Salient Object Subitizing: Supplementary Material

1. Visualizing the CNN Subitizing Classifiers

We provide some visualization results of our CNN-based subitizing classifiers to provide insight into the model learned by the CNN.

1. Sample prediction results, showing example true positives, false positives and false negatives produced by our Subitizing classifiers (Fig. 1).
2. Sample visualization results of our CNN subitizing classifiers using the method of [4] (Fig. 2).
3. 2D-embedding of the fc7 CNN feature: before and after fine-tuning using our SOS dataset (Fig. 3-4).

References

[1] A. Karpathy. t-SNE visualization of CNN. http://cs.stanford.edu/people/karpathy/cnnembed/.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems (NIPS)*, 2012.

[3] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge, 2014.

[4] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *International Conference on Learning Representations (ICLR), Workshop Track*, 2014.

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

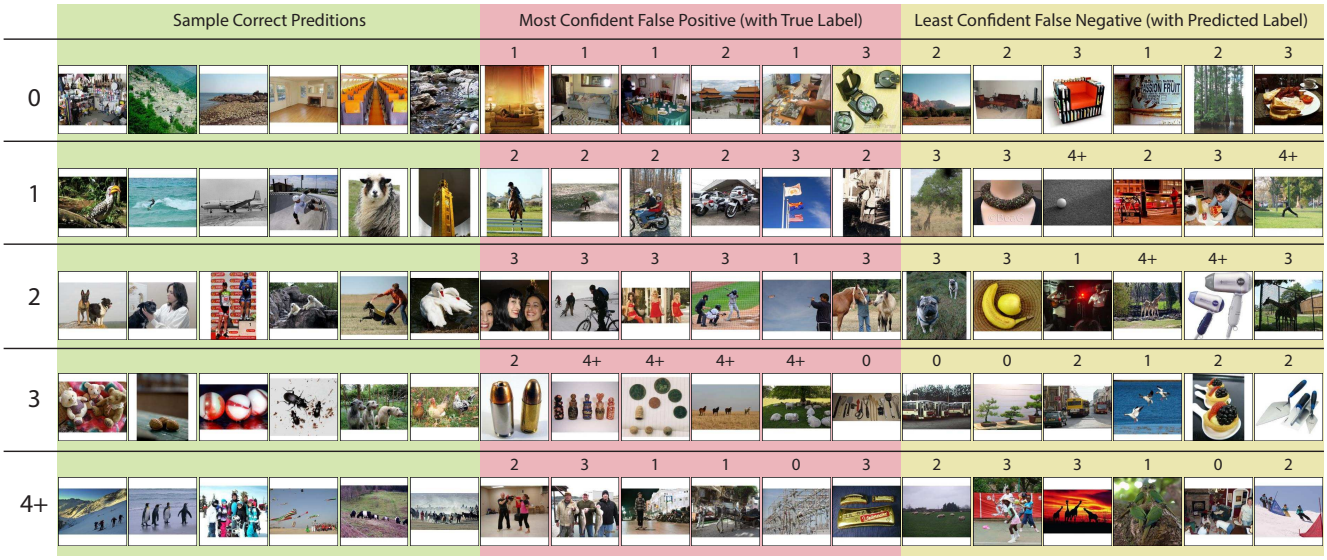

Figure 1: Sample prediction results, showing example true positives, false positives and false negatives by our Subitizing classifiers. Each row corresponds to a class. The green column shows sample true positives of our CNN classifiers for each class. The red column shows the most confident false positives, *i.e.* images incorrectly classified as the target class with highest confidence scores. The ground-truth labels are displayed on top. The yellow column shows the least confident false negatives, *i.e.* images that belong to the target class but receive the lowest confidence scores for this class, and thus are classified as other classes. The predicted labels are displayed on top.

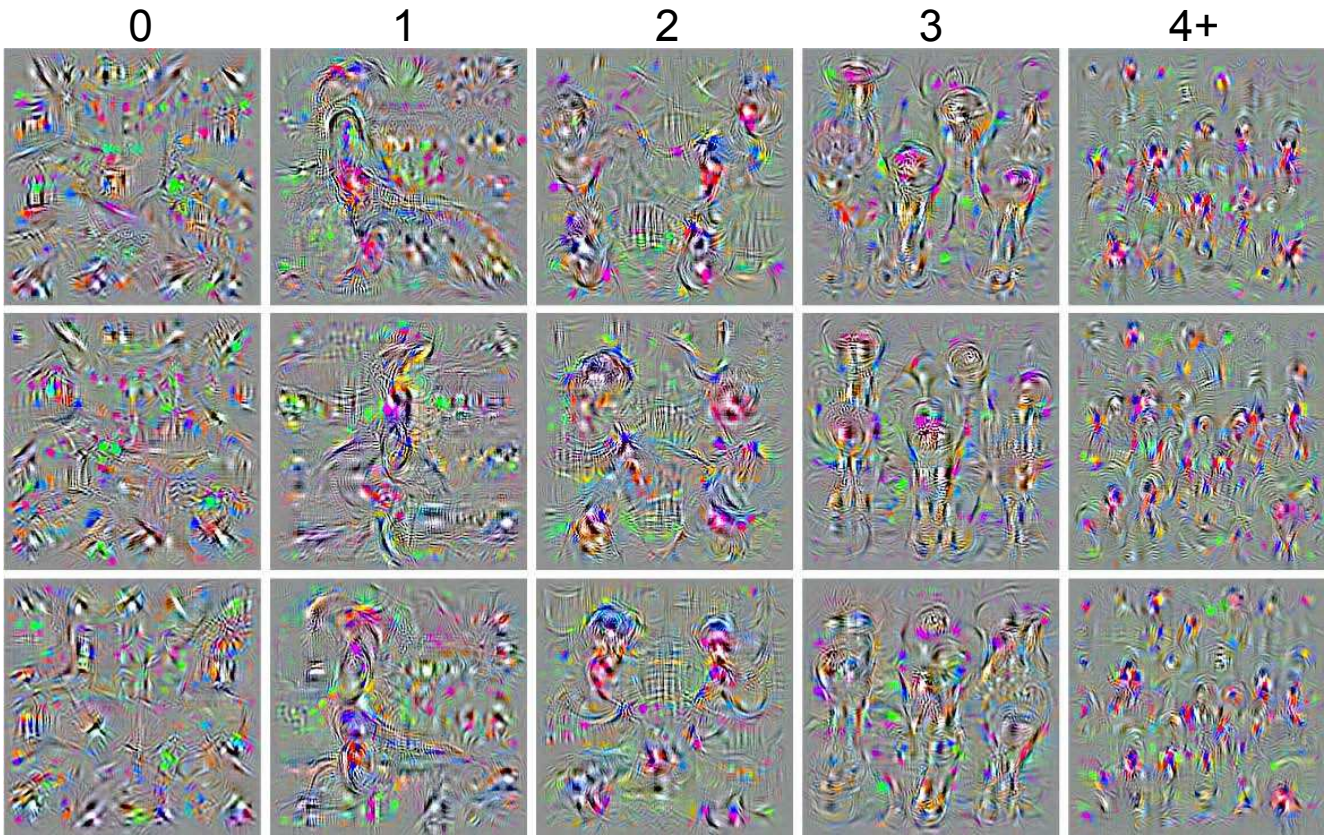

Figure 2: Sample visualization results of our CNN subitizing classifiers using the method of [4]. The method of [4] computes an image that maximizes a specific classification score with L2 regularization. By starting from a random initial image, the method of [4] optimizes the image using back-propagation. In the above figure, each column corresponds to a 1-vs-all classifier for a category, and three random samples are displayed. The visualization results indicate that the CNN captures some common visual patterns for each category, especially for categories 2, 3 and 4+.

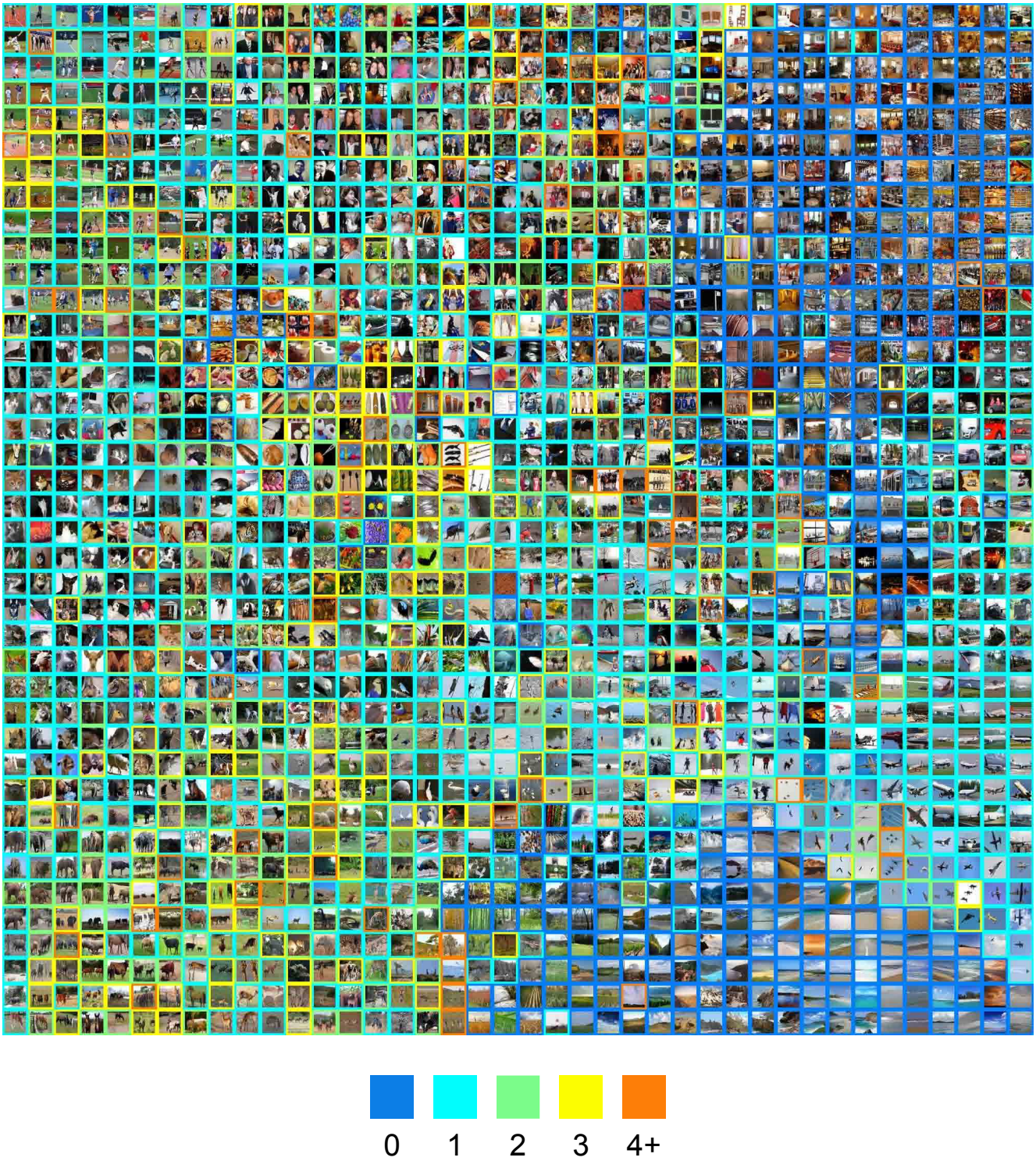

Figure 3: We use the t-SNE visualization code from [1] to visualize the 2D embedding of the fc7 layer in our fine-tuned CNN model for SOS. The code of [1] first computes the 2D embedding of the source feature space and then, for each point on a sample grid in the embedded 2D space, the nearest instance in our SOS dataset is shown. Here we show the visualization result for the CNN model of [2] trained on ImageNet [3]. Ground-truth labels of images are indicated by the color of the borders. Without fine-tuning on our SOS dataset, the background images (class 0) are pretty well separated from the other classes in the fc7 feature space. This explains the good performance of our CNN_wo_FT baseline. However, other classes are still mixed.

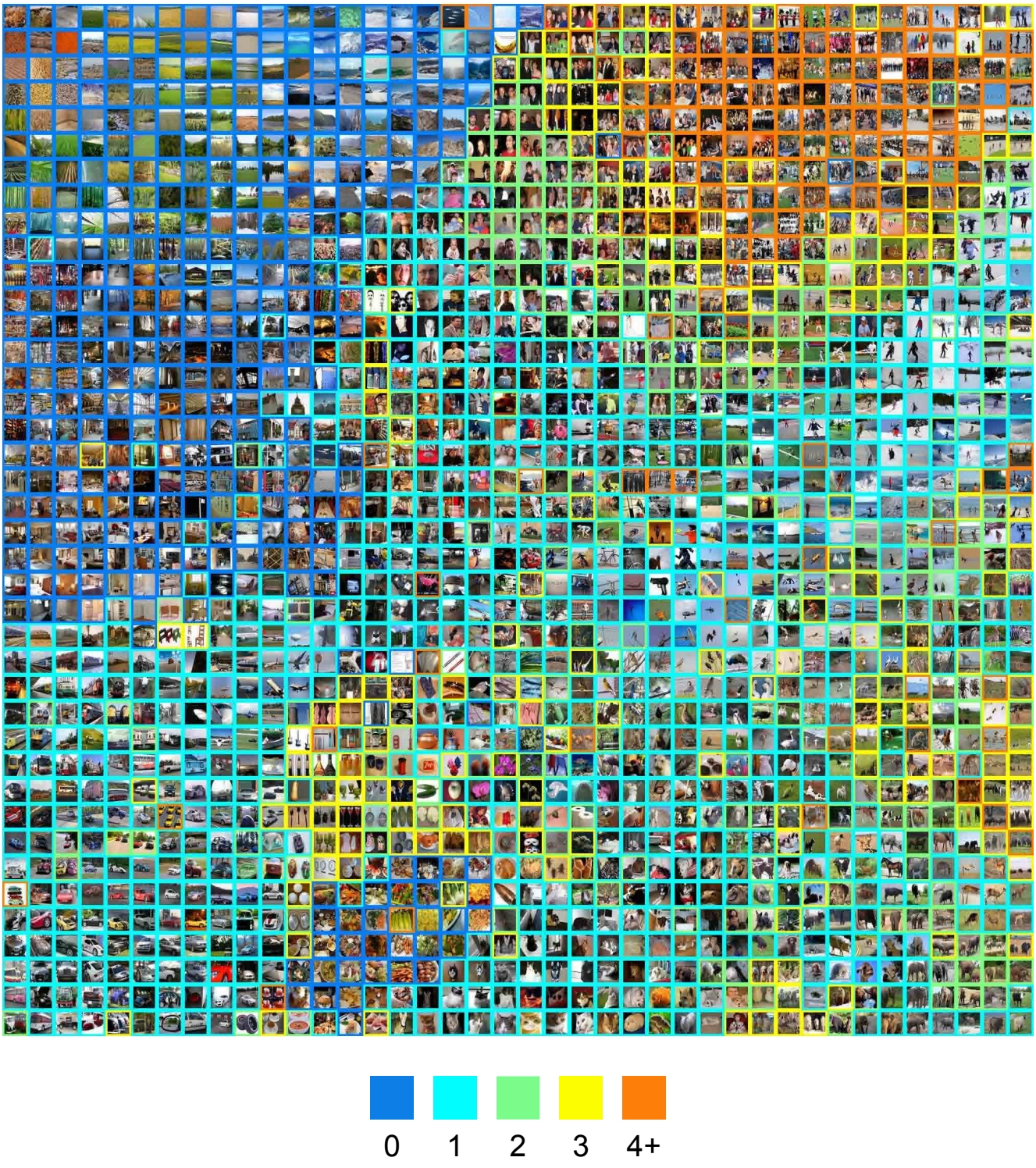

Figure 4: The 2D embedding of our fine-tuned CNN model. The classes are much better clustered in the 2D embedding space. Images with similar visual content and composition tend be close to each other in the fc7 embedding space.