A Machine Teaching Framework for Scalable Recognition Supplementary Materials

Pei Wang
UC, San Diego
pew062@ucsd.edu

Nuno Vasconcelos
UC, San Diego
nuno@ucsd.edu

1. Experimental settings

Because of the page limitation, some details are not included in the paper. Here we supply them.

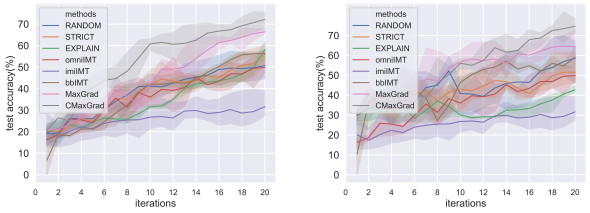
Data preprocessing On all three datasets used in the paper, the images are subject to standard normalizations. Training images are first randomly resized and cropped to 224×224 and then randomly flipped with a probability of 0.5. For the testing images, they are first resized to 256×256 and then center-cropped to 224×224. All images are also first converted to $[0.0, 0.1]$ from $[0, 255]$, and then normalized by subtracting the mean $[0.485, 0.456, 0.406]$ and dividing by the standard deviation $[0.229, 0.224, 0.225]$ of each RGB color channel.

Network training The pre-trained ResNet-18 used is provided by PyTorch[1]. For all experiments, the student networks are trained 20 epochs by gradient descent with batch size of $|\mathcal{L}^t \cup \mathcal{C}^t|$ and weight decay of $1e-4$. The learning rate is set to $1e-4$ with 0.9 momentum.

Detailed set-ups for training MTurkers On the evaluation of student teaching, each experiment was performed with 40 students, on Amazon MTurk. The teaching process had two phases, teaching and testing. Before teaching, workers were shown a brief introduction of the teaching set-up, illustrating how our web-based teaching interface works. During teaching, they were shown a sequence of 20 images. At each iteration, they were asked to select a category label from a list of options (five for butterflies and three for Characters). Once they made a choice, they received feedback. For correct labels, they were told "Your choice is correct." Otherwise, true label and counterfactual explanations were presented as in Figure 4 of the paper. Upon receiving feedback, workers waited for a minimum of 2 seconds before they could proceed to the next teaching image. After teaching, 20 randomly selected testing set images were assigned to each learner. These random images were different for each learner and no feedback was given.

On the evaluation of scalable recognition, the similar process is conducted but 30 unlabeled images are assigned after the teaching phase.

[1] https://pytorch.org/vision/stable/models.html



(a) butterflies (b) Chinese Characters

Figure 1: Test set labeling accuracy of simulated students.

2. Results for simulated student

While the ultimate goal is to teach humans, experiments with simulated students are important to enable replicable method comparisons. In these experiments the student is a network and the goal to select teaching examples that optimize training speed. Figure 1 shows the evolution of student test set labeling accuracy vs. teaching iteration, for the various machine teaching algorithms of Figure 2 of the paper. It can be seen that CMaxGrad significantly improves on MaxGrad [1], the prior state of the art, especially on Butterflies. This suggests the importance of counterfactual explanations.

3. Ablation study

To evaluate the effect of counterfactual explanations [2] on the teaching algorithm and CMaxGrad, we did the ablation study by simply associating the counterfactual explanations to the teaching images selected by RANDOM and MaxGrad. The results are reported in Table 1 that shows that adding counterfactual explanations (CE) to MaxGrad is better than MaxGrad but weaker than CMaxGrad. This indicators the importance of jointly selecting teaching images and their explanations because of the correlations.

4. Selected samples

In Figure 2 and 3, we show the selected teaching sets of MaxGrad and CMaxGrad on two datasets for student teach-

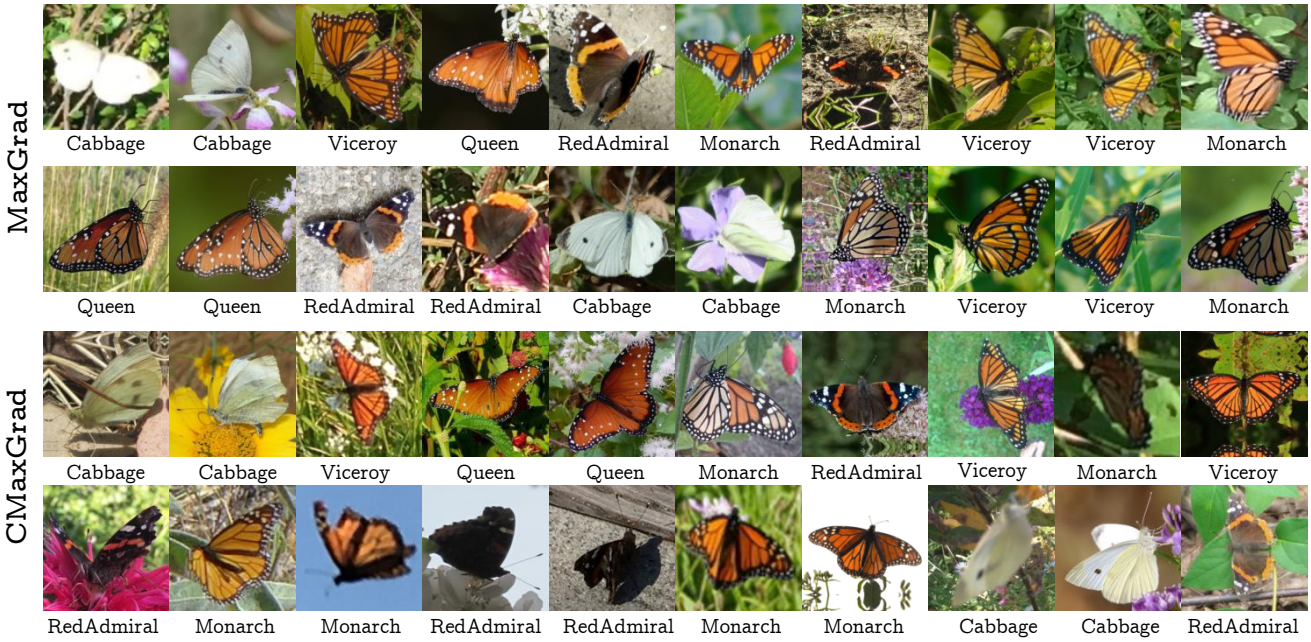


Figure 2: Comparisons of selected teaching examples with raster scan order on Butterflies.

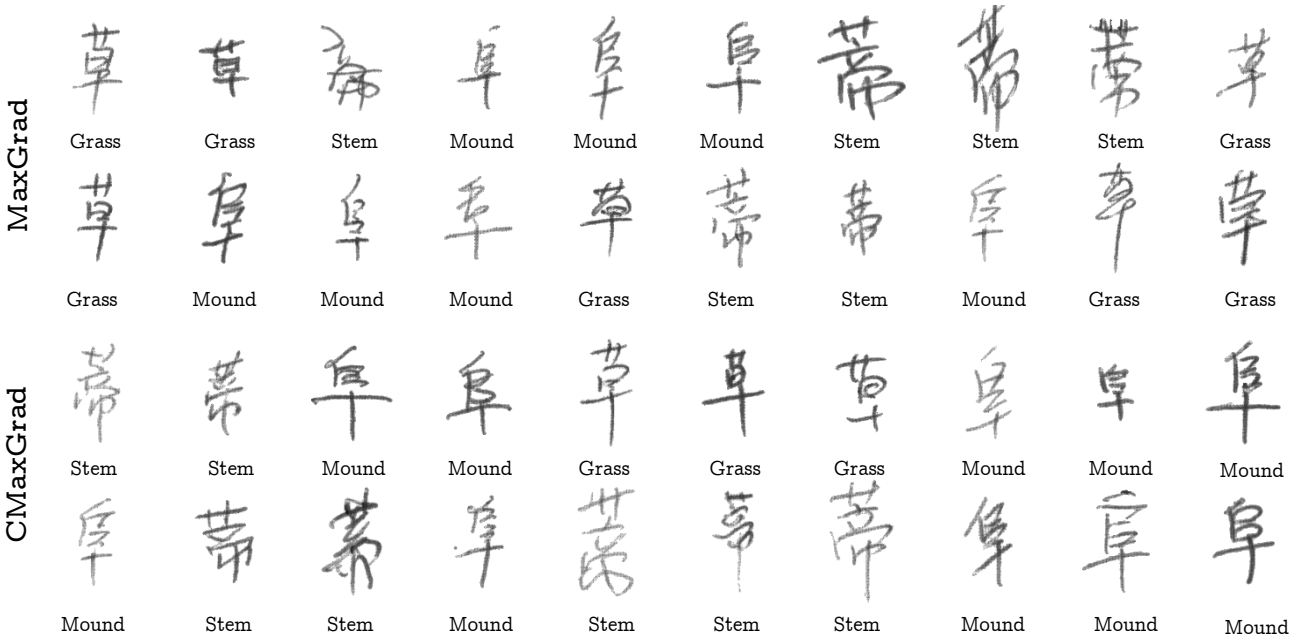


Figure 3: Comparisons of selected teaching examples with raster scan order on Chinese Charaters.

	Butterflies	Chinese Char.
RANDOM	65.2	47.05
RANDOM+CE	74.60(18.58)	71.13(20.90)
MaxGrad+CE	83.00(24.20)	83.75(19.89)
CMaxGrad	84.10(18.24)	84.63(20.18)

Table 1: Test set Labeling accuracy, mean (std).

ing evaluation. Also, in Figure 4 and 5, 5 (3) randomly selected teaching images and their potential counterfactual explanations are shown.

Next, in Figure 6 and 7, we show the selected examples when evaluating the whole framework by using "CMaxGrad+SimCLR" of Table 2 of the paper on two datasets, and some randomly selected counterfactual explanations in Figure 8 and 9.

At this time, we can not give a sufficient intuition about the teaching set, but we observe that , beyond some "normal" images that teach students the appearance of the

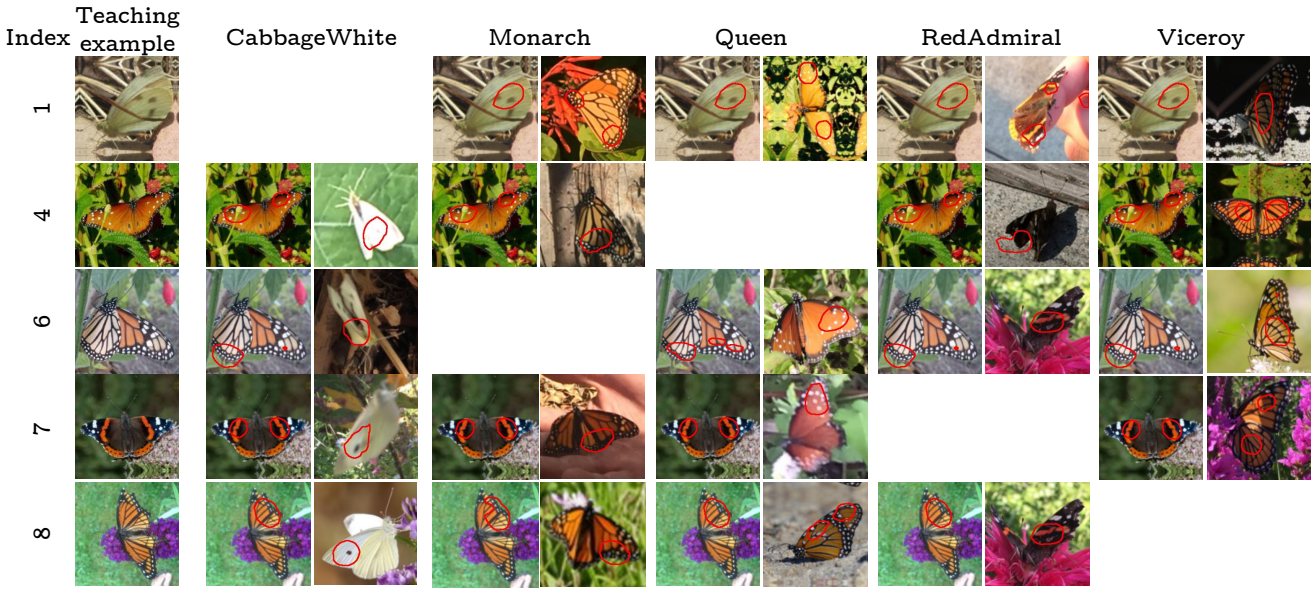


Figure 4: Counterfactual explanations on Butterflies of 5 randomly selected teaching samples on Butterflies. The column index indicates the potential counterfactual classes of the turker.

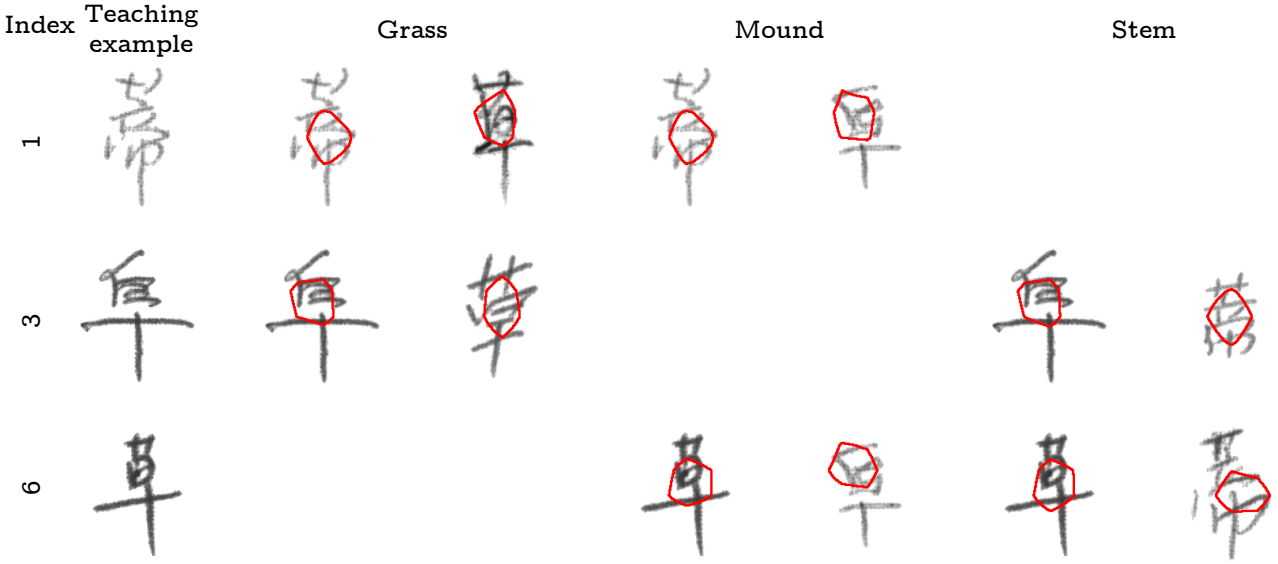


Figure 5: Counterfactual explanations on Butterflies of 3 randomly selected teaching samples on Chinese Characters. The column index indicates the potential counterfactual classes of the turker.

concepts, the algorithm tends to select images with unusual poses, low-resolution, occlusion, and even camouflage. This is consistent with MaxGrad and CMaxGrad's example selection based on large negative margins. These images force the students to focus attention on features that are essential for class discrimination, speeding up the learning process.

References

[1] Pei Wang, Kabir Nagrecha, and Nuno Vasconcelos. Gradient-based algorithms for machine teaching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1387–1396, June 2021. 1

[2] Pei Wang and Nuno Vasconcelos. Scout: Self-aware discriminant counterfactual explanations. *CVPR*, 2020. 1

Figure 6: Selected teaching examples with raster scan order on Butterflies.

Figure 7: Selected teaching examples with raster scan order on Gull.

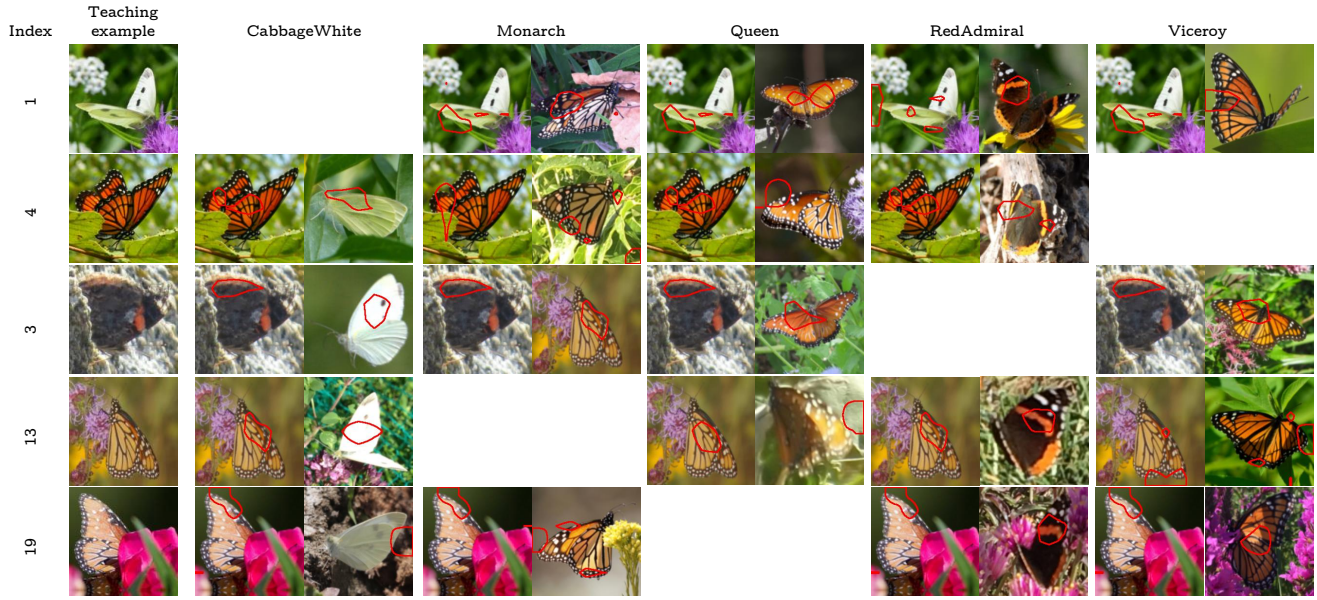


Figure 8: Counterfactual explanations on Butterflies of 5 randomly selected teaching samples on Butterflies. The column index indicates the potential counterfactual classes of the turker.

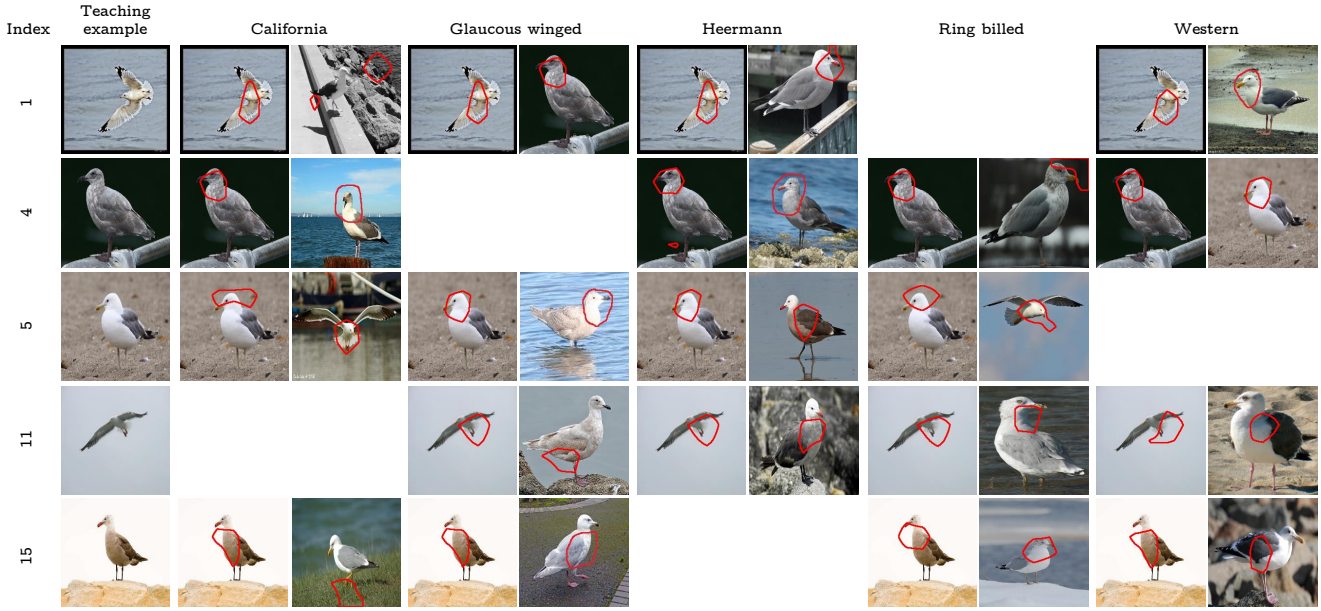


Figure 9: Counterfactual explanations on Gull of 5 randomly selected teaching samples on Gull. The column index indicates the potential counterfactual classes of the turker.