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
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...)
The diagram explains the process of a collaborative image reconstruction task. It involves two main stages: Describer and Reconstructor.

- **Describer**:
  - Original image
  - Links of Public Images from the Internet
  - Only Describer sees the original image
  - Text chat (Skype chat)
  - Audio feedback (voice chat)
  - Visual feedback (screen-share on Skype)
  - Both have access to the internet

- **Reconstructor**:
  - Photoscape X
  - Final Reconstruction

This workflow emphasizes the interaction between the Describer and Reconstructor, utilizing various communication tools for effective collaboration.
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#.Wx7BZlOUvGI
cut this out some how
like maybe screenshot it?
check that works
```

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.

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).

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

Original

WebP

Human Compressed
example ii:

Original

WebP

Human

Compressed
example iii:
example iv:

Original

WebP

Human Compressed
example v:
example vi:
### Results

➢ Mturk scores for Human and WebP reconstruction

<table>
<thead>
<tr>
<th>Image</th>
<th>Original size (KB)</th>
<th>Compressed chat size (KB)</th>
<th>WebP size (KB)</th>
<th>Mean score</th>
<th>Median score</th>
</tr>
</thead>
<tbody>
<tr>
<td>arch</td>
<td>1119</td>
<td>3.805</td>
<td>3.840</td>
<td>4.04</td>
<td>3</td>
</tr>
<tr>
<td>balloon</td>
<td>92</td>
<td>1.951</td>
<td>2.036</td>
<td>6.22</td>
<td>7</td>
</tr>
<tr>
<td>beachbridge</td>
<td>3263</td>
<td>4.604</td>
<td>4.676</td>
<td>4.34</td>
<td>4</td>
</tr>
<tr>
<td>eiffeltower</td>
<td>2245</td>
<td>4.363</td>
<td>4.394</td>
<td>5.98</td>
<td>6</td>
</tr>
<tr>
<td>face</td>
<td>1885</td>
<td>2.649</td>
<td>2.762</td>
<td>2.95</td>
<td>3</td>
</tr>
<tr>
<td>fire</td>
<td>4270</td>
<td>2.407</td>
<td>2.454</td>
<td>6.74</td>
<td>7</td>
</tr>
<tr>
<td>giraffe</td>
<td>5256</td>
<td>3.107</td>
<td>3.144</td>
<td>6.28</td>
<td>7</td>
</tr>
<tr>
<td>guitarman</td>
<td>1648</td>
<td>2.713</td>
<td>2.730</td>
<td>4.88</td>
<td>5</td>
</tr>
<tr>
<td>intersection</td>
<td>3751</td>
<td>3.157</td>
<td>3.238</td>
<td>6.8</td>
<td>7</td>
</tr>
<tr>
<td>rockwall</td>
<td>4205</td>
<td>6.613</td>
<td>6.674</td>
<td>4.41</td>
<td>4</td>
</tr>
<tr>
<td>sunsetlake</td>
<td>1505</td>
<td>4.077</td>
<td>4.088</td>
<td>5.08</td>
<td>5</td>
</tr>
<tr>
<td>train</td>
<td>3445</td>
<td>1.948</td>
<td>2.024</td>
<td>6.85</td>
<td>7</td>
</tr>
<tr>
<td>wolfsketch</td>
<td>1914</td>
<td>0.869</td>
<td>0.922</td>
<td>8.25</td>
<td>9</td>
</tr>
</tbody>
</table>
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)
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)
<table>
<thead>
<tr>
<th>Song</th>
<th>Original MP3 file size</th>
<th>MIDI file size</th>
<th>Compressed MIDI size</th>
<th>Compression ratio MP3 → Compressed MIDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axel (Crazy Frog)</td>
<td>1MB</td>
<td>34KB</td>
<td>9KB</td>
<td>0.0087</td>
</tr>
<tr>
<td>Hey Brother</td>
<td>2MB</td>
<td>77KB</td>
<td>69KB</td>
<td>0.033</td>
</tr>
<tr>
<td>Sweet Home Alabama</td>
<td>2MB</td>
<td>53KB</td>
<td>24KB</td>
<td>0.011</td>
</tr>
<tr>
<td>Take me home country roads</td>
<td>3MB</td>
<td>13 KB</td>
<td>3KB</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
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

Police Sketch Artist Case study:

~ KB of Data

40 questions
toward automated reproducible HAAC
**Compression Method**

1. Original Image
2. Describer searches for matching elements
3. Describer edits elements via script to make a composite
4. Script parsed into string, improving compression

**Decompression Method**

1. String translated back into script
2. Script executed to produce a composite image

**Data Compression Results**

- Script-Based Compression: 318 bytes
- Natural Language Compression: 3183 bytes
- WebP Compression [2]: 3231 bytes
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 ?
thank you!

questions?