Human-Centric Compression:

What can lossy compression learn from humans?

Soham Mukherjee (Monta Vista), Sean Yang (St. Francis) and Ashu Bhown (Palo Alto)
Shubham Chandak, Irena Fischer-Hwang, Kedar Tatwawadi and Tsachy Weissman (Stanford)

Lossy image compression

- Explosion in digital images requires increasingly more storage space
  - Example: 12 megapixel camera on iPhone X
    - Total of 36 (RGB) or 18 MB per image (YUV 4:2:0)
    - Sharing a photo album with just 100 pictures takes at least 1.8 GB data to be transmitted

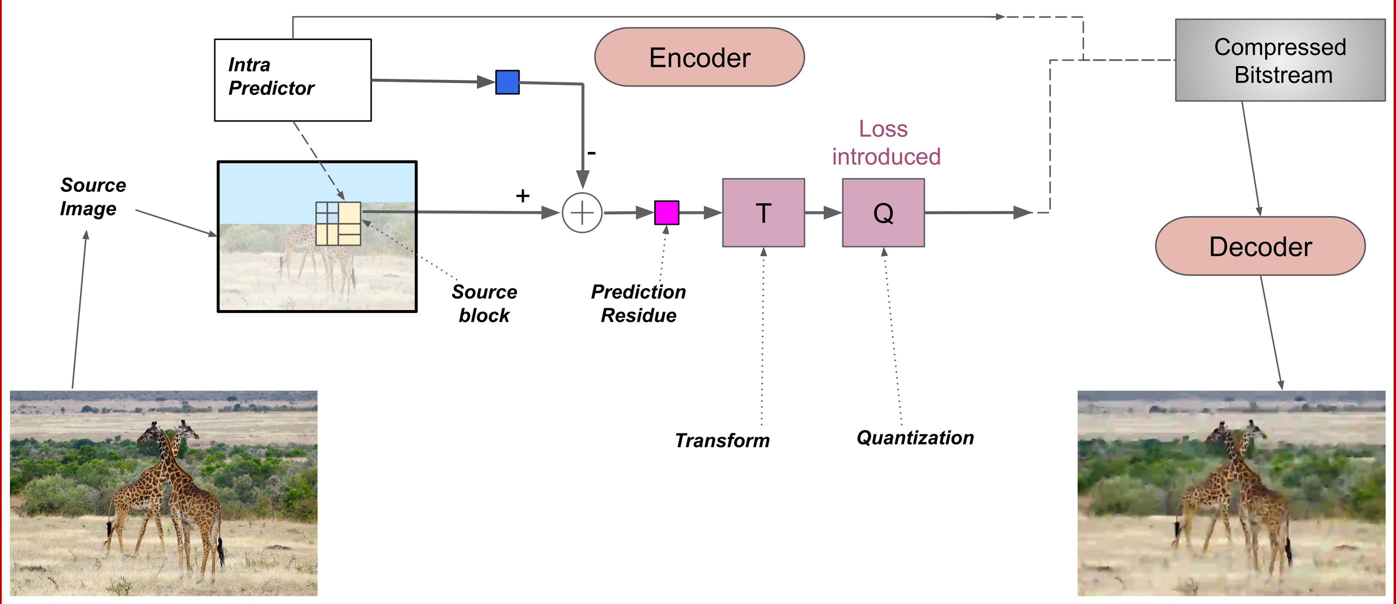


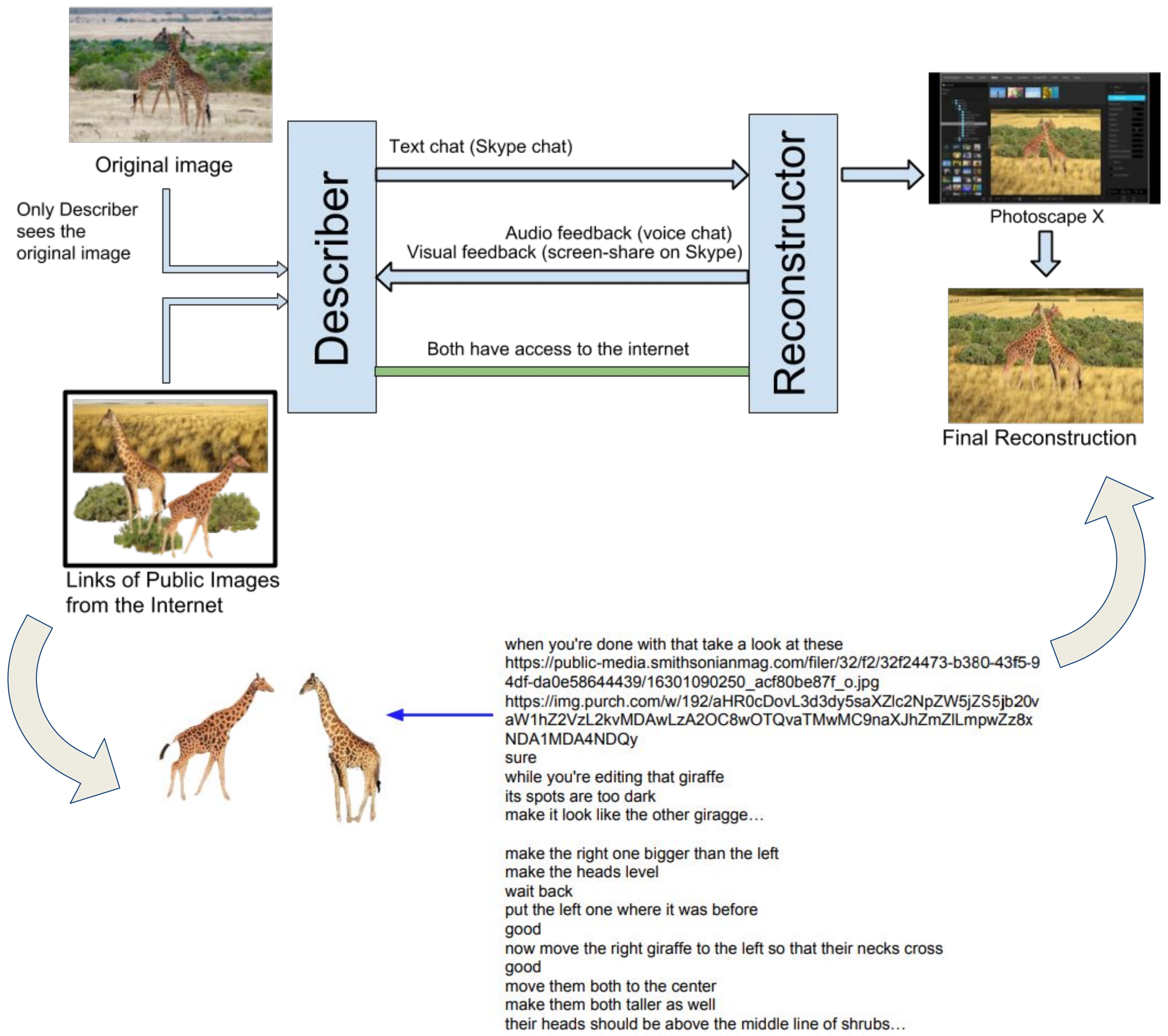
Figure: Typical lossy image compression framework

- Lossless compression gives only ~2:1 compression (on average)
  - Some loss must be tolerated
- Flaws of traditional lossy compressors, especially at low bit rates:
  - Staircase noise (aliasing) along curved edges
  - Blockiness
  - Posterization
  - Generation Loss

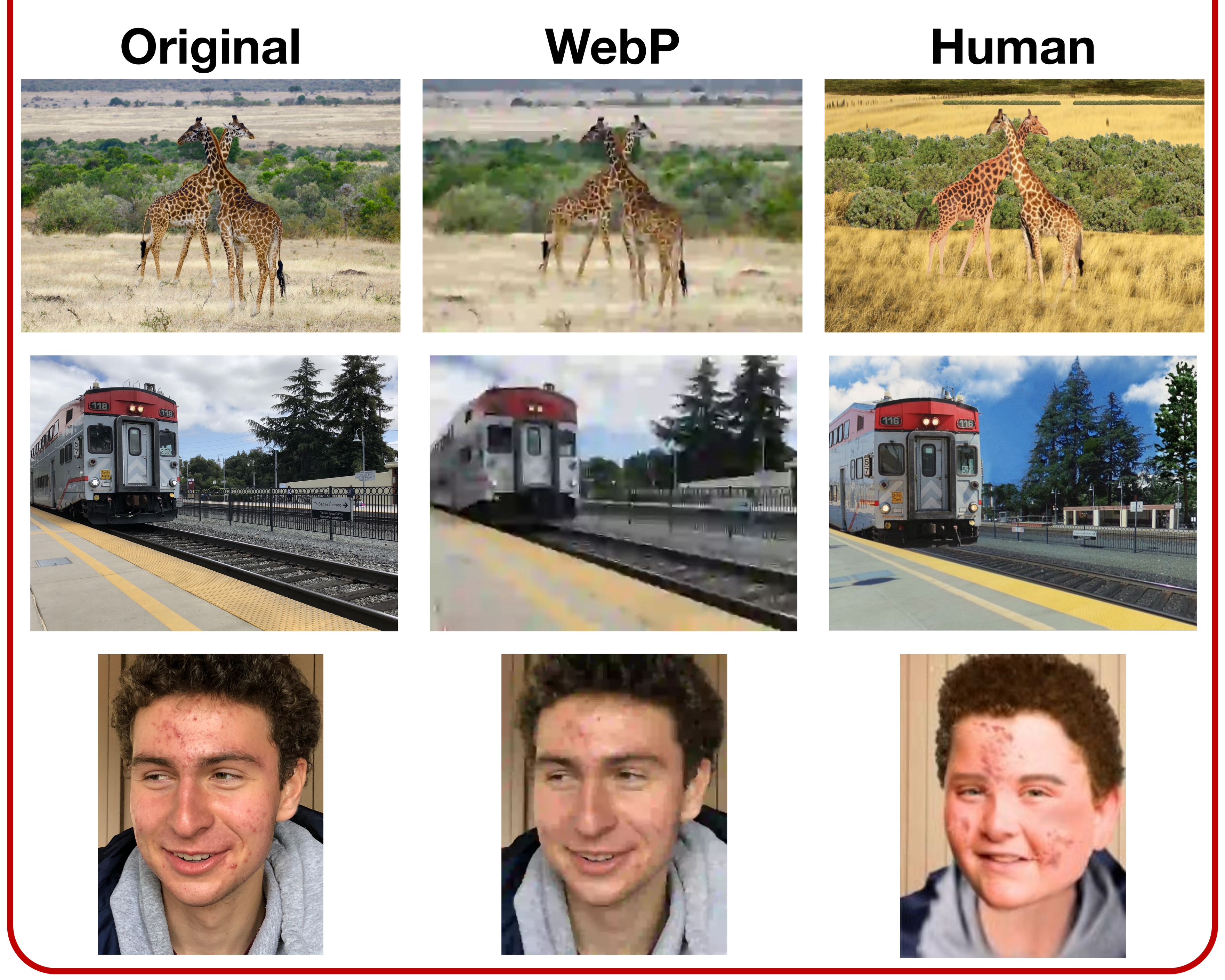
Human-centric compression

- Question: Can we create more efficient lossy compressors by preserving only what humans perceive as important?
- Goals:
  - A more human-centric approach to image compression
  - Use of the vast public repository of images already available on the Internet
- The human-centric approach:
  - Optimize for what humans care about by prioritizing high level semantic descriptions rather than arbitrary loss functions (RMSE, SSIM, etc.)
  - Leverage the efficiency of human language (rather than encoding and decoding pixels)

Methods



- When reconstruction has been completed to the describer's satisfaction, the compression experiment is stopped
- The processed and bzip2-compressed text transcript is the compressed representation of the original image



Testing and Results

- Used WebP to lossily compress the original image to a size similar to the text-based representation
- Used Amazon Mechanical Turk to evaluate images

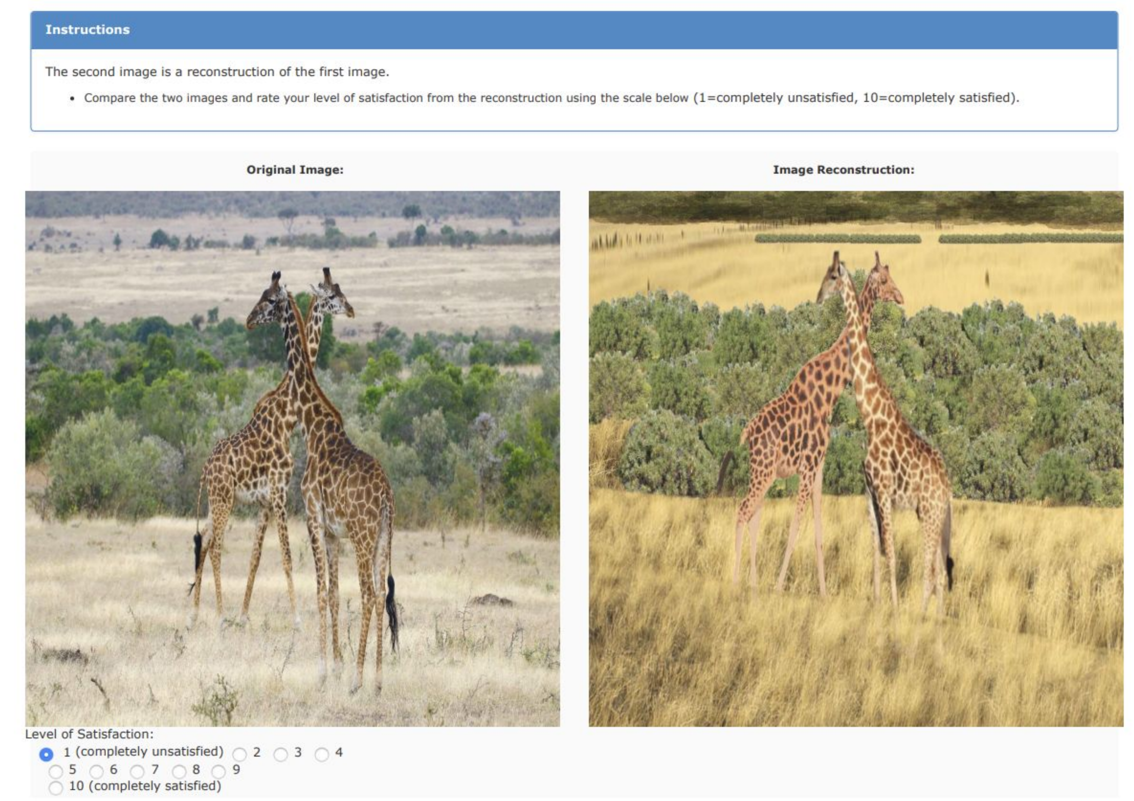


Image	Original size (KB)	Compressed chat size (KB)	WebP size (KB)	Mean score		Median score	
				Human	WebP	Human	WebP
arch	1119	3.805	3.840	4.04	**5.1**	3	**5**
balloon	92	1.951	2.036	**6.22**	5.45	**7**	6
beachbridge	3263	4.604	4.676	**4.34**	3.92	4	4
eiffeltower	2245	4.363	4.394	**5.98**	5.77	6	6
face	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

- Human compression can outperform traditional compression at very low bit rates
- Using semantically and structurally similar images from a large database can dramatically improve compression ratio
- Demonstrated room for growth in lossy compression

Future work

- Human compression framework is useful as an exploratory tool, but not practical:
  - Use GANs to perform description and reconstruction[1]
  - Use neural networks to predict human scores[2]

References

1. Agustsson, Eirikur, et al. "Generative Adversarial Networks for Extreme Learned Image Compression." arXiv preprint arXiv:1804.02958 (2018).
2. Chinen, Troy, et al. "Towards A Semantic Perceptual Image Metric." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

Our paper:
Bhown, Ashutosh, et al. "Humans are still the best lossy image compressors." arXiv preprint arXiv:1810.11137 (2018).