# <u>Bottleneck-based Encoder-decoder AR</u>chitecture (BEAR) **Tor Learning Unbiased Consumer-to-Consumer Image**

## Representations

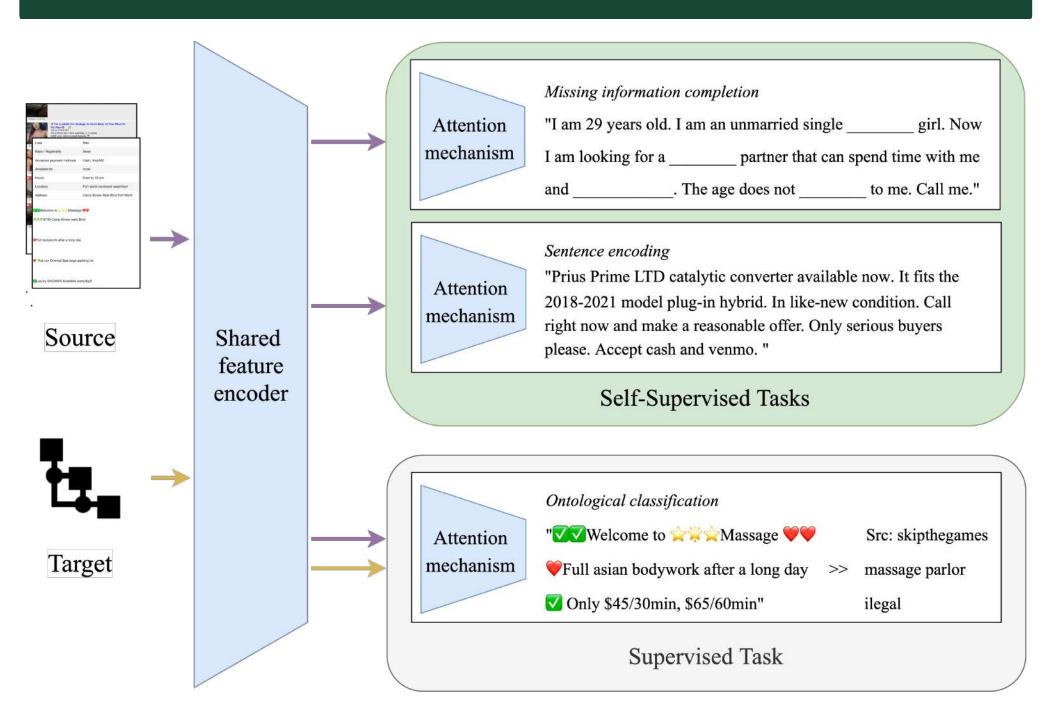
Pablo Rivas,<sup>1</sup> Gisela Bichler,<sup>2\*</sup> Tomas Cerny,<sup>1\*</sup> Laurie Giddens,<sup>3\*</sup> and Stacie Petter<sup>4\*</sup>

\*Equal Contribution. <sup>1</sup>Department of Computer Science, School of Engineering and Computer Science, Baylor University, Texas, USA. <sup>2</sup>School of Criminology and Criminal Justice, California State University, San Bernardino, California, USA. <sup>3</sup>Information Technology and Decisions Sciences Department, University of North Texas, USA. <sup>4</sup>School of Business, Wake Forest University, North Carolina, USA

#### **Contributions**

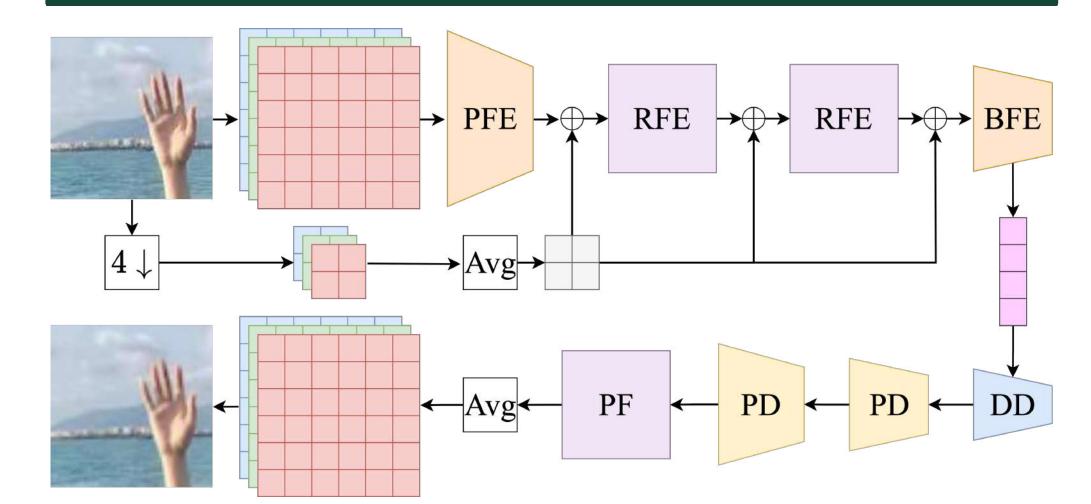
- ► We design an autoencoder model that learns rich features that produce latent representations of images.
- ➤ Our model uses modern machine learning techniques to learn low-dimensional representations at scale with few parameters in comparison to other similar models.
- ► The proposed model hides personal identifiers from plain sight while keeping necessary information for downstream trafficking detection tasks.

#### **Background and Motivation**

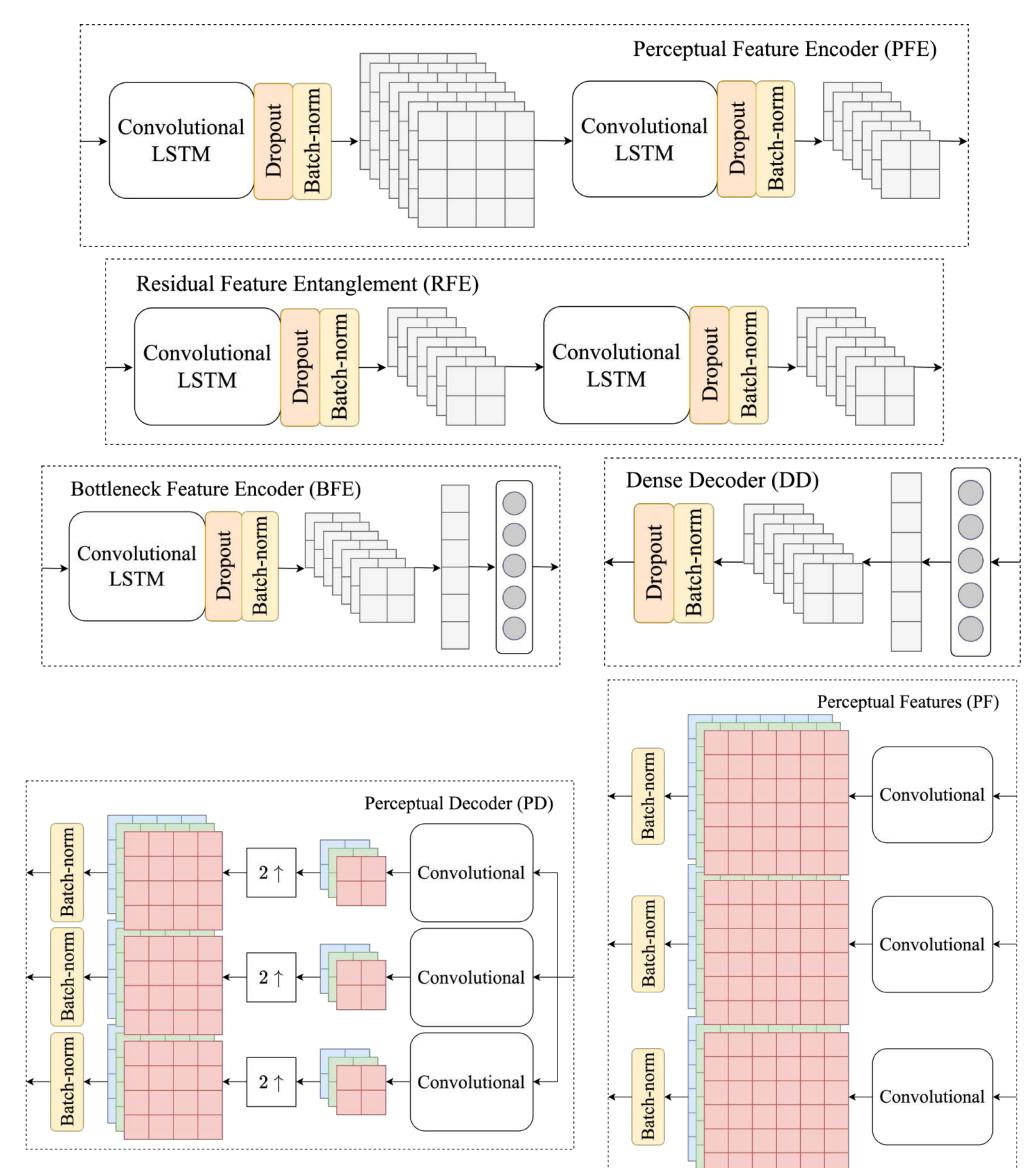


- ➤ Consumer-to-Consumer Marketplaces. With the COVID-19 global pandemic, C2C marketplaces became heavily used, fostering an increase in illegal activities such as human trafficking and trafficking of stolen goods.
- ▶ **Big Picture**. We want to extract multimodal features from C2C data using attention mechanisms on self-supervised, contrastive, and supervised learning tasks.
- ▶ **SOTA**. Most transformer-based approaches such as BYOL or ViT, are data-hungry and too large to train from scratch or fine-tune. Other big models such as CLIP or GPT can cost hundreds of millions of dollars to train. We need lightweight, scalable, image representations that <a href="https://hichestor.org/hichestor.org/">hide</a> personally identifiable information from plain sight.

#### **BEAR Design Paradigm**

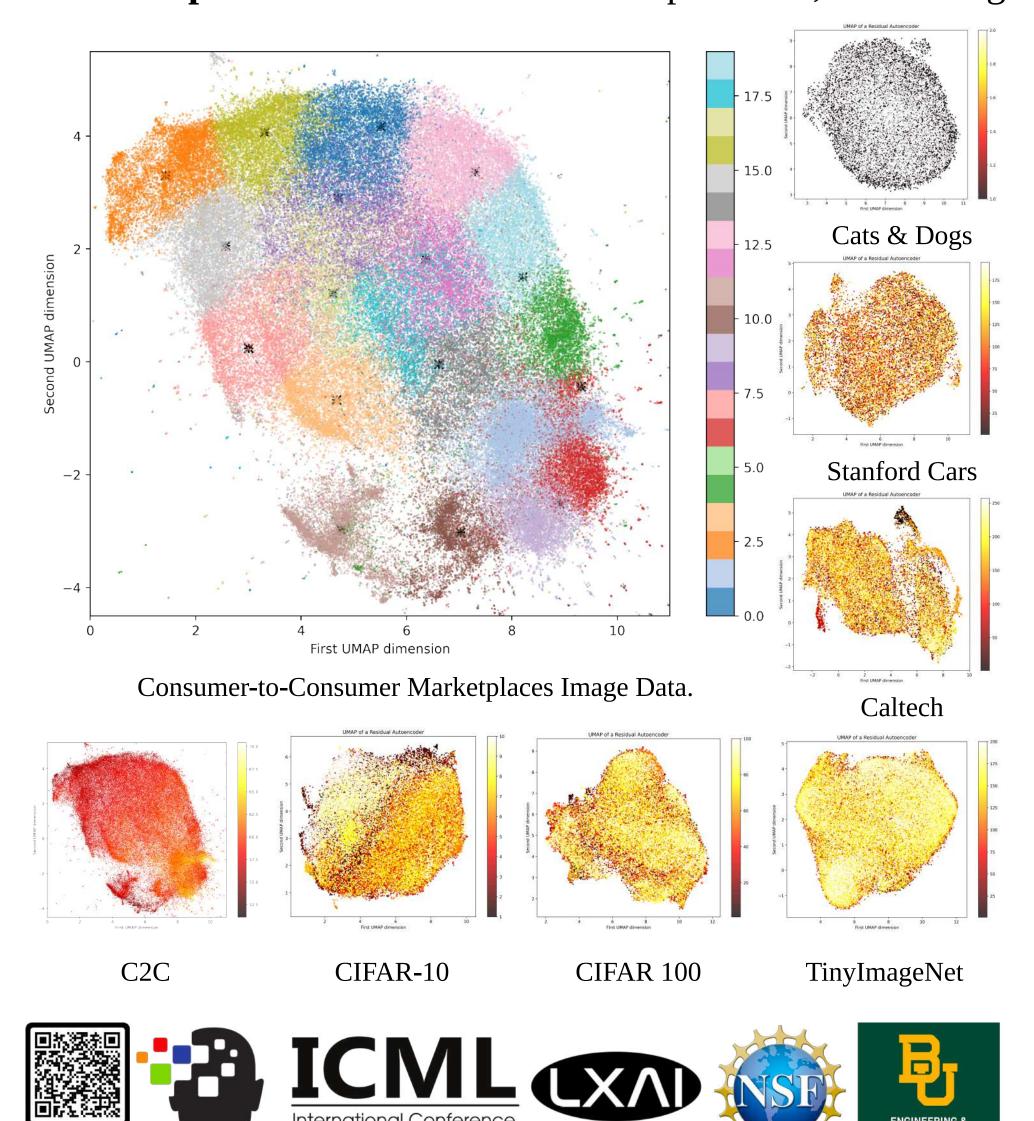


This preliminary study was funded by the National Science Foundation under grants CHE-1905043, CNS-2136961, and CNS-2210091.



### Why Does It Work?

- ▶ Data augmentation was key to successful self-supervision.
- ► High-quality latent space due to perceptual feature learning.
- ► Scalability granted by convolutional-based design.
- ▶ Baseline model is only 4M, cf. ViT 86M, BYOL 250M.
- ▶ **Next steps**: label-based self & weak supervision, fine-tuning.



On Machine Learning