BeliefGAN : Towards A More Informative Loss Function For Conditional GANs Using Approximate Inference

Varad Gunjal, Shubhakar Tipireddy

Carnegie Mellon University

{vgunjal, stipired}@andrew.cmu.edu

10-618
MACHINE LEARNING FOR STRUCTURED DATA

Introduction & Motivation

- GANs are a popular modern approach to generative modeling that learn the underlying structure of the data.
- They comprise of a Generator (G) & a Discriminator (D) network, engaged in a minimax game trying to optimize opposing loss functions.

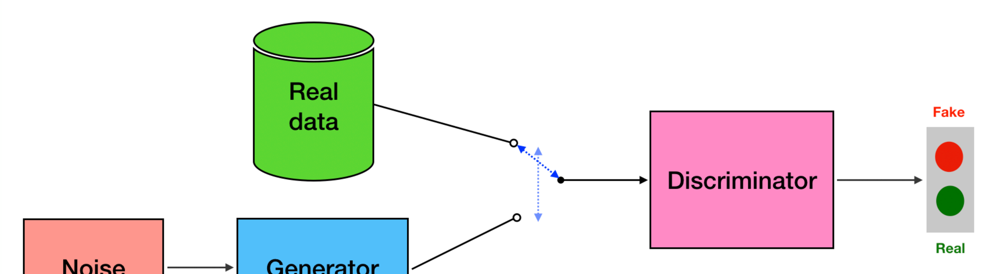

Fig: 1 General block architecture of a Generative Adversarial Network (GAN)

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

Goal: The objective is to enable generation of photo-realistic images with specific features - which are described by a set of text attributes and conform to artificially imposed relational constraints.

Hypothesis : Increasing the amount of information fed back in the loss function by evaluating each expected feature of the generated image individually enables the GAN to learn latent structure of the data faster, and it can generate images with desired visual features in fewer iterations with enhanced stability.

Related work

- **Conditional DCGAN** [4] showed that conditioning the GAN on auxiliary information about the data such as class labels could be used as an effective means to control features of the output data.
- In **Wasserstein GAN** [3], modifying the loss function of a GAN, by using Wasserstein (earth mover) distance, rather than JS divergence, drastically improves the stability of learning and helps address issues like mode collapse.
- These ideas motivate our approach to use auxiliary labels as input to the convolutional GANs as well as using them towards a **more informative feature-specific loss function** that enhances the feedback provided to the discriminator.

Dataset

- We run our experiments with the popular CelebA dataset - a large-scale face attributes dataset with more than **200K celebrity images**, each with annotations of **40 attributes**.
- We worked with a reduced set of 100k images displaying a wide variety of poses, diversity in features and can be asserted to constitute a representative sample of the dataset.

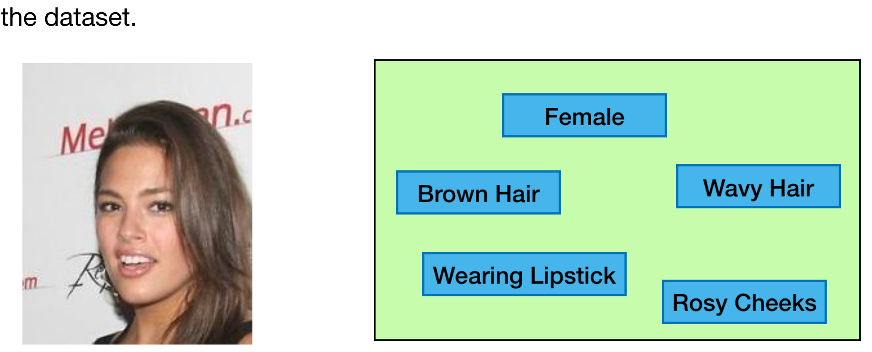

Fig: 2 (Left) Image of a celebrity from CelebA dataset; (Right) Attribute values that are marked 1 for the image, rest of features are marked -1

Methods

Models

The baseline model is a Conditional Deep Convolutional GAN (cDCGAN) which uses a binary feature vector of 40 attributes as additional input to the generator and discriminator.

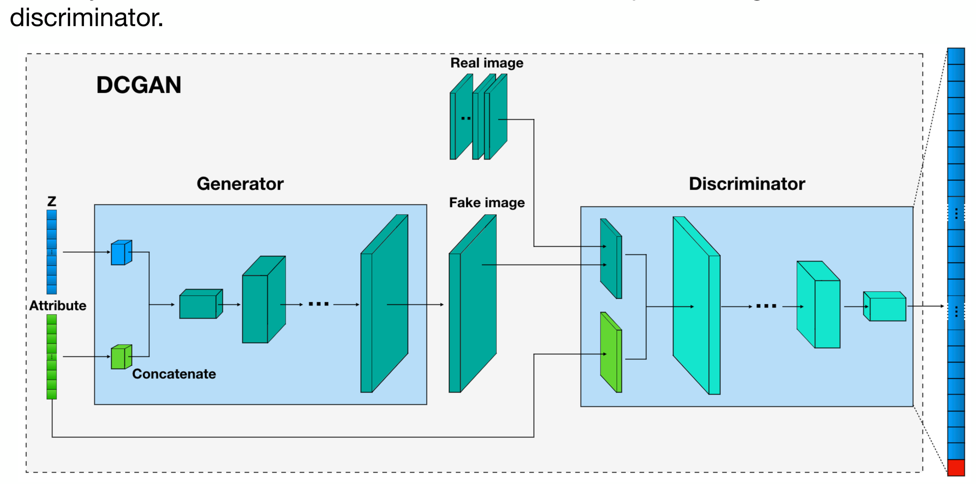

Fig: 3 Baseline architecture of conditional DCGAN + linear layer to predict the confidence of features + image classification

The final output of the cDCGAN is interpreted as a confidence value in features and of the image being real/fake by passing through a Sigmoid layer.

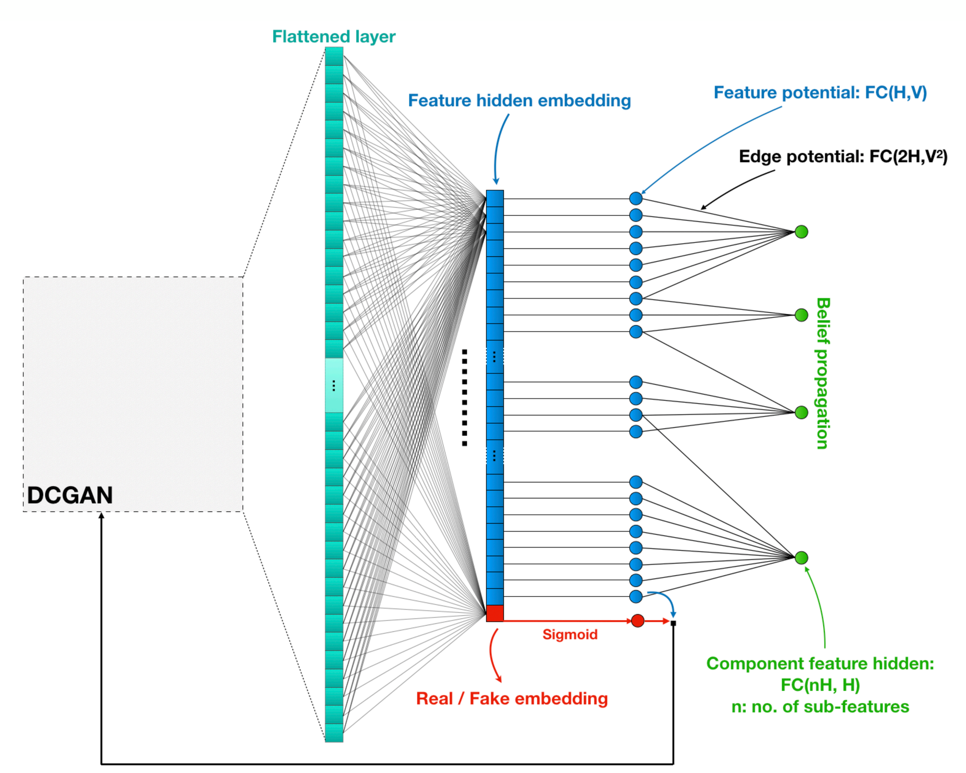

Fig: 4 Proposed architecture of BeliefGAB to estimate the confidence of features; Image classification is given by the FC layer + sigmoid

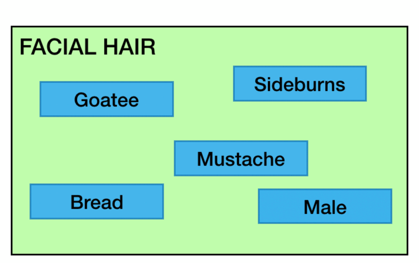

Fig: 5 Compound feature and sub-features associated with it

- Our model expands on the baseline by passing the output of the final convolutional layer through a FC layer.
- The elements of the output feature vector are interpreted as a **feature embedding** which are used to calculate **unary potentials -** the variable belief for a particular binary value of the feature at that index.

Algorithm 1 Forward-Backward Belief Propagation Algorithm

```
Trees = [Tree for each compound node]
N = number of compound nodes
for i from 1 to N do
  BeliefPropogation(Trees[i]) {All the features that are nodes of Trees[i] are updated}
end for
for i from N to 1 do
  BeliefPropogation(Trees[i])
end for
```

Fig: 6 Psuedocode for belief propagation in forward-backward fashion

- Every compound feature and its associations are manually constructed from prior knowledge about the attributes of the dataset.
- We run **belief propagation**, using a strategy similar to the **forward-backward** algorithm, to update the variable beliefs.
- **Ground truth** of each **compound feature** is calculated under the assumption that each of the **sub-features equally contribute** to the compound feature and vice versa.
- Loss is calculated using feature-wise **Binary Cross Entropy**.

Results

Evaluation Metrics

- We use 2 popular metrics for evaluating the performance of GANs - the Frechet Inception Distance **(FID)** and Inception Score **(IS)**.
- **Inception Score**
  - The Inception Score is a measure of 2 properties of the generator in a GAN:
    - Whether it **generates** images with **meaningful** objects.
    - Whether it **generates** images with **diverse content**.
  - **Higher is better**.
- **FID Score**
  - The FID is supposed to **improve on the IS** by actually *comparing* the statistics of generated samples to real samples, instead of evaluating generated samples in a vacuum.
  - **Lower is better.**

	FID	Inception Score
DCGAN	310.58	2.73
WGAN	426.66	1.02
Cond-DCGAN	302.81	2.18
BeliefGAN	222.25	2.63

Table: 1 Table indicating the evaluation metrics for each model after 20,000 mini-batch iterations

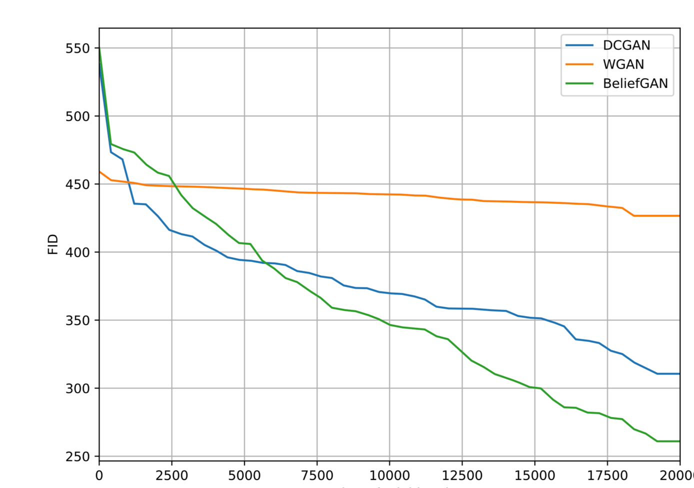

Fig: 6 FID score observed for each of the models after each mini-batch iteration

- DCGAN and BeliefGAN start off with similar FID scores but DCGAN improves quickly and performs better than BeliefGAN till ~6000 iterations.
- Eventually BeliefGAN exhibits better performance among all three models.
- Results from Cond-DCGAN were close to DCGAN and inconclusive and hence were omitted.

- DCGAN and BeliefGAN start off with similar inception scores and continue to exhibit similar performance.
- Performance evaluations based on inception score were inconclusive.
- WGAN started off similar to other models but never showed improvement.

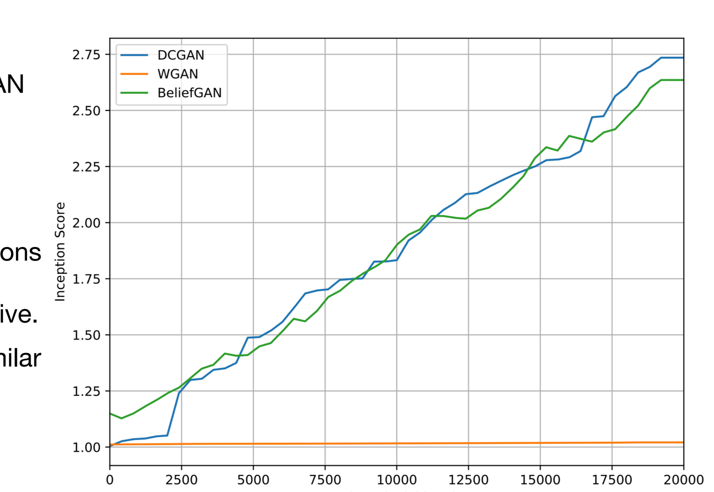

Fig: 7 Inception score observed for each of the models after each mini-batch iteration

Fig: 8 Generated images for tags "WEARING_MAKEUP", "BLONDE_HAIR"

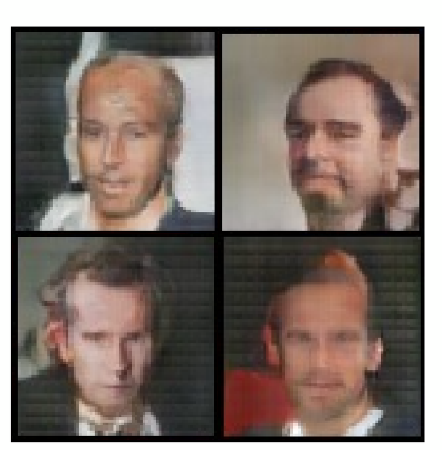

Fig: 9 Generated images for tags "RECEDING_HAIRLINE", "STUBBLE"

Discussion

- The effect of the relational factors can be clearly observed in the gender of images generated by conditioning on obviously gender-specific attributes.
- Although we observed a general trend towards more performance improvement per epoch for the BeliefGAN architecture, this comes at the expense of significantly longer (~3-4x) training time when compared to existing architectures like DCGAN.
- One important issue with the current implementation is the lack of interpretability of the expected facial features in the image by visual inspection. This is due to inherent limitations of the DCGAN architecture which result in poor output resolution.
- Over the course of this work, we also explored some other implementation details within our approach -
  - **Using LDA to learn correlated features from data** : Rather than using a manually defined relation between attributes on the dataset, we attempted to use LDA to identify clusters of correlated features from the data.
  - **Parallelized BP with barrier synchronization** : To improve the training time of the algorithm, we implemented a basic parallel version of belief propagation that simultaneously processes messages and belief updates within the same level of the tree.

Conclusion

- In this project, we showed how augmenting the loss function to reflect feature-specific information could potentially help Conditional GANs train faster and express attributes that the data is conditioned upon.
- We observe that our modifications enable expressing control over generation of specific attributes in the output, by trading off for training time.
- Future Work
  - One obvious direction to explore is evaluation of the described strategy of feature-specific loss, on modern architectures like StyleGAN, which have a proven ability to generate higher resolution images.
  - Another promising direction to pursue would be to use the belief propagation strategy with individual feature-based loss computed with the Wasserstein distance rather than BCE.
- Where can this be used?
  - This approach to augmentation of the loss function is a generic improvement strategy that can be adapted to any scenario where correlated attributes of data are available.

References

[1] Goodfellow, et al. "Generative Adversarial Networks." *ArXiv.org*, 10 June 2014, https://arxiv.org/abs/1406.2661.

[2] Radford, et al. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." *ArXiv.org*, 7 Jan. 2016, https://arxiv.org/abs/1511.06434.

[3] Arjovsky, et al. "Wasserstein GAN." *ArXiv.org*, 6 Dec. 2017, https://arxiv.org/abs/1701.07875.

[4] Mirza, et al. "Conditional Generative Adversarial Nets." *ArXiv.org*, 6 Nov. 2014, https://arxiv.org/abs/1411.1784.