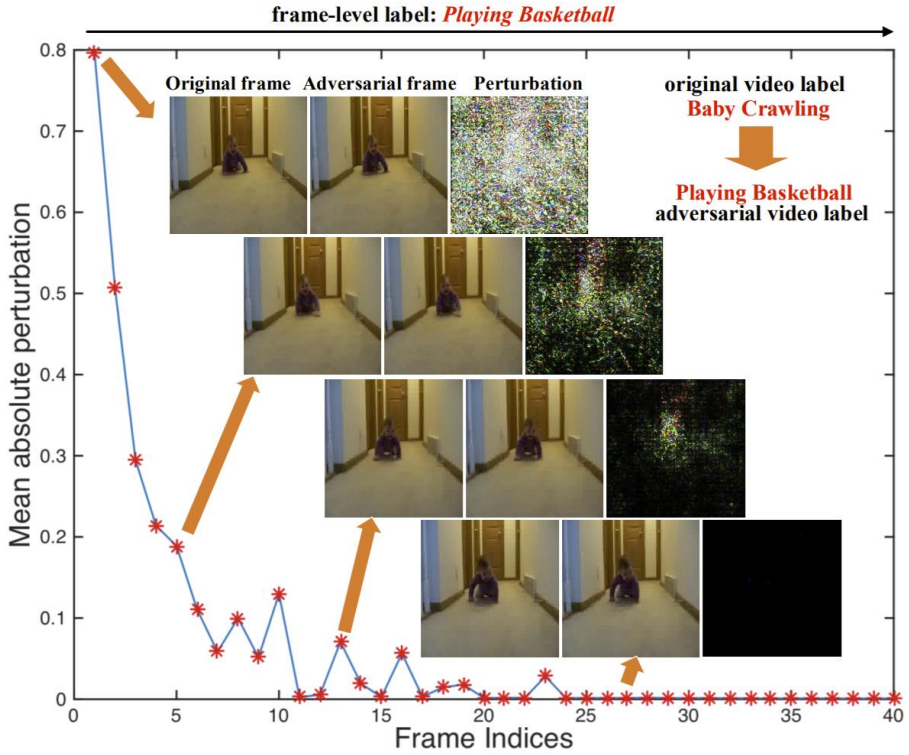
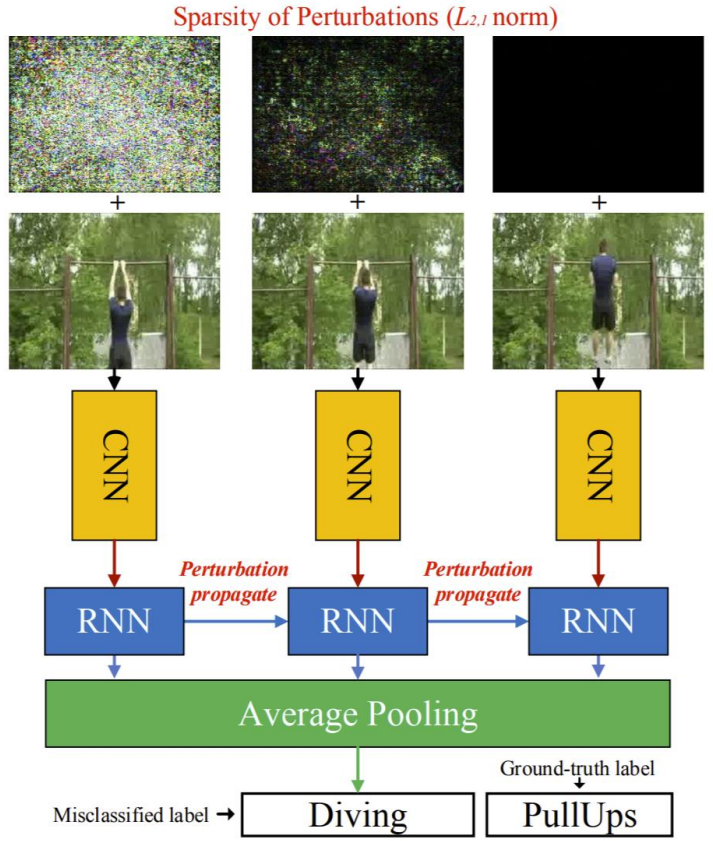
Sparse Adversarial Perturbations for Videos

Introduction





Methodology

$$\arg\min_{\mathbf{E}} \lambda||\mathbf{E}||_{2,1} - \ell(\mathbf{1}_y, J_\theta(\hat{\mathbf{X}}))$$

why $l_{2,1}$ norm?

$$\arg\min_{\mathbf{E}} \lambda||\mathbf{E}||_{2,1} - \frac{1}{N}\sum_{i=1}^{N} \ell(\mathbf{1}_{y_i}, J_\theta(\hat{\mathbf{X}}_i))$$

cross videos

$$\arg\min_{\mathbf{E}} \lambda||\mathbf{M}\cdot\mathbf{E}||_{2,1} - \frac{1}{N}\sum_{i=1}^{N} \ell(\mathbf{1}_{y_i}, J_\theta(\mathbf{X}_i + \mathbf{M}\cdot\mathbf{E}))$$

propagation across frames

Experiment

Table 1: The results of fooling rates versus different sparsities.

S	0%(40)	80%(8)	90%(4)	97.5%(1)
F	100%	100%	91.8%	59.7%
P	0.0698	0.6145	1.0504	1.9319

Table 2: Time for computing perturbations in one iteration.

S	0%	50%	75%	87.5%	97.5%
Time	2.853s	1.367s	0.612s	0.346s	0.0947s

Experiment

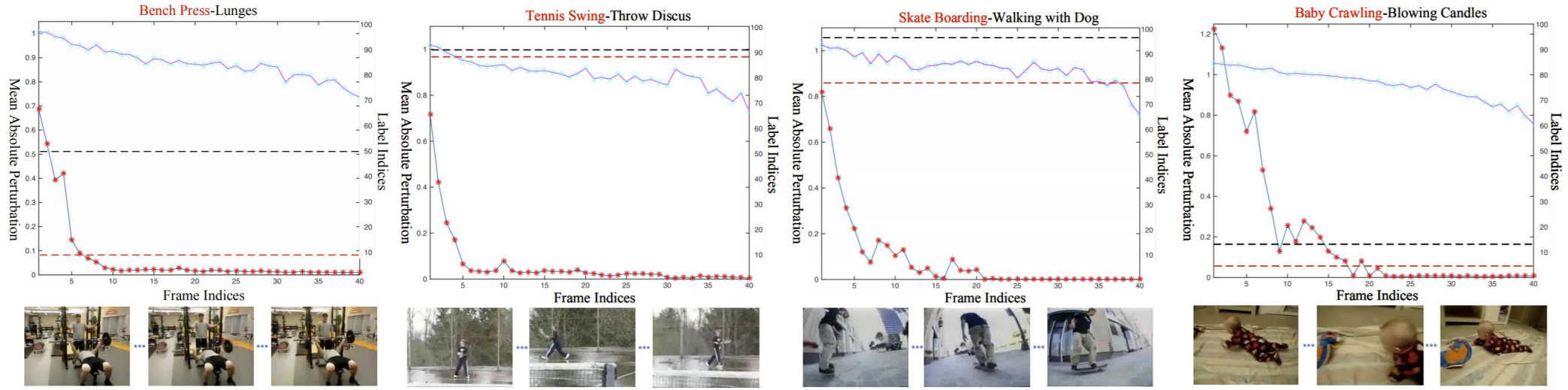


Figure 4: Four examples for showing perturbation propagation on UCF101 dataset. The x-axis denotes the frame indices in a video. The left y-axis denotes the Mean Absolute Perturbation (MAP) value of each frame's perturbations, and the right y-axis is the label indices. The blue line with stars is the curve of MAP values with $l_{2,1}$ norm, and magenta line with circles is the result with l_2 norm. The red dotted line is the predicted frame-level label indices for the clean video, and black dotted line is the predicted frame-level label indices for the adversarial video, both by the action recognition networks (the video-level labels are listed in the top of each figure with the same color). In the bottom of each figure, we give the corresponding video frames. For detailed discussions, please see the texts.

Experiment

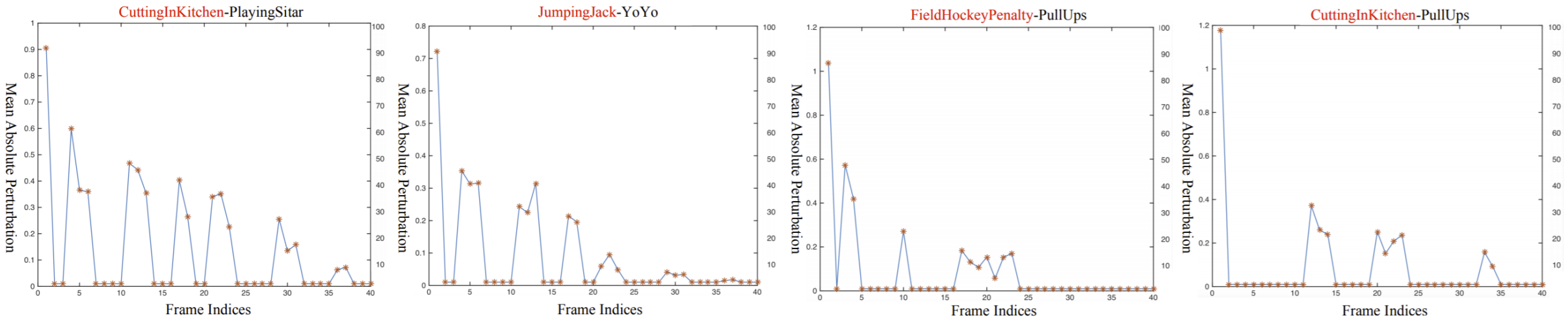


Figure 6: Perturbations can be added on any discrete frames with a random margin. These are four adversarial examples under different sparsities conducted on UCF101 dataset. For detailed discussions, please see the texts.

Experiment

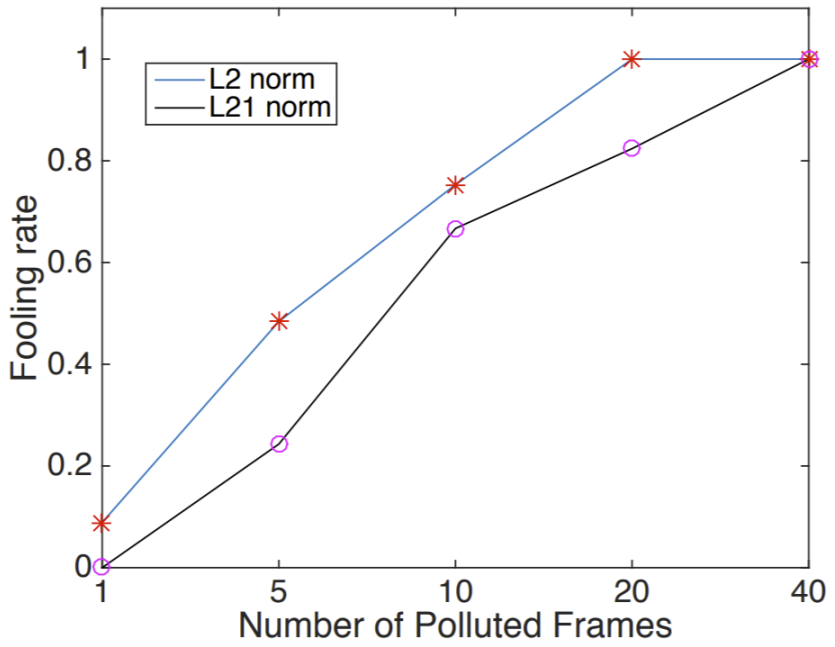


Figure 8: Comparisons between l_2 and $l_{2,1}$ norm versus Fooling rate on UCF101 dataset. We here report the results when $N = 1, 5, 10, 20, 40$, respectively. The total number of frames is 40.

Conclusion

Advantages:

Reduced the time cost
Low Mean Absolute Perturbation(MAP)
Provided an idea that use $L_{2,1}\ norm$ to achieve sparsity

Disadvantages:

Specific architecture
High perturbation rate for specific frames
Randomly select frames(may not be the best)