Machine Learning for Human Learning & Teaching

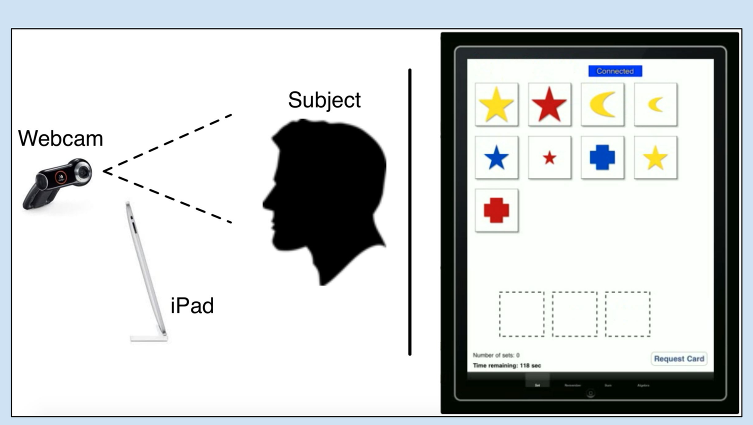
P.I.: Jacob Whitehill (jrwhitehill@wpi.edu) (WPI). **Collaborators & students**: Arkar Min Aung (WPI), Aysha Foster (VCU), Zeqian Li (WPI), Yi-Ching Lin (VCU), Jennifer LoCasale-Crouch (UVA), Javier Movellan (UCSD), Erin Ottmar (WPI), Anand Ramakrishnan (WPI), Margo Seltzer (Harvard), and Zewelanji Serpell (VCU)

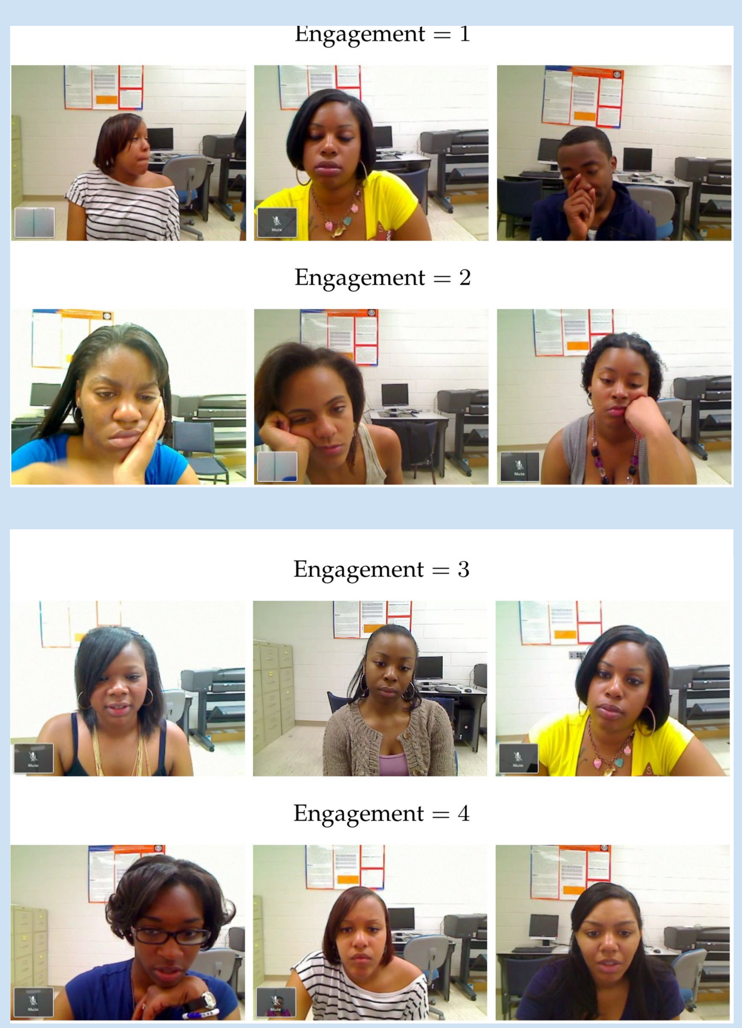
Abstract

- The goals of our research are to:
(1) Automatically perceive students' learning and emotions with accuracy comparably to that of good human teachers.
(2) Model, predict, and optimize human learning using machine learning and optimal control theory.
(3) Make the large quantities of human knowledge already existing on the Internet more easily accessible to learners.

Automatic Student Engagement Recognition from Video

- Since the 1980s, student engagement has been recognized by teachers and researchers as both an indicator of effective teaching and an important outcome itself.
- In a collaboration between UCSD, VCU, and WPI, we developed a video-based automatic engagement detector for use with iPad-based intelligent tutoring systems.
- Accuracy of the system for binary judgments is comparable to that of human coders.
- Posttest scores were predicted just as accurately from engagement judgments as with pretest scores.
- Soft label distributions resulted in better prediction accuracy than hard labels.
- See Whitehill, Serpell, Foster, Lin, & Movellan (2014), and Aung & Whitehill (2018), for more details.

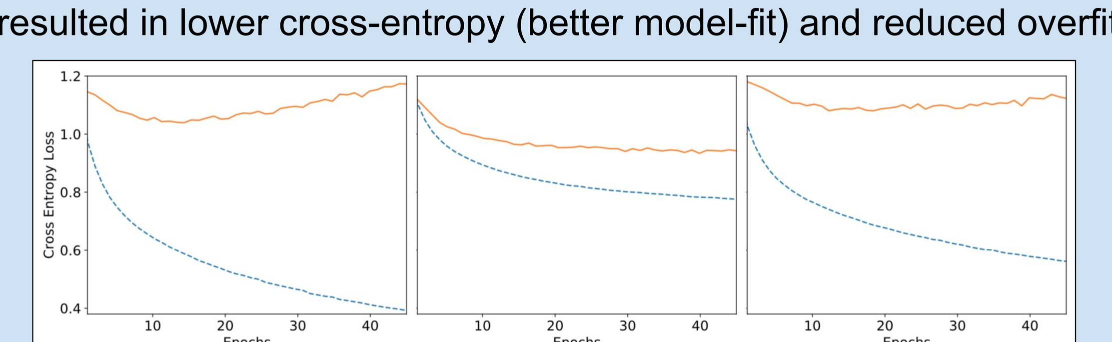




Correlations of Engagement with Test Scores

	Pre-test	Post-test
Human labelers		
Mean engagement label	0.52*	0.37
$P(\text{Engagement} = 1)$	−0.39	−0.22
$P(\text{Engagement} = 2)$	−0.32	−0.18
$P(\text{Engagement} = 3)$	−0.34	−0.40
$P(\text{Engagement} = 4)$	0.57*	0.47*
Automatic classifier		
$P(\text{Engagement} = 4)$	0.64*	0.27

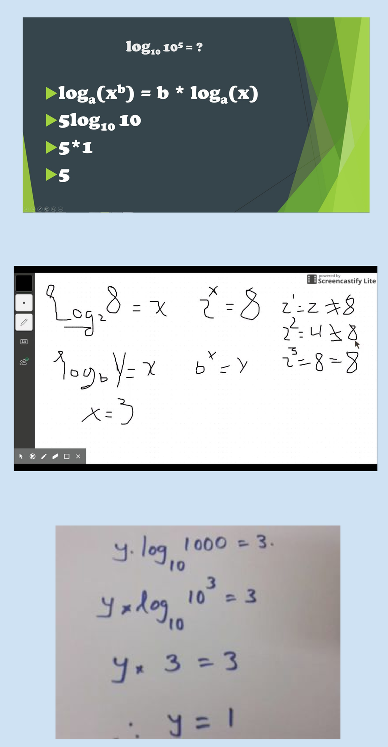
Training on soft labels (a probabilistic ground-truth label for each image) resulted in lower cross-entropy (better model-fit) and reduced overfitting:

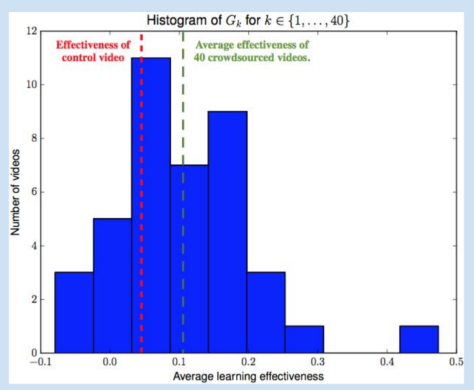


(a) Hard (b) Soft (c) Smoothed

Crowdsourcing Tutorial Videos

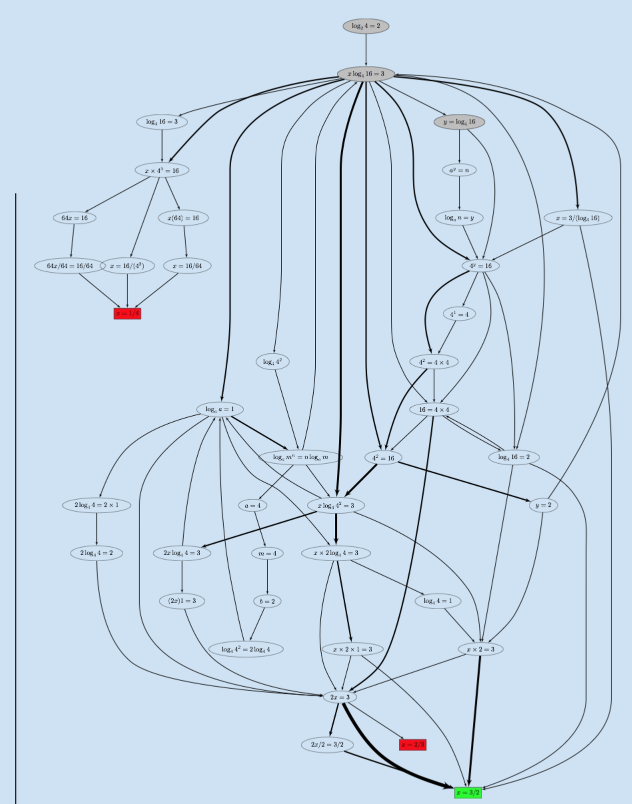
- In this project we explore whether ordinary people on the Web can be incentivized to author and contribute novel math tutorial videos.
- In a 4-week study on Amazon Mechanical Turk, we collected 399 videos on 18 different math problems about logarithms.
- The videos varied in terms of pedagogical approach & learning style.
- The best crowdsourced videos were just as effective at helping students learn (pretest -> posttest) as a popular Khan Academy video.
- See Whitehill & Seltzer (2017) for more details.





Video	Participants	G_k
1	58	0.1416
2	42	0.1140
3	57	0.0942
4	35	0.0932
Khan	58	0.1506

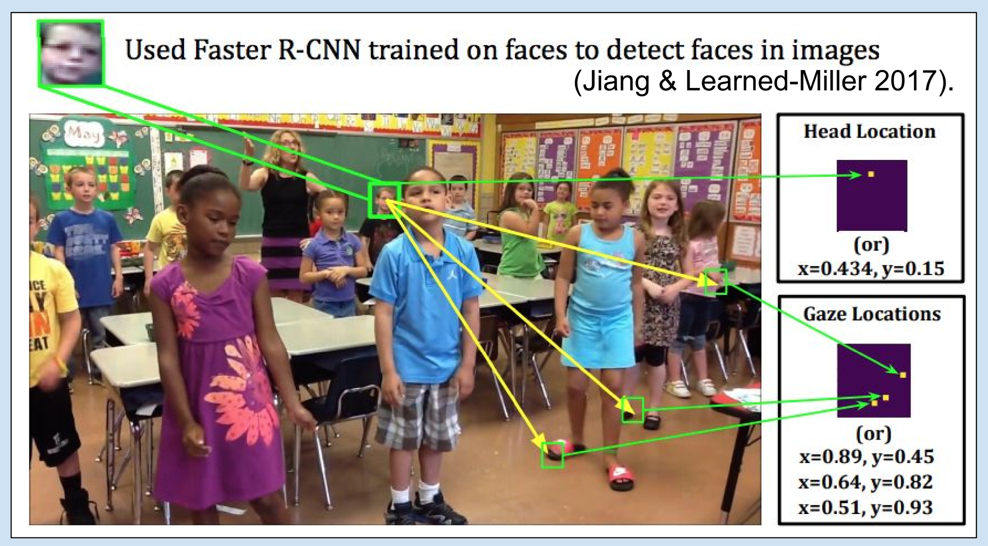
Table 1. Average learning gains G_k as measured in Experiment 3, for the 4 videos were estimated to be highest in Experiment 2, compared to the average learning gains of a popular Khan Academy video on logarithms.

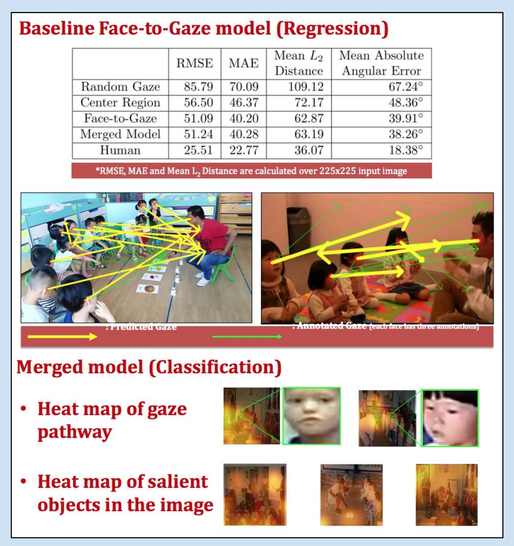


Graph of different solution pathways of the same math problem.

Towards Automatic Classroom Observation Analysis

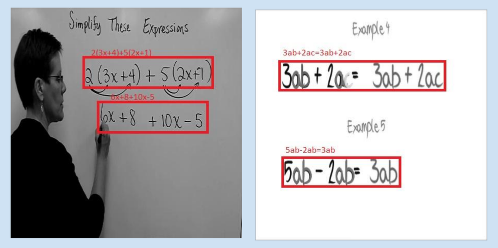
- Classroom observation is used in schools throughout the USA to provide teachers with feedback for professional development.
- Manual classroom observation is very expensive and laborious.
- The goal of this project (collaboration between WPI and UVA) is to provide an automated assessment of teacher-student interactions in school classrooms using multi-modal machine learning.
- We are currently working on automatic 2-D eye-gaze following of teachers & students to assess whether teachers are attending to those students who need may need help; the approach is based on computer vision architecture by Recasens, et al. (2015).



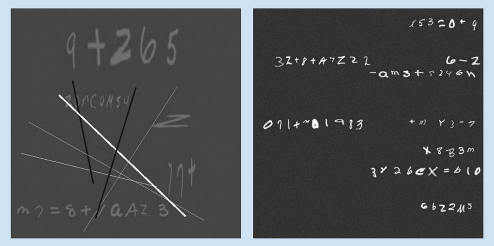


Large-Scale Visual Search through YouTube Math Tutorials

- YouTube contains over 1M tutorial videos on algebra alone.
- Many thousands of these videos contain high-quality educational content that could help learners to understand how to solve specific math problems.
- Harnessing the diversity of YouTube tutorials may help to enable personalized learning.
- In this work, we are using algorithms for simultaneous object detection & recognition to create a large-scale search index over video-based math content.

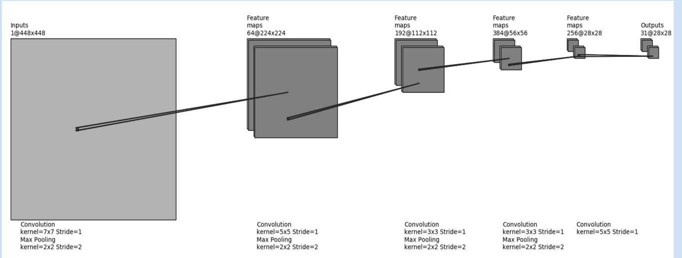


Expected annotation of frames from math tutoring videos. We hope our model can detect math expressions and annotate them correctly.

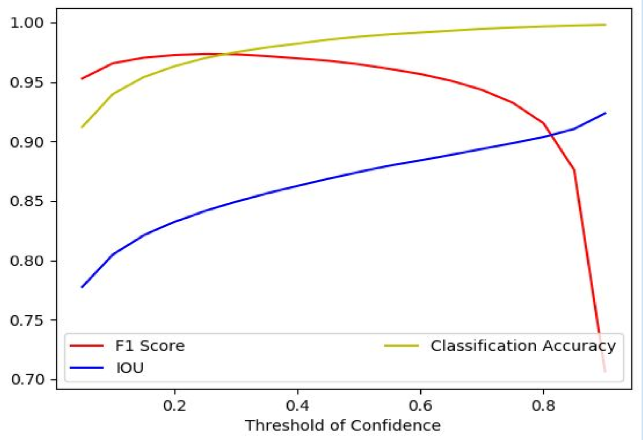


We can generate infinite training/testing data by placing handwritten digits and symbols on an image.

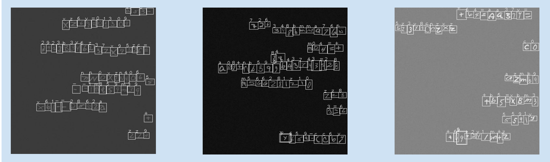
$$\underbrace{\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\left[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2\right]+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\left[\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2\right]}_{\text{Bounding Boxes}}+\underbrace{\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}(C_i-\hat{C}_i)^2+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{noobj}(C_i-\hat{C}_i)^2}_{\text{Confidence}}+\underbrace{\sum_{i=0}^{S^2}\mathbb{1}_{ij}^{obj}\sum_{c\in classes}(p_i(c)-\hat{p}_i(c))^2}_{\text{Classification}}$$



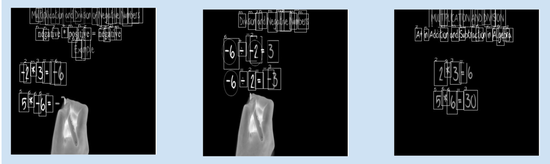
We use a recognition architecture similar to the YOLO (Redmon, et al. 2016).



Curves plotted on average F1 score, IOU and classification accuracy for different threshold of confidence. We use 1,000 images as validation set to test different threshold for confidence.



Test results on synthetic images. The threshold of confidence is set to 0.4.



Test results on frames from real tutoring videos. All frames are resized to 448*448 and reversed in grayscale.

Acknowledgements

- This material is based upon work supported by the National Science Foundation under grants 1551594, 0542013, and 1041755.
- This work was partially supported by a grant from the Harvard Initiative for Learning & Teaching.
- This work was partially supported by a Microsoft Azure for Cloud Computing grant.

References

1. Whitehill, J., & Seltzer, M. (2017, April). A Crowdsourcing Approach to Collecting Tutorial Videos--Toward Personalized Learning-at-Scale. In Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale (pp. 157-160). ACM.
2. Recasens, A., Khosla, A., Vondrick, C., & Torralba, A. (2015). Where are they looking?. In Advances in Neural Information Processing Systems (pp. 199-207).
3. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
4. Jiang, H., & Learned-Miller, E. (2017, May). Face detection with the faster R-CNN. In Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on (pp. 650-657). IEEE.
5. Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. IEEE Transactions on Affective Computing, 5(1), 86-98.
6. Aung, A., and Whitehill, J. "Harnessing Label Uncertainty to Improve Modeling: An Application to Student Engagement Recognition". IEEE Face & Gesture Recognition (FG) (accepted), 2018.