NeurIPS 2025

Saliency-guided deployment-adaptive compression for wildlife camera traps

Tianhong Xie[1], Justin Kay[2], Timm Haucke[2], Sara Beery[2]   tixie@hmc.edu

1 HARVEY MUDD COLLEGE   2 Massachusetts Institute of Technology

NEURAL INFORMATION PROCESSING SYSTEMS

Computer Vision for Camera Traps

Automatically triggered cameras, also called **camera traps**, are an indispensable tool for studying **animal ecology**.

Since they are often deployed far away, data either has to be retrieved physically or transmitted via low-bandwidth transmission.

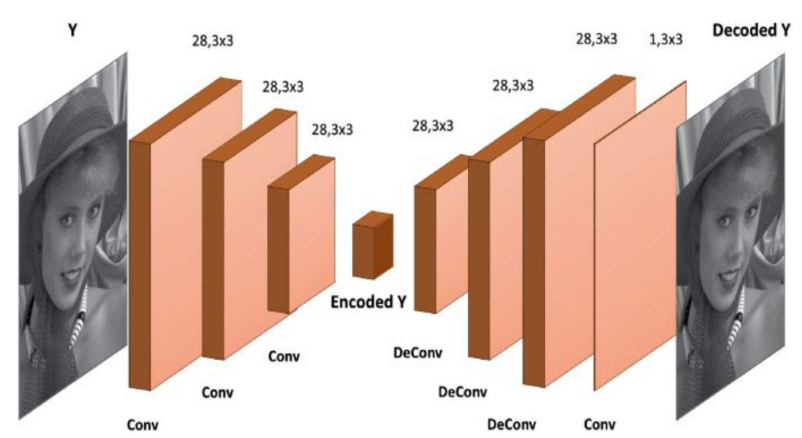
Real-time analysis is sometimes necessary: for example, to detect invasive species. Efficient on-device compression is critical.

Unfortunately, **traditional image codecs (JPEG) fail to achieve sufficient compression ratios for satellite transmission**. We thus explore autoencoders.

Autoencoders for image compression

Autoencoders are great for image compression because they consistently outperform algorithms like JPEG in terms of compression ratio and reconstruction quality.

We propose a method for efficient deep-learning based image compression to address this challenge, combined with full / low-rank adaptation (LoRA) fine-tuning and saliency-guided compression.



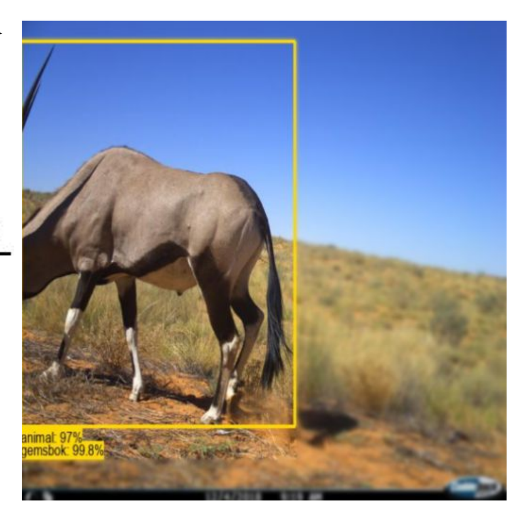
Saliency-Guided Compression

$$Loss = \lambda \star 255^2 \star D_{new} + R$$

$$D_{new} = \frac{\lambda_{animal} * MSE_{animal} + \lambda_{background} * MSE_{background}}{total\_weighted\_pixels}$$

A weighted MSE loss allows the model to assign lower penalties to reconstruction errors in less important background regions.

Gaussian Blur



MegaDetector is used to intentionally blur out the background.

Doing so reduces image entropy, which means each pixel requires fewer bits to represent, essentially increasing the compression ratio.

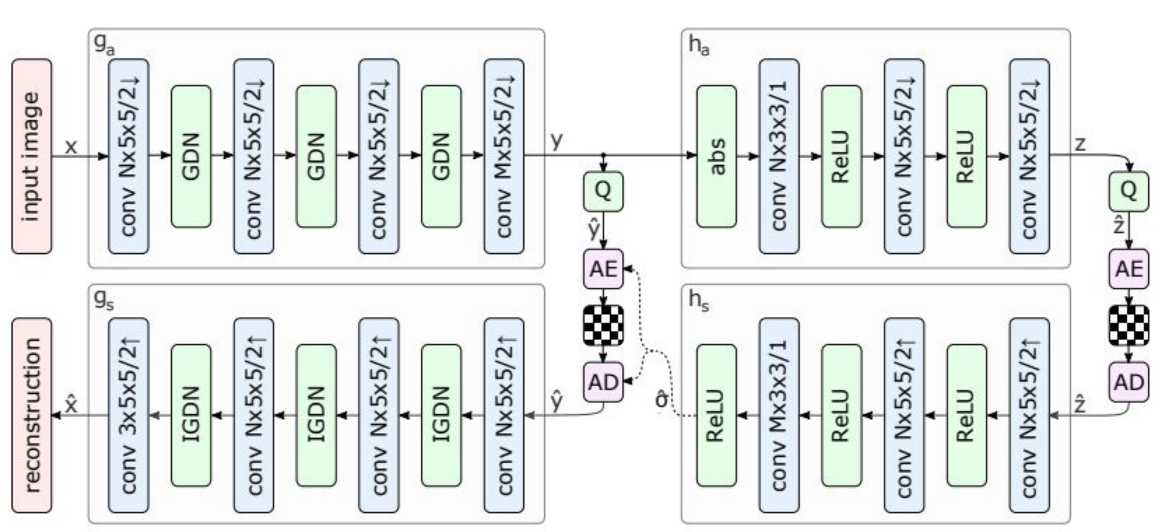
We also derive a new weighted MSE loss based on our method.

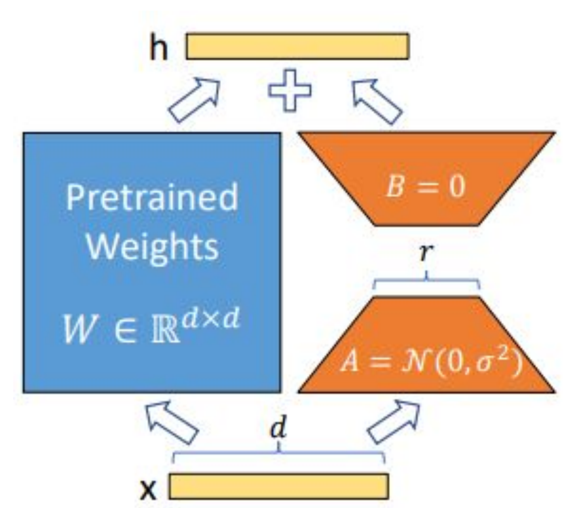
Low-Rank Adaptation fine-tuning

Low-ranked adaptation (LoRA) is a parameter efficient technique typically used for lightweight LLM fine-tuning.

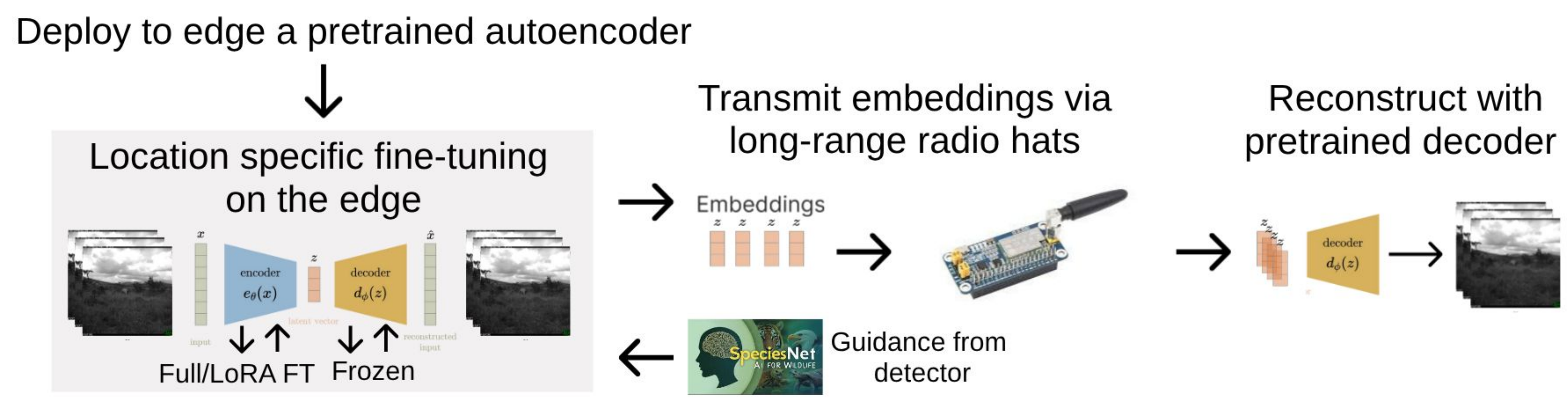
Decoder: Frozen   Encoder: full / LoRA fine-tune

The convolutional layers are decomposed into smaller matrices of A and B, reducing GPU memory usage by up to 70% in our experiments. GDN and deconvolutional layers stay untouched.



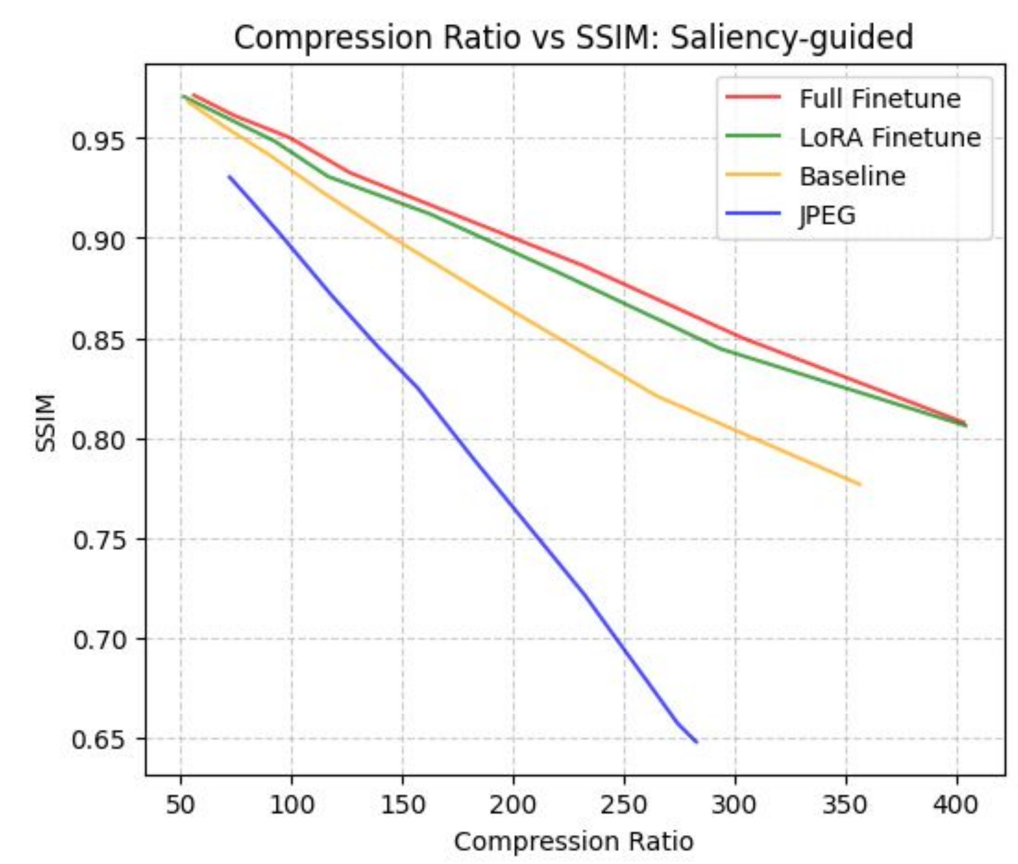


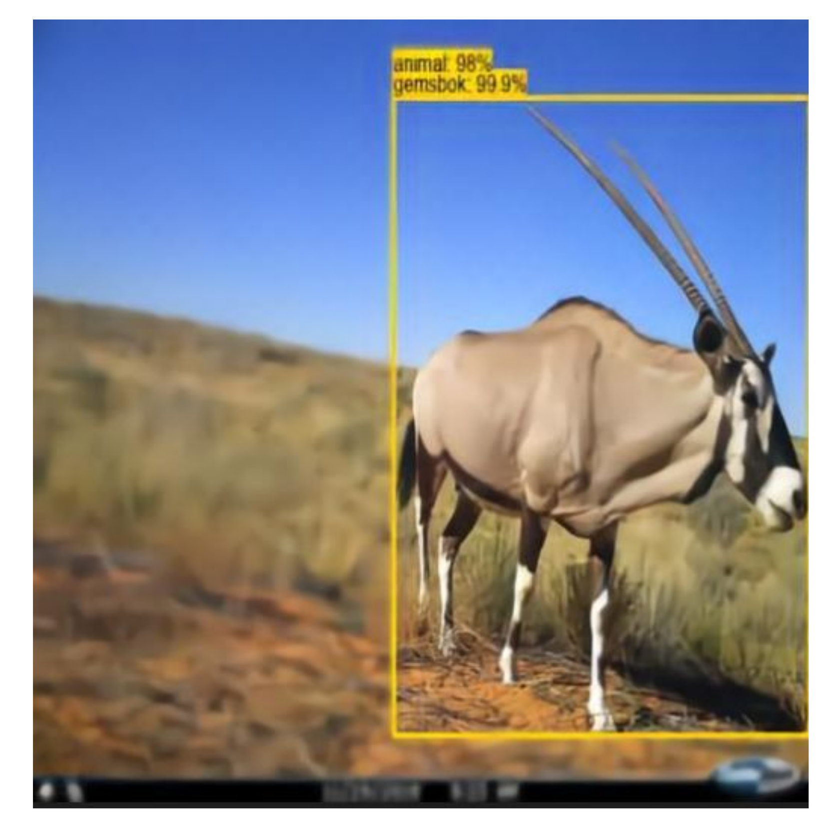
Proposed Approach: Site-Specific On-Device PEFT

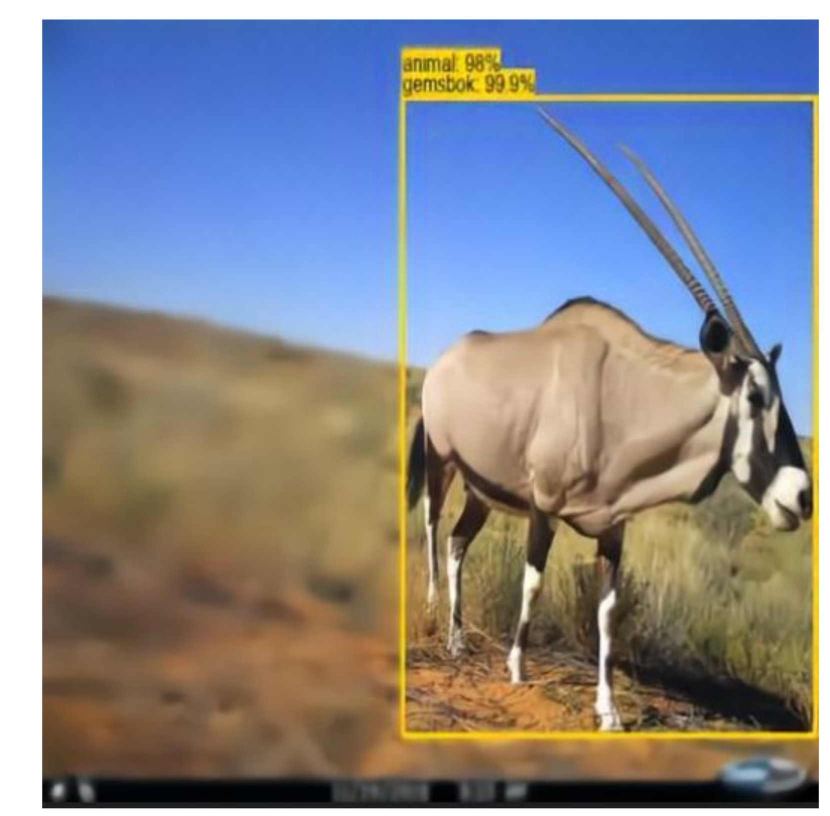


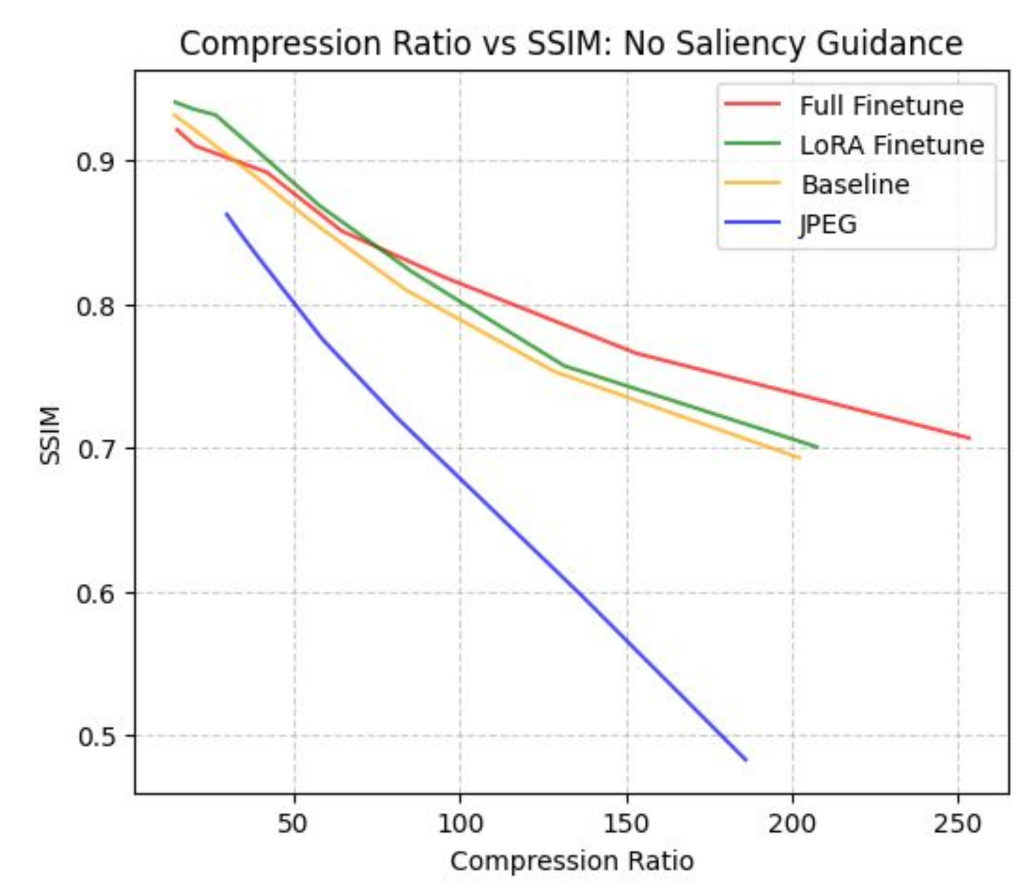
- As a case study, we transmit embeddings via long-range radio (LoRa, not to be confused with low-rank adaptation). We use SX1262 LoRa hats connected to a Raspberry Pi.
- LoRa hats are inexpensive and do not require external Wifi support.

Results









- Saliency-guided compression (0.001:1 background vs animal weight) significantly increased the maximum compression ratio (~250 -> ~400).
- Saliency-guided compression has a better animal region quality despite being more compressed in the background.
- LoRA fine-tuning yields slightly worse performance, but it is a trade-off between performance and efficiency.
- Overall, our method significantly improved upon the existing baseline in image compression and demonstrates suitability for environmental camera-trap applications.

References:

Johannes Ballé, Valero Laparra, and Eero P. Simoncelli. End-to-end optimized image compression. International Conference on Learning Representations (ICLR), 2017. https://arxiv.org/abs/1611.01704.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. arXiv:2106.09685, 2021.

Acknowledgement:

I would like to thank members of MIT Beery Lab for their mentorship (Justin Kay, Timm Baucke, Prof Sara Beery). I would also like to thank Harvey Mudd College OCE for its funding of Ben Huppe '14 Memorial Internships for a Sustainable World