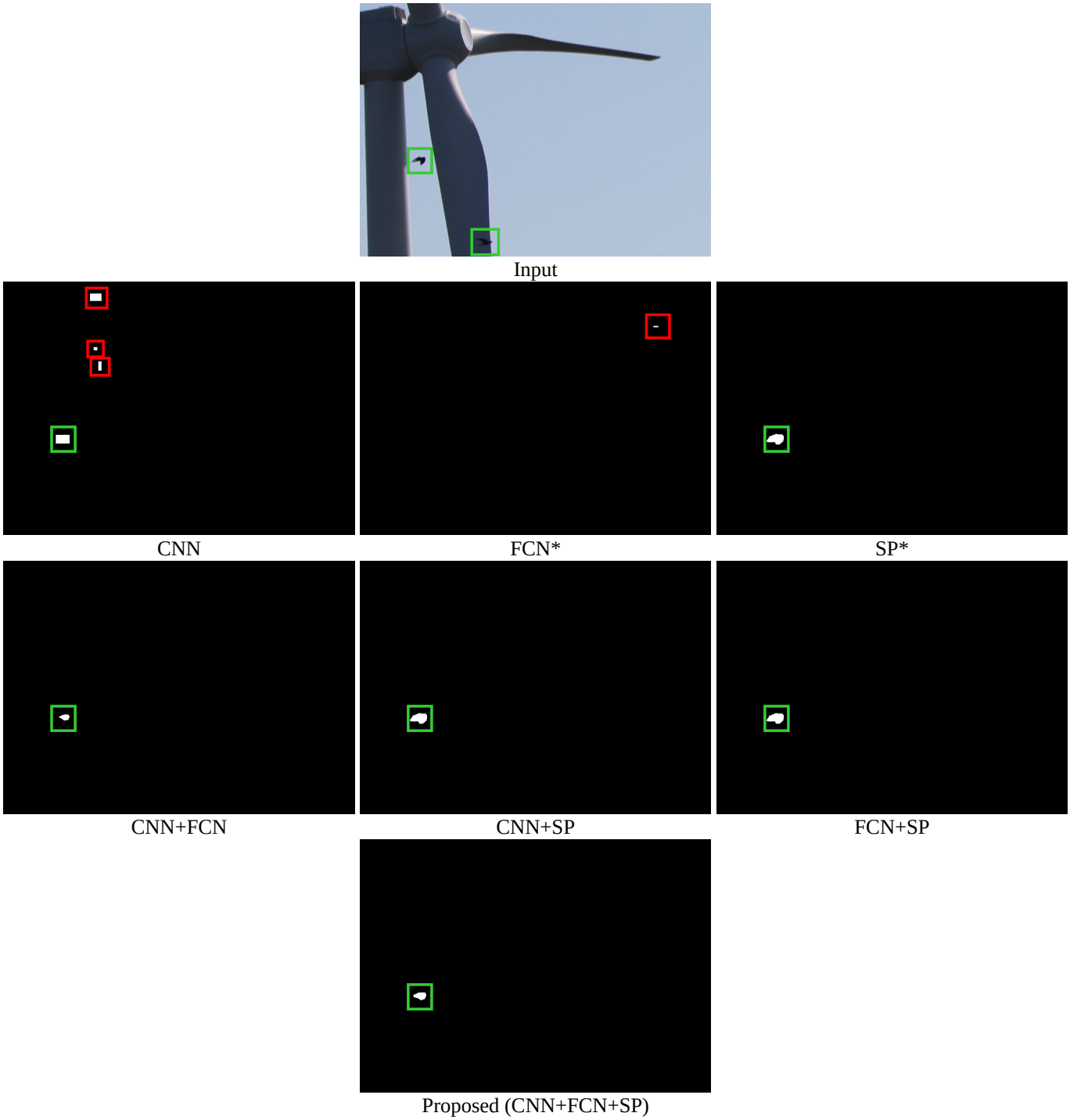


**Fig. 4**. Examples of typically misdetected regions. The red squares mean false positives. CNNs are prone to the regions around the hub. FCNs are prone to the tips of the blade.

| Size | Method | Precision | Recall | F-measure |
|---|---|---|---|---|
| tiny | SP* | - | - | - |
| | FCN* | 0.093 | 0.060 | 0.073 |
| | CNN | 0.541 | **0.896** | 0.674 |
| | CNN+FCN+SP | **0.860** | 0.642 | **0.735** |
| small | SP* | **1.000** | 0.763 | 0.865 |
| | FCN* | 0.932 | 0.863 | 0.896 |
| | CNN | 0.667 | **0.875** | 0.757 |
| | CNN+FCN+SP | **1.000** | 0.863 | **0.926** |
| normal | SP* | 0.970 | 0.889 | 0.928 |
| | FCN* | **1.000** | 0.944 | 0.971 |
| | CNN | 0.583 | **0.972** | 0.729 |
| | CNN+FCN+SP | **1.000** | **0.972** | **0.986** |

**Table 1**. F-measure of individual methods are shown when the input images are categorized by image size. In the *tiny* group, the best recall was achieved by CNNs. The proposed method achieved the second best recall, but achieved the highest F-measure owing to its high precision. In the group of *small*, the highest precision and recall was given by SP* and CNNs, respectively. However, FCN* achieved higher F-measure than SP* or CNNs because FCN* achieved the second highest precision and recall, which shows FCN* can simultaneously detect birds and recognize the background. In *normal* size, FCN* had the highest precision, while CNNs had the best recall. In all cases, FCN* achieved the highest F-measure in individual methods when they are solely used. The proposed method always produces the highest F-measure.