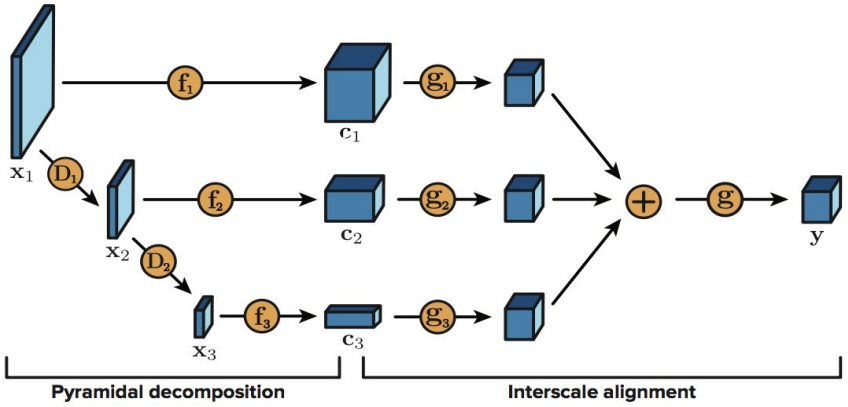
Real-Time Adaptive Image Compression (Cons)

Oren Rippel & Lubomir Bourdev
WaveOne Inc.

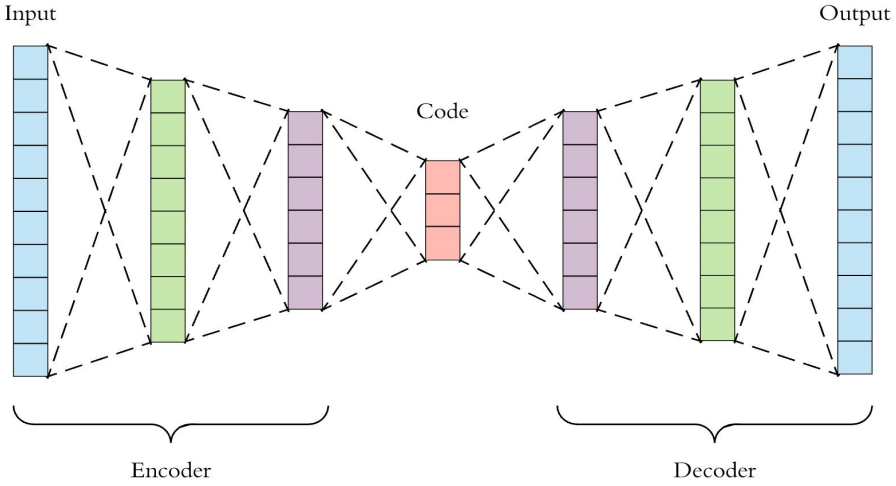
Presenter: Abhinav Deep Singh

#1 - Decisions not backed up empirically (1)

- Why pyramid Structure used for encoding?
- Performance on simple encoder/decoder model vs pyramid structure?



vs



#1 - Decisions not backed up empirically (2)

- Why B=6 used for quantization? What's the effect of other values?

3.2.1. QUANTIZATION

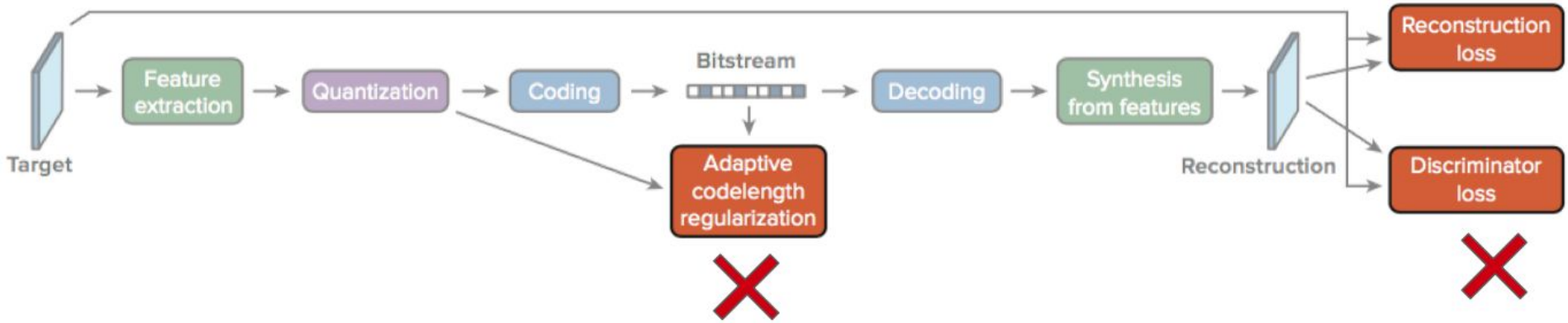
Given a desired precision of B bits, we quantize our feature tensor $\mathbf{y}$ into 2^B equal-sized bins as

$$\hat{y}_{chw} := \text{QUANTIZE}_B(y_{chw}) = \frac{1}{2^{B-1}} \left\lceil 2^{B-1} y_{chw} \right\rceil .$$

For the special case $B = 1$, this reduces exactly to a binary quantization scheme. While some ML-based approaches to compression employ such thresholding, we found better performance with the smoother quantization described. We quantize with $B = 6$ for all models in this paper.

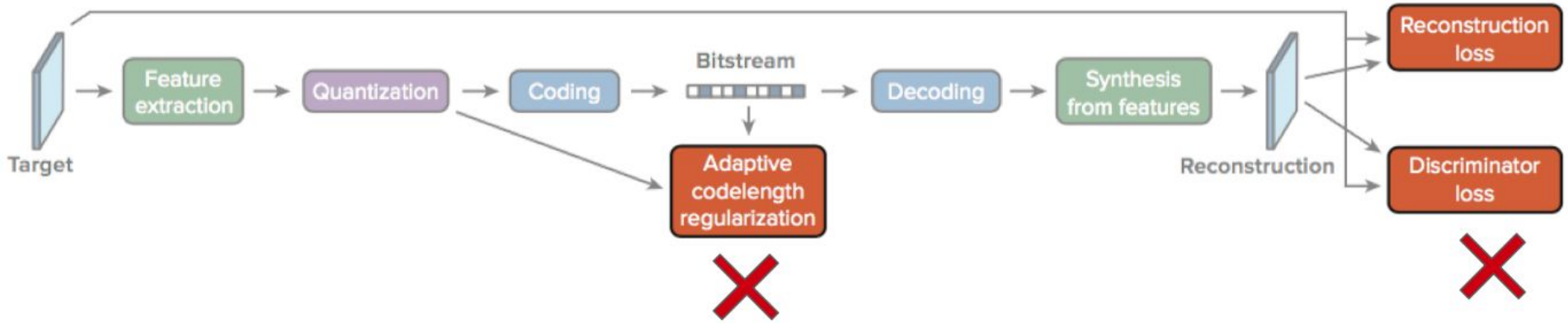
#1 - Decisions not backed up empirically (3)

- What's the performance of baseline model?
- Without ACR Loss performance?



#1 - Decisions not backed up empirically (4)

- Why use GAN loss at all? Why reconstruction loss not enough?
- Without GAN Loss performance?



#2 - Training GANs

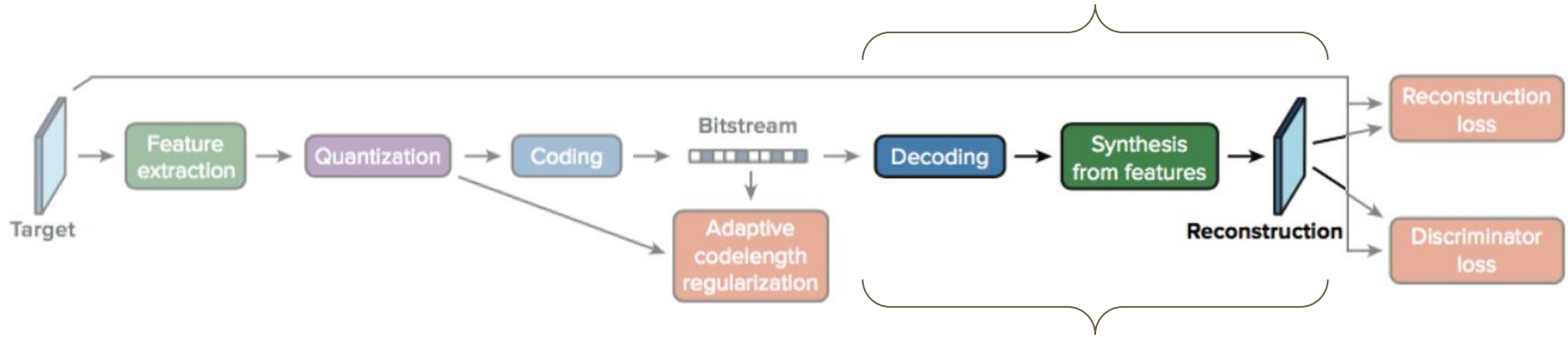
- Complex training mechanism, probably to avoid model collapse
- Should have used some other GANs like Wasserstein GAN
- Also, again not mentioned how lower and upper bounds are selected
- LapGANs

- If $a < L$: freeze propagation of confusion signal through the reconstructor, and train the discriminator.
- If $L \leq a < U$: alternate between propagating confusion signal and training the disciminator.
- If $U \leq a$: propagate confusion signal through the reconstructor, and freeze the discriminator.

In practice we used $L = 0.8, U = 0.95$. We compute the accuracy a as a running average over mini-batches with a momentum of 0.8.

#3 - Complete Pipeline Explanation Missing in Paper

- Final half is not explained in paper
- Simply mention that it is inverse of encoder (feature extraction).
- But encoder has skip connections, so decoder does not seem trivial
- A simple diagram would have been enough

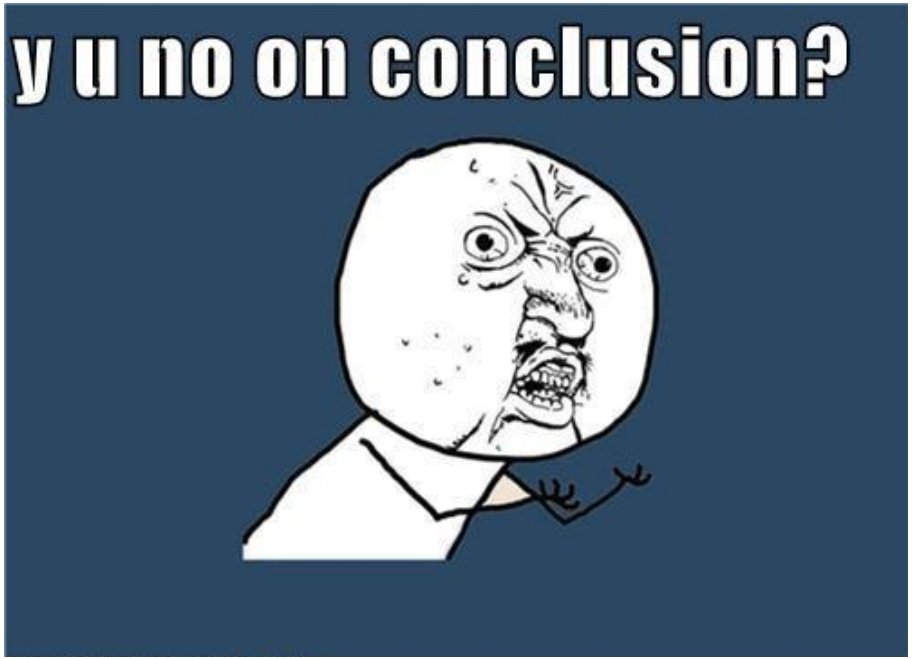


#4 - Fixed bit rate - Fixed model?

- The ACR targets a fixed bit rate during training
- This means that for different compression rates, different models are needed to be trained
- Not trained a universal model (like the first paper)

#5 - Not Structured

- No Conclusion / Discussion



Thank you