A Appendix

A.1 Discussion on Framework Robustness

As the proposed framework relies on the SCC from pretrained models, a natural question is raised: **can it still be effective when the pretrained model is weak?**

Table A1 explores the performance of the proposed method using different backbones on the Food-101N dataset. Although ResNet18 has around 2% deficiency on the pretrained model comparing to ResNet50, our framework can still help boost the performance. Moreover, we found that although the advantages from our framework might be affected by a weaker model, it stays effective even with a weaker model AlexNet on Food-101N, showing the robustness of framework.

Table A1. Dependence on Pretrained Model Ability

Backbone	S1 Top-1	S2 Top-1 (ours)	S2 Top-1 (ours+GBA)
ResNet18 Vanilla	82.16	83.44	83.72
AlexNet Vanilla	72.55	73.01	73.67

A.2 Algorithm

We provide the algorithm block in Algorithm 1 to help readers understand details of our framework.

Algorithm 1 Proposed Method

Input: Dataset $\mathcal{D}$, a pretrained model $\mathcal{M}_{\theta_0}$ with model parameter θ_0.
Function: Binary cross entropy $\mathcal{H}(p, q)$.
Output: Optimal model parameter θ .

Let $\theta = \theta_0$. # *$\mathcal{M}_\theta$ will be finetuned from $\mathcal{M}_{\theta_0}$*
for $(x, y^*) \in \mathcal{D}$ **do**
  # *prepare for confidence c:*
    Predictions of $\mathcal{M}_{\theta_0}$ on x : $p(y|x, \theta_0) = \mathcal{M}_{\theta_0}(x)$.
    Obtain confidence $c = p(y_{(\omega)}|x, \theta_0)$.
  # *prepare for loss functions:*
    Obtain augmented $x' \leftarrow Augment(x)$.
    Predictions of $\mathcal{M}_\theta$ on x': $p(y|x', \theta) = \mathcal{M}_\theta(x')$.
    Webly supervised loss $\mathcal{L}_w = \mathcal{H}(y^*, p(y|x', \theta))$.
    Self-label supervised loss $\mathcal{L}_s = \mathcal{H}(p(y|x, \theta_0), p(y|x', \theta))$.
  # *compute final loss for training:*
    $\mathcal{L} = c \times \mathcal{L}_w + (1 - c) \times \mathcal{L}_s$.
    $\theta \leftarrow \arg\min_\theta \mathcal{L}$
end for
return θ