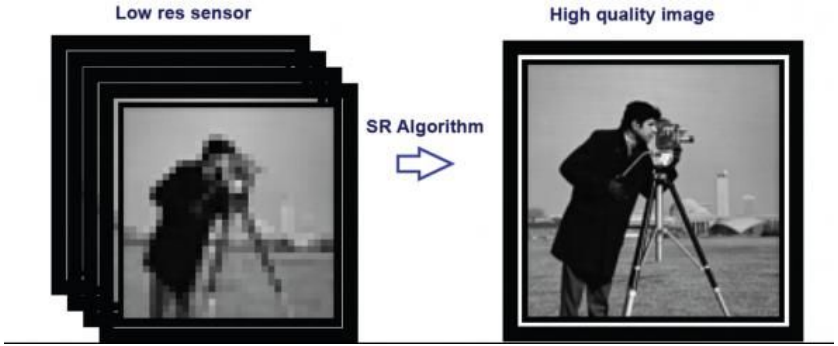
# Adversarial Techniques for SISR

Adam Vest

# Single Image Super Resolution

- Challenging problem → difficult to infer the values for multiple pixels from a single pixel
- Significant advances in the last two years due to deep learning methods
  - First deep learning systems used ConvNets, recent state-of-the-art uses GANs
- Goal is to find some mapping $f(I^{LR}) = I^{SR}$ where $I^{SR}$ is as close as possible to $I^{HR}$



# Adversarial Learning and SISR

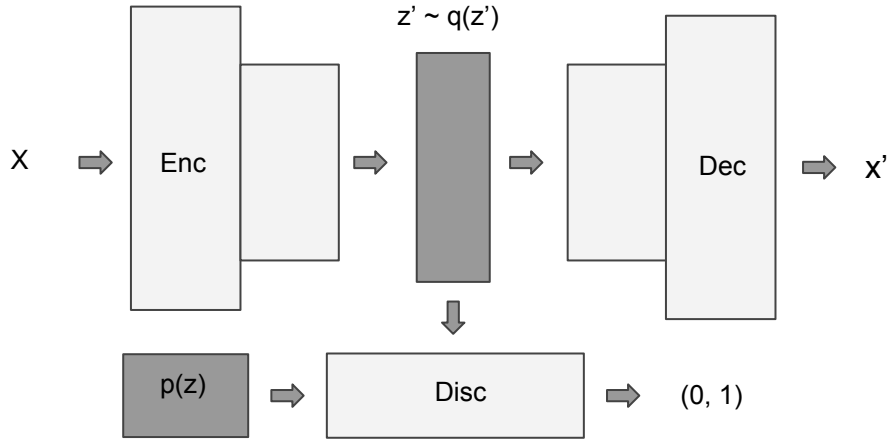
- Recent work has demonstrated the ability of adversarial learning methods to produce higher quality images than a simple ConvNet
- Potential Methods…
  - GANs
  - Adversarial Autoencoders (AAE)
  - Adversarially Learned Inference (ALI)
  - Adversarial Generator-Encoder Networks (AGE)

# GAN

- Utilized in “Photo-Realistic SISR Using a GAN,” a milestone paper from Twitter
- Process:
  - Generator accepts a LR image input and produces a SR image output
  - Discriminator attempts to determine if input image is a HR (real) or SR (fake) image
  - Note: generator no longer accepts random noise as in the original GAN formulation
- Decent PSNR, state-of-the-art Mean Opinion Score
- Adversarial loss can easily be augmented to include a content loss (e.g., MSE or a perceptual loss function) to improve the feasibility of reconstructions

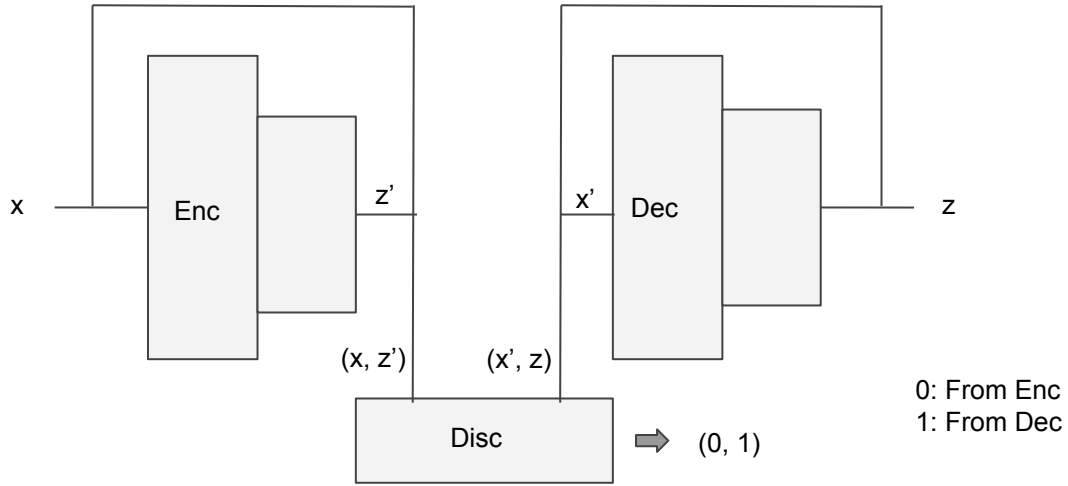
# AAE

- Architecture described in Makhzani et al. “Adversarial Autoencoders”
- Adversarial learning process forces Enc(x) = z’ ~ q(z’) to match some simple, chosen prior distribution for the p(z) for the latent space
- Personal Question: Why not give the LR images directly to the decoder?



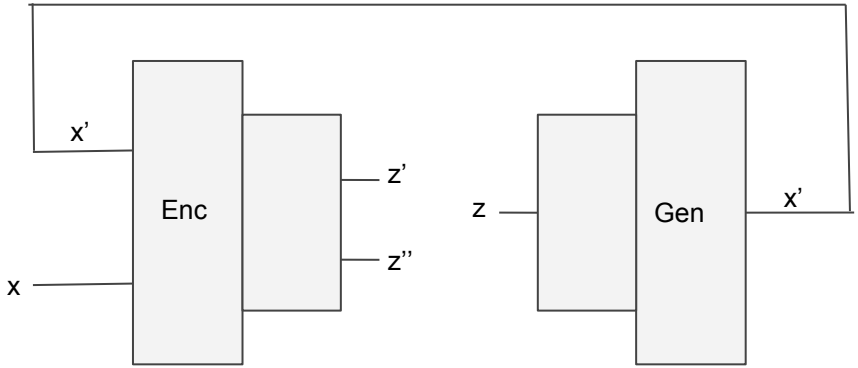
# ALI

- Model described in Dumoulin et. al “Adversarially Learned Inference”
- Allows for inference and generation by aligning the Dec(z) w/ the data distribution X and Enc(x) w/ the latent space distribution Z



# AGE

- Architecture described in Ulyanov et al. “Adversarial Generator-Encoder Networks”
- Adversarial game between the Encoder and Generator
- Generator tries to match Gen(z) to X by minimizing the divergence of z’ and z’’ in the latent space

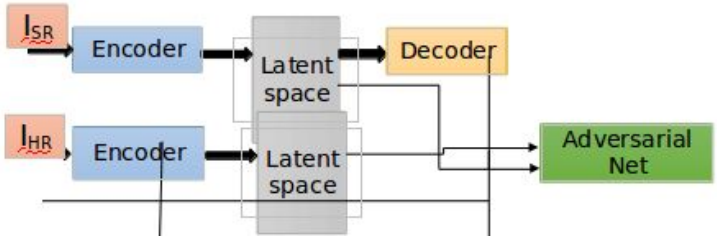


# Additional Considerations...

- Datasets?
  - Training: Most papers use ImageNet. Randomly crop the image several times, then downscale each crop (or sub-image) and use these to train
  - Testing: Set5, Set14, BSD100, Urban 100 are common, generally come w/ LR and HR images
- Loss functions?
  - In addition to adversarial loss, a “perceptual” or content loss function is often used in order to improve image reconstructions. VGG Loss used in Twitter GAN paper outperforms MSE in terms of image quality, but underperforms in terms of PSNR
- Evaluation Metrics?
  - PSNR, SSIM, MOS all good options, PSNR most common but may not be a good choice if we use a non-MSE content loss

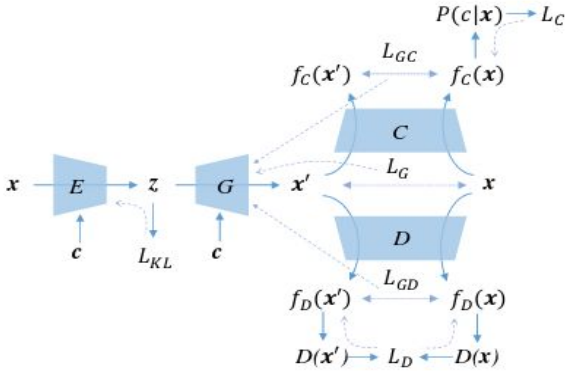
# Where Are We Now?

- Meeting this week to discuss different architectures, loss functions, evaluation metrics, etc.
- Current Proposal (From REU Project Proposal List):
  - Utilize some adversarial encoder-decoder architecture
  - Investigate MSE vs VGG loss
  - “We propose a new adversarial approach for the image super-resolution problem in which the adversarial network not only tries to match the low-resolution image and high-resolution image but also the latent variables of both sets”



# Conditioning?

- Would need to condition based on the LR image (not a one-hot encoded class vector, may be difficult)
- CGAN → need to investigate
- CVAE-GAN introduced 2 days ago:
  - Reconstructs images using information about its class -- excellent results



# Thank you! Questions?