*Technical accomplishments*

Our group concentrates on video and image processing on compressed domain in general. Despite numerous lossy compression codecs like JPEG and MPEG, we are still lack of understanding of the boundary of lossy compression. We research on data-driven deep learning-based compression techniques. In fact, we manage to build a variable compression rate network and outperform classical codecs like JPEG on well-known benchmarks. With more bits of code, our trained network can perform better. The state-of-art of data-driven codec, especially for video, are still quite rudimentary and we will keep investigating this track.

Deep-learning based codecs also open a path for performing image processing on compressed domain. We manage to jointly train a variable compression rate network which is capable of compressing and classifying images at the same time. On MNIST dataset, we can achieve more than 50 dB PSNR with 1 bit per pixel and 98.5% accuracy. We are also investigating more complex dataset like CIFAR10 and ImageNet. For CIFAR10 dataset, we can achieve 35 dB PSNR with 2 bit per pixel but the accuracy is not as good as the state-of art. For full resolution images like ImageNet and MS COCO dataset, patches of an image are currently trained independently. We believe it is possible to take advantage of the dependency among different patches by using time-series models like LSTM and GRU. We are also researching on distributed codec systems. Distributed encoders are isolated from each other and compress images in their own way. Under the assumption that each encoder takes in similar images, the hypothesis is that we can perform better in terms of compression by decoding the codes jointly rather than having a decoder for each encoder separately.

*Plans for next year*

We plan to research more on videos in addition to images next year. Deep learning-based video codecs are rudimentary and most of them only serve as auxiliary functionality for MPEG. With the time-series models for patches of images, we should be able to incorporate dependent patches from neighboring frames with proper motion adjustment. For independent images, we plan to build a web crawling software which can take in unlimited amount of data. Unlike image classification and detection which are limited by labeled data, compression can naturally generate features and create an embedding without any supervised data.

As for video and image processing, we plan to enlarge our scope of objectives. Apart from compression and classification, detection and segmentation can also be jointly optimized in a single neural network. Since a human can do multiple jobs with a single brain, it would be interesting to push the limit of multi-objective optimization in deep learning models. Furthermore, object detection and segmentation also play a crucial role in video surveillance. For example, we can use our classifier to determine the existence of anomaly in compressed domain, use our variable rate decoder to reconstruct the images, and use our detector to detect the anomaly pixel-wisely, and this will all happen in a single neural network. To jointly optimize the network, we may also encounter data that are unlabeled for some tasks. Therefore, we also plan to research on semi-supervised multi-objective training. The hypothesis is that we can label our data by utilizing the knowledge from the network trained with other objectives. One example is that images from the same class should also have similar codes.