Crowdsource Study and Predictor of Compressed Picture Quality

Spring 2020

Arkan Abuyazid, Shehryar Ahmed, Brandon Lee,

Akash Shukla, Meenakshi Swaminathan, Kan Vanthanasuksan

Faculty Sponsor: Dr. Alan C. Bovik

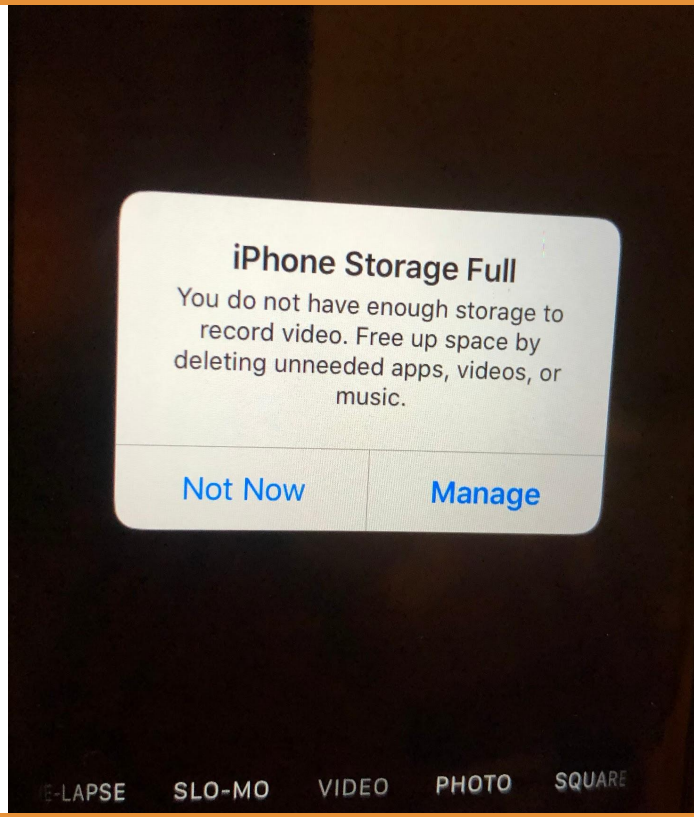
Special Thanks: Jerry Yu

The University of Texas at Austin
Electrical and Computer Engineering
Cockrell School of Engineering

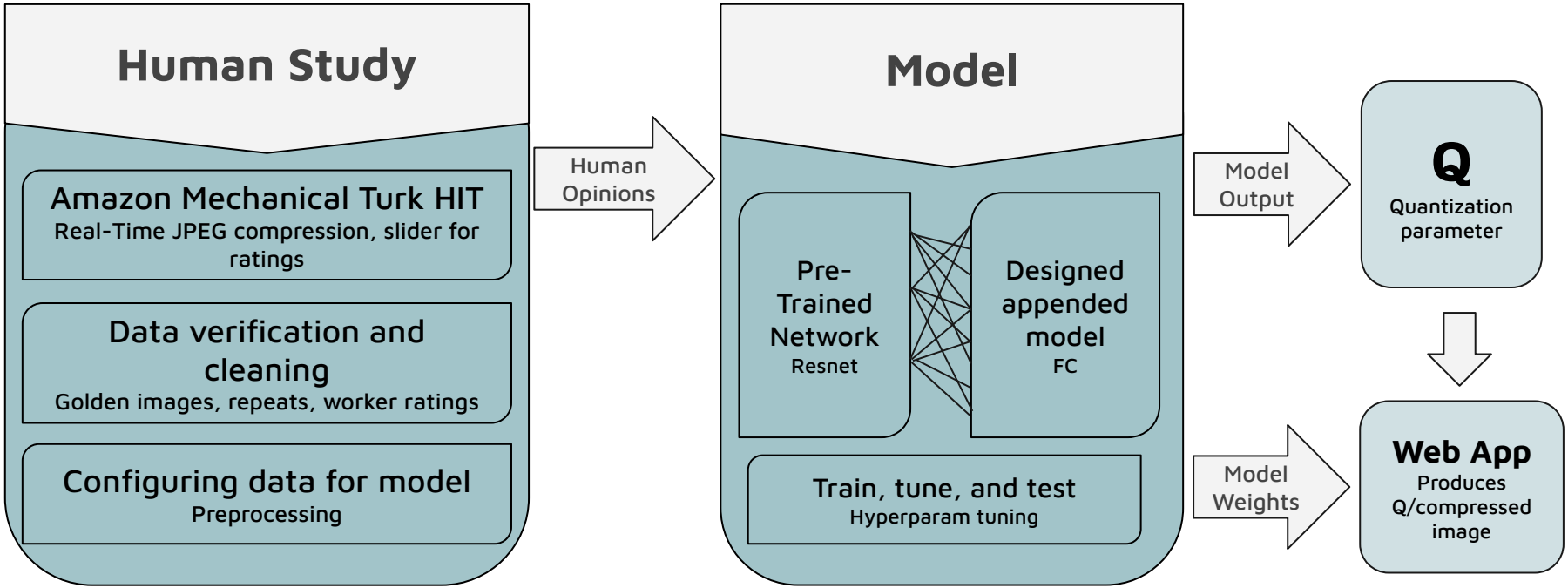
Table of Contents

Design Problem

- Millions of images are uploaded each day, using up a lot of bandwidth (petabytes)
- Pseudo-perceptual metrics, like SSIM, do not accurately represent human-perception
- Our goal: design a system that automatically determines the optimal JPEG quantization parameter for an image based on human perception

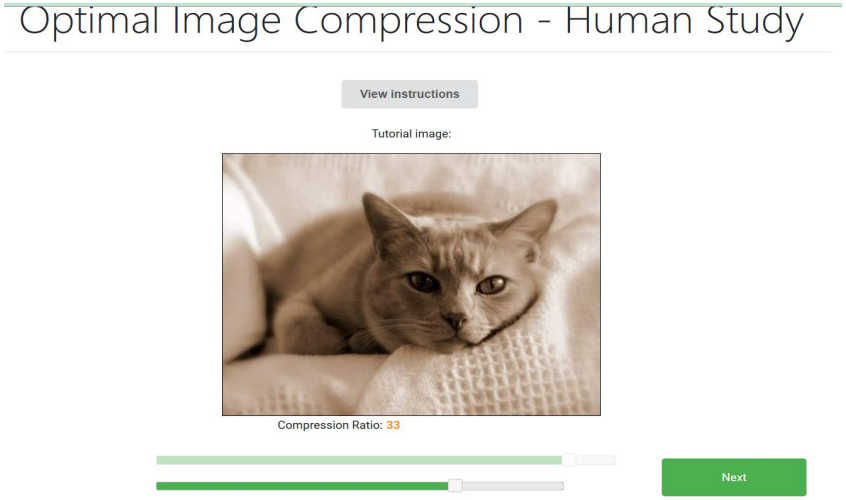


Design Solution



Amazon Mechanical Turk Human Study

- Use MTurk to crowdsource image compression scores from workers
- 8,500 images in dataset
- Workers use slider to control image compression
- Collect 2 slider values for each image: point where compressed image first changes and point where compressed image's quality slightly drops



Human Intelligence Task (HIT) Webpage

- Built using HTML for visual elements and Javascript for compression, recording scores, communicating with S3 bucket
  - S3 bucket: cloud storage service to store our images, data files
- Page contents: Instructions + quiz, 3 tutorial images, 55 actual images for data collection
- Integrity checks: make sure workers are taking the HIT seriously
  - Ex: slider values too low
  - Failing an integrity check rejects the worker mid-task

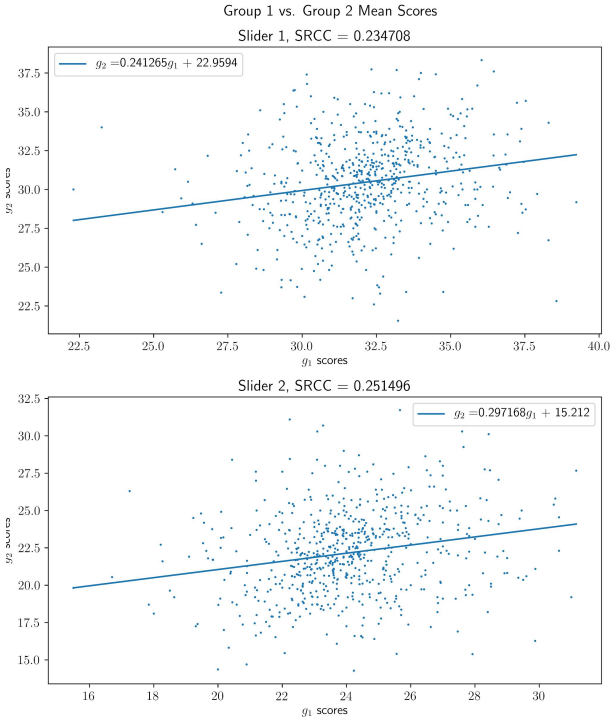
Conducting the Study

- Initial testing with small batches
- Over time, increased size of batches and number of concurrent batches
- Worker submissions require manual approval/rejection
  - Submissions auto-approve after 3 days
  - Aided by maybe_fail flag: flags worker if they pass original integrity checks but fail secondary (more strict) thresholds
  - Usual reasons for rejection: slider values too high/low/similar, task completed too quickly, SRCC score too low
- 6600 HITs approved, 360000 scores total, average of 30 scores per image

Predictive Model

- 3 main aspects to building the predictive model and its requirements
  - Data visualization and worker correlations
  - Data denoising
  - Model building and testing
- Interestingly - data correlation and denoising are some of the most important aspects to model building

Data Visualization and Correlations



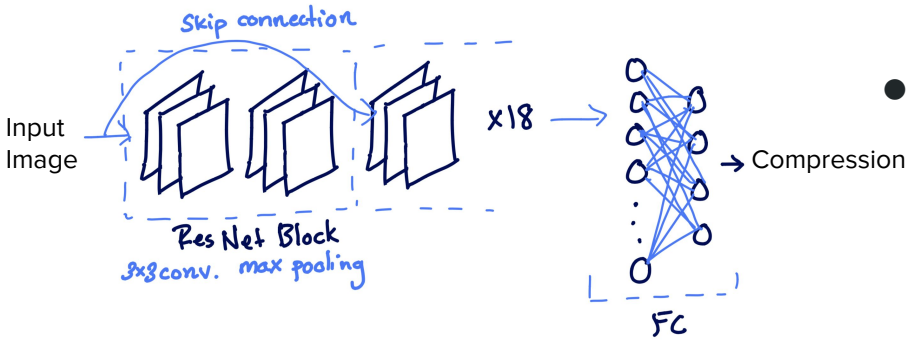
- Averaged the Spearman ranking of worker scores against our Golden Image Lab Study
  - Obtained a value of 0.5
- Inter-correlated the workers SROCC and LCC (Bovik et al. 2010)
  - SROCC and LCC - around 0.2 both sliders

Data Denoising and Preprocessing

- Crowdsourced data is noisy
- Two types of denoising
  - Standard Deviation Based Outliers
  - SRCC Outliers
- Used a ranking mechanism to eliminate bottom percentage of workers
  - Assigned penalty score to workers based off of deviance from either metric
  - Ranked workers according to penalty score
  - Eliminated bottom threshold percentage of workers

Model Building and Testing

Architecture Overview



- Final architecture: ResNet18 (fine tuned) with 2 layers of FC
  - Input: Source Image
  - Output: Optimal Compression Parameter
- Obtained final architecture by cross-validating over FC architectures and other hyperparams
  - Used multiprocessing - outsourced compute to use all available CPU and GPU cores
  - Used TACC

Web Application

- UI where users can go to submit their images, and returns the compressed version.
- Stats returned:
  - Size in KB
  - Compression parameter
- Order of operations:
  - Image is passed to the NN
  - It returns the compression parameter for that image
  - Image is compressed to that level

Image Compression Page

Browse... JPEGImages_2009_001078.jpg

Compress Image

Original Image:

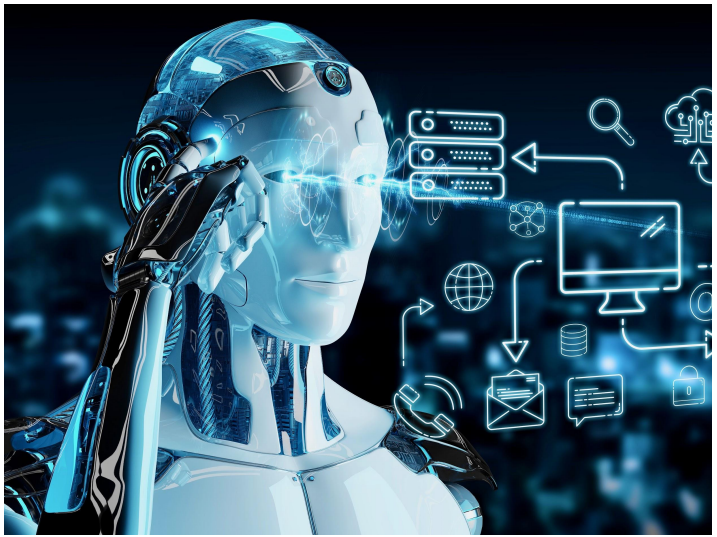
Size in Bytes + Padding: 278.78 KB

Compressed Image:

Size in Bytes + Padding: 87.63 KB

Compression Level: 69/100

Test and Evaluation (Human Perception)



- To test our model, we partitioned our dataset and occluded the test set from model
- Evaluated our model based off of one thing:

Did our model perform like a human?

- This means a successful model will do two things
  - Do our predictions fall within a standard deviation of the human image scores
  - Minimize deviance from mean score for each image
- 96% of the time our predictions fall within one standard deviation of the image scores.

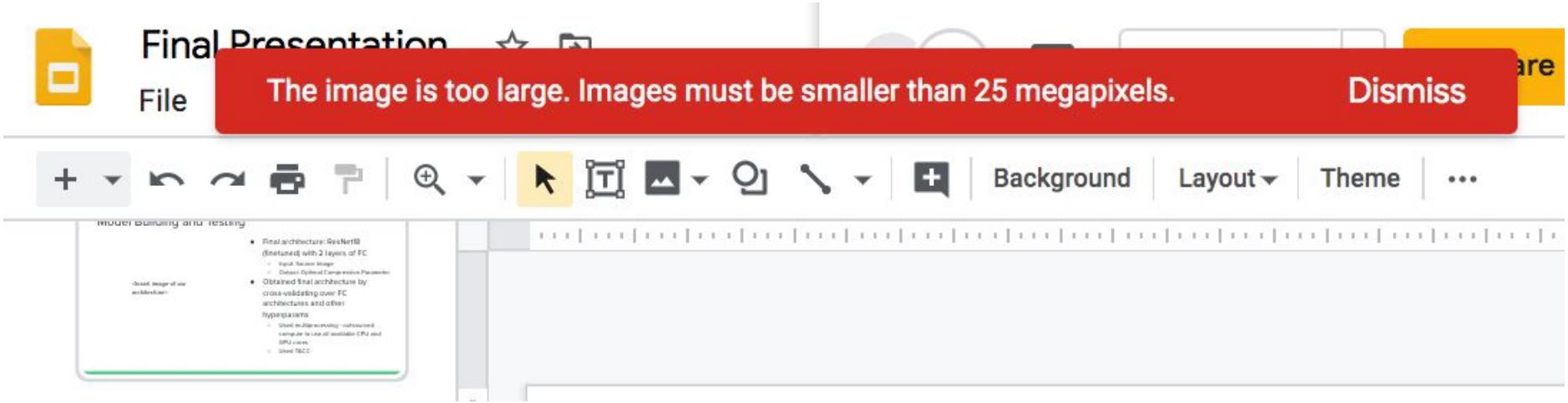
Test and Evaluation (Data Reduction)

- Used KonIQ-10K dataset (Hosu, et al. 2020) to determine compression performance
- Average Percent Reduction = 7.25%
- Standard Deviation of Percent Reduction = 14.41%
- Amounts to A LOT of savings over large databases
- Top = original, Bottom = compressed

Recommendations

- Things we would do differently:
  - Budget more time for building HIT page and conducting MTurk studies
  - Determine a way to verify correctness of a worker's slider scores, instead of having only methods to determine if they are wrong
- Future work:
  - Larger golden image study and MTurk study
  - Try out different models and architectures

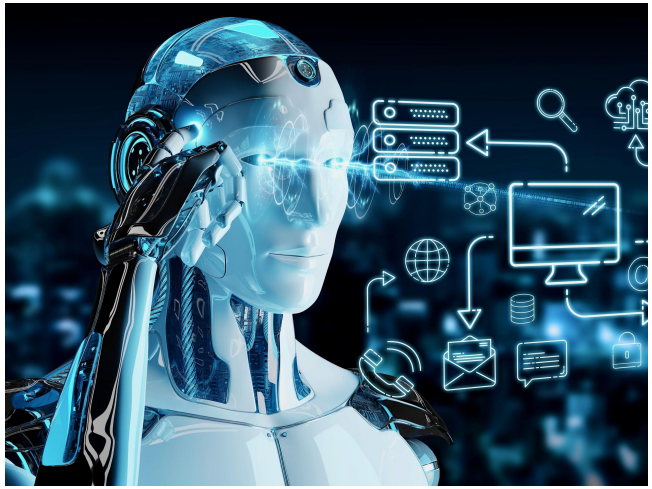
In Conclusion - Why is this Project Even Important?



When we tried to upload an image to use in this presentation...

In Conclusion - Why is this Project Even Important?

We used our tool to compress the images for this presentation!



A 43% reduction in memory and 0% drop in visual quality!

HIThere.jpg	5/6/2020 12:07...	JPG File	3,271 KB
HIThere_COMPRESSES.jpg	5/6/2020 12:08...	JPG File	1,892 KB

Compressed Again

HIThere_COMPRESSES.jpg	5/6/2020 12:08...	JPG File	1,892 KB
HIThere_COMPRESSES_COMP...	5/6/2020 1:46 ...	JPG File	1,890 KB

Thank for for watching and all of your incredible support this semester to make this possible!