Class-Conditional Superresolution with GANs

Vincent Chen, Liezl Puzon, Christina Wadsworth. ***Advisor: Ben Poole***

{vschen, puzon, cwads}@stanford.edu. benpoole@stanford.edu

Overview

- Super resolution enhances the resolution of a low-res image.
- Many deep learning models today work fairly well with an upscaling factor of 4x but use only the downscaled image as input.
- We demonstrate **two class-conditional GAN models** that outperform state-of-the-art GANs for superresolution:
  - Explicitly passes in a known class parameter as input
  - Uses a separately trained classifier to inform a ResNet about character attributes
- Both improve the quality of the super-resolved output.

Related Work

Generative Adversarial Networks (GAN)

In 2014, Goodfellow et. al [1] introduced the GAN framework for deep learning, in which a generative model competes with a discriminative adversary in a **two- player minimax game**:

$$\min_G \max_D(D,G) = E_{x\ p_{data}}[\log(D(x)] + E_{z\ p_z(z)}[\log(1-D(G(z)))]$$

- **Discriminative model** determines whether a sample is generated or a data example.
- **Generative model** attempts to fool the discriminative model.

GANs for Super-Resolution (SRGAN)

Ledig et. al [2] proposed a model with the capability to implement large **upscaling factors (4×)** with photo-realistic reconstructions of low-resolution images:

- **Feedforward network** as the generating function
- **Perceptual loss** as weighted combination of components
- Features a **deep residual network**

SRGAN (4X upscaling)

Original

SRGAN results are nearly indistinguishable compared to the original. [2]

Conditional GANs (CGAN)

- New frameworks introduce additional information to improve the performance of GANs.
- We model work on conditional GAN, introduced by Mirza et. al [3]:
  - Discriminator (*D*) and generator (*G*) models are conditioned on auxiliary information ***y***.

GANs + Auxiliary Class Information

We introduce two methods for implementing the conditional GAN framework.

1. Conditional GAN

We use a conditional predictive distribution for the discriminator and generator based on class information, ***y***:

$$\min_{\Theta_G} \max_{\Theta_D}(D,G) = E_{I^{HR}\sim p_{\text{train}}(I^{HR})}[\log(D_{\Theta_D}(I^{HR}|\mathbf{y})] + E_{I^{LR}\sim p_G(I^{LR})}[\log(1-D_{\Theta_G}(G_{\Theta_G}(I^{LR}|\mathbf{y})))]$$

2. GAN + Class Loss

As another method to introduce class information, we implemented a loss function that comprises GAN loss as well as an **added class loss term**.

- Chose gender (Male/Female)* as our conditional attribute for experiments
- Trained a **gender classifier** with **95.6% test accuracy**, and used this checkpointed model to compute class loss in the super-resolution model.
- <u>**Class loss**</u> = cross entropy loss between predicted classes of true image and generated image
- In effect, **penalize the generator for predicting an image that is not aligned with the original image's class** (gender, in this case)
- <u>**Hypothesis:**</u> Penalty causes the model to maintain some of the finer class details. For example, a woman with short hair should be more likely to have feminine features than a man with short hair.

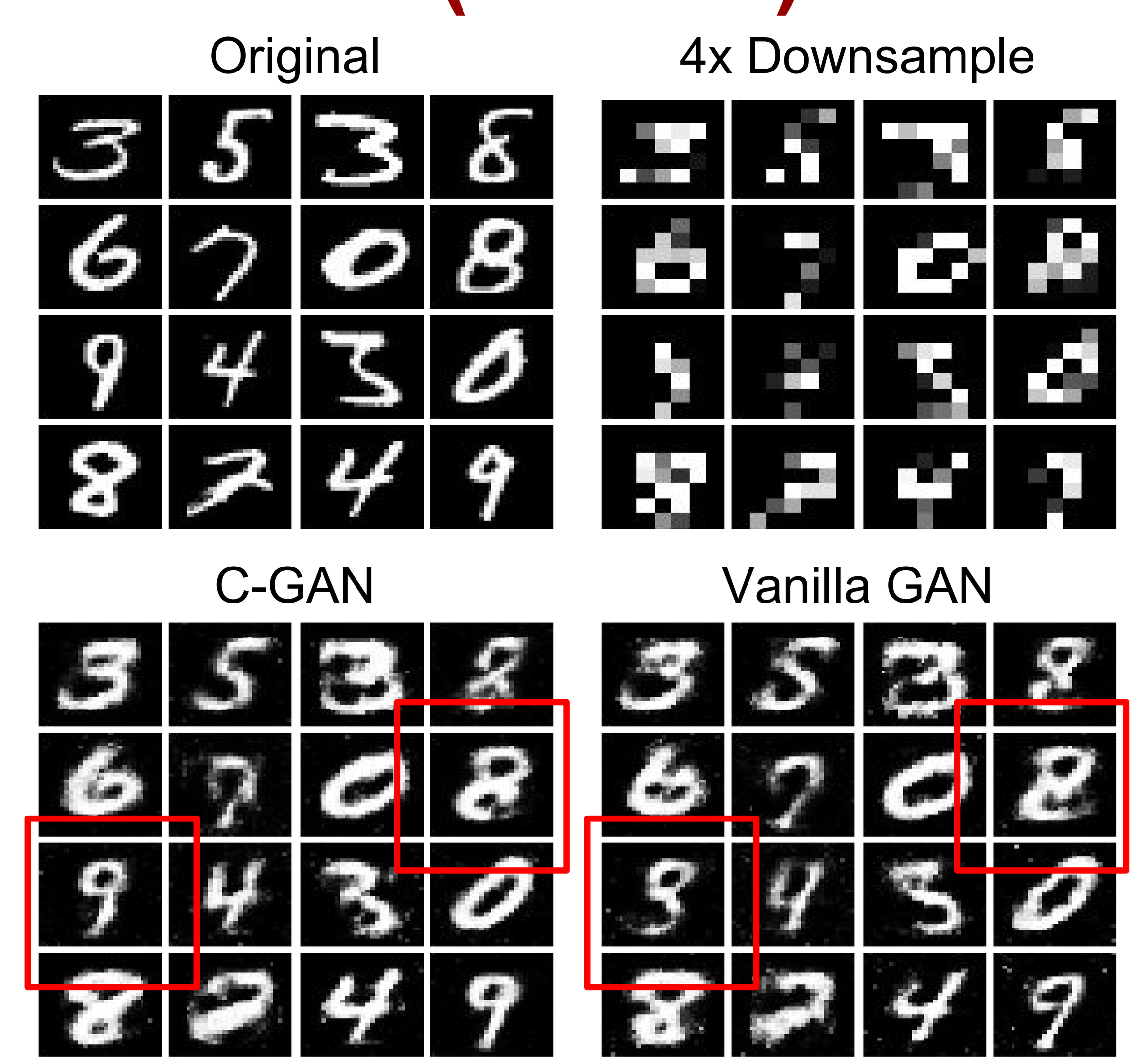
Conditional GAN (MNIST)

We trained the Conditional GAN model on the MNIST dataset.

- Concatenate class information, ***y***, to the inputs of feed-forward *G/D* models

Observations

- The Vanilla GAN produces artifacts that make certain digits appear ambiguous (red highlight)
- For same examples, the C-GAN produces results that are more recognizable as digits.

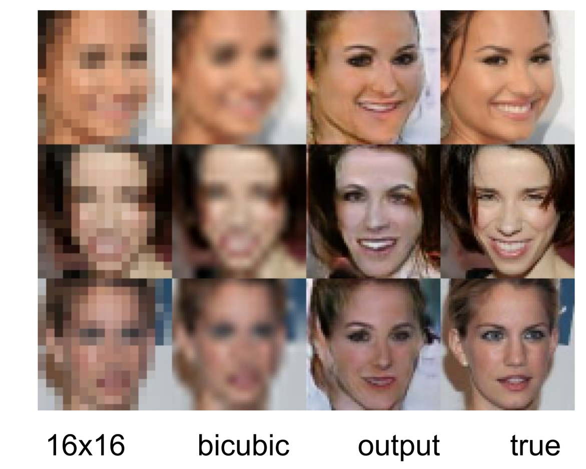


CGAN vs. Vanilla GAN for super-resolution on MNIST.

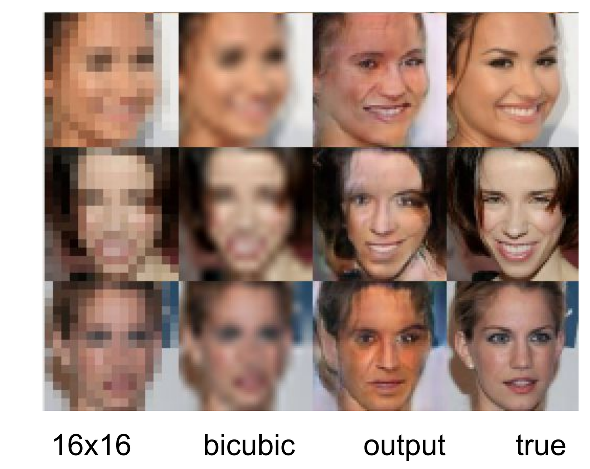
*Due to the pre-labeled CelebA dataset, there exists a dichotomy between faces labeled " masculine/feminine". We hope not to reinforce these conventions, and instead suggest that more thoughtful approaches must be taken to train real-world models based on subjective attributes.

GAN + Class Loss (CelebA)

We trained the GAN + class loss model with CelebA, and saw significant differences between our model and the plain GAN superresolution model.

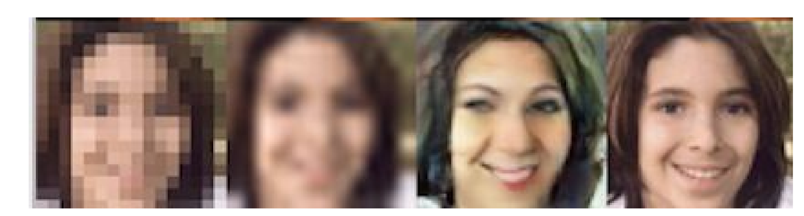


Our GAN + class loss model

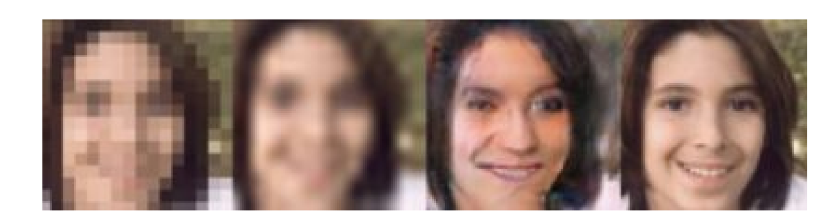


SRGAN output

- Generally, both model outputs are similar in overall quality.
- **When the original GAN generates an output image that has features from the opposite gender**, (see bottom image above).
- **Left**: GAN + class loss model produced image with feminine features
- **Right**: GAN model output resembles a face with male features
- Our image, on the left side, has much more feminine features and face structure. This is consistent with the true gender classification of the image: female.



Our GAN + class loss model



SRGAN output

Above we see a case in which our classifier was wrong.

- The (wrongly) **predicted gender of the true image is female**, and we see how that affected the generated image above.
- **Left**: GAN + class loss model produced an image with female features
- **Right**: GAN model output resembles the original image.
- Added classifier loss produces outputs that more resemble the attributes of the different genders it learned from the dataset. Luckily, our classifier only had 4.4% error rate, so these examples are rare.

Takeaways

- Introducing auxiliary information into GANs for super-resolution helps maintain true attributes of images.
- GANs are notoriously difficult to train. Introducing class loss omits the need to train entirely new models. Instead, we fine-tune existing, working models.
- Would be interesting to continue exploring different methods of introducing auxiliary information to models (i.e. InfoGAN [4])

References

[1] Goodfellow, I., et. al (2014). Generative Adversarial Nets.
[2] Letdig et. al (2016). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.
[3] Mirza, M. & Osindero, S (2014). Conditional Generative Adversarial Nets.
[4] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets.