## humans are awesome\*

#### \*compressors

(or: what machines can learn from humans about lossy compression)

AOMedia Symposium, October 21st, 2019

Tsachy Weissman Stanford

#### joint work (mainly) with:

Ashu Bhown (U of Michigan, until recently Palo Alto high school) Soham Mukherjee (UC Berkeley, until recently Monta Vista high school) Sean Yang (UC Berkeley, until recently St. Francis high school)



#### and

• Shubham Chandak, Irena Hwang & Kedar Tatwawadi (Stanford)



• Judith Fan (UCSD)





## image compression

- lossless: GIF, PNG
- lossy: JPEG, JPEG2000, WebP

## should we be happy?



### realistic to aim for this kind of a picture?



### what would Shannon do?



## entropy/compression of English text

- can we talk about fundamental limits?
- we can talk about achievability

Claude E Shannon, "Prediction and entropy of printed english," Bell system technical journal, vol. 30, no. 1, pp. 50–64, 1951.

#### Prediction and Entropy of Printed English By C. E. SHANNON

(ManuscriptReceived Sept. 15, 1950)

A new method of estimating the entropy and redundancy of a language is described. This method exploits the knowledge of the language statistics possessed by those who speak the language, and depends on experimental results in prediction of the next letter when the preceding text is known. Results of experiments in prediction are given, and some properties of an ideal predictor are developed.

## our goals

- provide a human centric approach to image compression:
- bring humans' shared language/experiences to bear
- utilize humans' shared knowledge (the Internet)
- tailor to what humans care about

#### understand what's achievable

## setup

- 2 humans with 2 distinct roles
- one is the "describer", the other the "reconstructor"
- describer gets a new image and sends a text describing it to the reconstructor
- reconstructor attempts to recreate the image

## enter



### set-up details

- Text Commands (Describer —> Reconstructor)
  - The describer is only allowed to send messages to the reconstructor through the built-in Skype text chat.
  - The describer must turn off their outgoing audio/video to avoid inadvertently leaking any information to the reconstructor.

- Feedback (Reconstructor —> Describer)
  - The reconstructor may talk to the describer through audio/video/text chat.
  - The reconstructor may share their partial reconstruction with the describer in real-time, by using the screen-share feature of Skype.

Experiment ends when describer is satisfied with the reconstruction (or wants to call it a day...)





Links of Public Images from the Internet

#### compressed representation

bzip2 encoded Skype transcript represents the final compressed representation of the input image

k nice ok gimme a sec just a heads up its a photo with a sunset and a bunch of balloon im trying to find similar sunsets and ballons rn \*hot air ballons https:// www.stockcutouts.com/Hot-Air-Balloon-Silhouette#.Wx7BZl0UvGI cut this out some how like maybe screenshot it?

balloons\_data.txt

# legit?

• "feedback" ok

• timing?

### Testing methodology

Evaluating the quality of the reconstruction by the human compressors vs WebP

1. Human compression: The given input image is compressed by the humans using the procedure described. The size (in bytes) of the compressed representation of the image (the text) is recorded.

2. WebP compression: We use the WebP compressor to lossily compress the input image to have a similar size as the human compression text representation.

3. Quality evaluation: We compare the quality of the WebP and human compressed images using human scorers on the Mechanical Turk platform.

### What a worker would see:

#### Instructions

The second image is a reconstruction of the first image.

• Compare the two images and rate your level of satisfaction from the reconstruction using the scale below (1=completely unsatisfied, 10=completely satisfied).

**Original Image:** 

Image Reconstruction:





Level of Satisfaction: 1 (completely unsatisfied) 2 3 4 5 6 7 8 9 10 (completely satisfied)

## examples

#### WebP

### example I:

#### Original



### example ii:

#### Original



Human Compressed

TRULE TO DE

11 11 11 11 11 -

WebP

### example iii:

#### Original







### example iv:

#### Original

I Zi

WebP



### example v:

#### Original



#### WebP





### example vi:

#### Original



#### WebP





## Results

#### ➢ Mturk scores for Human and WebP reconstruction

Image	Original	Compressed chat	WebP size	Mean score		Median score	
	size (KB)	size (KB)	(KB)	Human	WebP	Human	WebP
arch	1119	3.805	3.840	4.04	5.1	3	5
balloon	92	1.951	2.036	6.22	5.45	7	6
beachbridge	3263	4.604	4.676	4.34	3.92	4	4
eiffeltower	2245	4.363	4.394	5.98	5.77	6	6
face	1885	2.649	2.762	2.95	5.47	3	6
fire	4270	2.407	2.454	6.74	5.09	7	5
giraffe	5256	3.107	3.144	6.28	4.48	7	4
guitarman	1648	2.713	2.730	4.88	4.07	5	4
intersection	3751	3.157	3.238	6.8	4.15	7	4
rockwall	4205	6.613	6.674	4.41	4.85	4	5
sunsetlake	1505	4.077	4.088	5.08	4.82	5	5
train	3445	1.948	2.024	6.85	3.62	7	3
wolfsketch	1914	0.869	0.922	8.25	3.46	9	3

## reference

- "Towards improved lossy image compression: Human image reconstruction with public-domain images", Bhown et al., on arXiv
- see also "HAAC" website:

https://compression.stanford.edu/human-compression

### Conclusions thus far

- Our experiment shows much room for improvement over existing standards at low bit rate
- Effective utilization of semantically and structurally similar images that are publicly available can be key
- Humans care about different things (relevant loss function) and also, for humans, it's often less about fidelity and more about image quality

### what next?

- > HAAC for audio
- ➤ HAAC for facial images
- > automated and reproducible HAAC

(work in progress)

### Stanford Compression Forum

About • Events • People • Affiliate Members • Courses • Seminars • Projects • Outreach •

### Summer internships for high school students

STEM to SHTEM (Science, Humanities, Technology, Engineering and Mathematics)



details:

https://compression.stanford.edu/summer-internships-high-school-students



## HAAC for music

## existing audio compression standards

- "lossless": WAVE (.wav), FLAC (.flac), and APE (.ape)
- lossy: MP3 (.mp3) AAC (.mp4, .m4a), OGG (.ogg), and Musepack (.mpc)

### how does a human perceive/represent music?

- score
- lyrics
- voice of vocalist(s)



## conversion-tool

HOME CONVERTERS ~ CONTACT US

#### Convert Audio to MIDI

📫 Like 6 😏 Tweet		1	+ Share 7
SPONSORED SEARCHES			🔈
audio converter	٩) (	midi files	٩
audio recording	٩ 🦳	midi music	٩

Use this tool to convert audio files such as **mp3**, **wav**, **ogg**, **m4a**, **and many other formats** to **MIDI**. Note: The quality of the resulting MIDI file depends largly on the input music. The MIDI file quality can range from good quality to unusable. But it is great fun to listen to the result in either way.

Song	Original MP3 file size	MIDI file size	Compressed MIDI size	Compression ratio MP3 → Compressed MIDI
Axel (Crazy Frog)	1MB	34KB	9KB	0.0087
Hey Brother	2MB	77KB	69KB	0.033
Sweet Home Alabama	2MB	53KB	24KB	0.011
Take me home country roads	3MB	13 KB	3KB	0.0009

## listen

### Sweet home Alabama by Lynyrd Skynyrd

### some points

- humans can perceive and describe music succinctly
- garage band can produce reasonable reconstructions based on little (MIDI)
- humans often value "quality" over fidelity
- humans can produce exquisite reconstructions based on little (the score)

## HAAC for facial images



## toward automated reproducible HAAC



### some current/future directions

- ML & AI toward fully automated delivery on what we've shown is achievable
- construction of a good (offline) Side-Information database

## HAAC for video?



## user defined/specific metrics ?

![](_page_39_Picture_1.jpeg)

![](_page_39_Picture_2.jpeg)

# thank you!

## questions?