Term Paper Problem Focus

Cameron Fabbri
Computer Vision
2/20/2017

Our project's end goal is to evaluate and compare to several methods for automatically colorizing images. We plan to implement multiple deep and non-deep learning algorithms for this task. Currently, the state of the art for automatic colorization is a deep learning approach by [8], which uses a deep convolutional neural network to generate plausible colorized images. Generative Adversarial Networks (GANs) [1], have been shown to outperform previous probabilistic models for image generation. I will be focusing on the design and implementation of multiple adversarial-based networks applied to the task of colorization. Because of the instability and difficulty in training GANs, to be able to properly use them as an application for colorization, I will also be focusing on the more classic study of GANs for generating images given a noise prior. Although not directly applied to our project, it is important to understand GANs in a general sense in order to work towards colorization (e.g be able to generate images first, then apply towards colorization). I am already familiar with the original GANs paper [1], so that will inherently be one of the papers I will be focusing on.

The second paper I will be focusing on is "Towards Principled Methods For Training Generative Adversarial Networks" [2], which provides theoretical steps towards fully understanding the training dynamics of GANs. It is fairly well known that GANs are very difficult to train properly, and can be very unstable. [2] points out that in theory the discriminator should be trained as close as possible to optimality, and then train the generator. However, in practice this does not work. The main contributions of this paper are the answers to the following questions: Why do updates get worse as the discriminator gets better? Why is GAN training massively unstable? Is the new cost function following a similar divergence to the Jensen-Shannon Divergence (JSD)? Is there a way to avoid these issues? Finally, they analyze theoretical differences in using the traditional (towards GANs) JSD cost function as opposed to the more traditional generative modeling cost function, the Kullback-Leibler divergence.

The "Wasserstein GAN" [3] follows nicely by comparing multiple distances between probability distributions. They cover the Total Variation (TV) distance, the Kullback-Leibler (KL) divergence, the Jensen-Shannon (JS) divergence, and the Earth-Mover (EM) distance or Wasserstein. This leads to a practical approximation of optimizing the EM distance, and how it can be applied to GANs, ending with their Wasserstein GAN (WGAN) algorithm. Unlike many previous GAN architectures, WGAN shows to offer stable training, as well as a meaningful loss metric correlating with the generator's convergence and sample quality. The model they use is very similar to the Deep Convolutional Generative Adversarial Networks [6] (DCGAN) model. This has shown to be a stable model in training GANs with deep architectures.

"Energy-Based Generative Adversarial Networks" (EBGANs) [4] treat the discriminator as an energy function that is trained to assign high energies to samples generated by the generator. This allows for a wider variety of architectures and loss functions in addition to the binary classifier as used with normal GANs. They bridge two classes of unsupervised learning models, namely GANs and auto-encoders, to generate higher resolution images than regular GANs while offering a more stable training behavior. For these reasons, I will also be focusing on this as a possible route for colorization.

The majority of research done in GANs has been towards either generating images from noise, or editing images conditioned on a set of attributes. "Image-to-Image Translation with Conditional Adversarial Networks" [5] investigates the use of conditional adversarial networks as a mapping between images (e.g synthesizing real images from drawings). One such task is colorizing a black and white image. One important point they note is that image-to-image translation problems map a high resolution input grid to a high resolution output grid. Solutions to this difficult task, especially for image colorization, incorporate skip connections into the network to allow information to flow through all of the layers of the network. This is done by concatenating the channels at layer i with those from layer $n - i$. The architecture used is also the DCGAN model.

The papers I will be focusing on [1,2,3,4,5,6] all address the task of generating images. I will be leveraging the work done towards stabilizing the training of GANs, as well as investigate methods shown in [5] for image colorization. I will also be referencing [7], which directly tackles skip connections to allow the flow of information. With these I will implement a stable adversarial network for the task of colorizing images. I will also be comparing the results of colorization across the three main vanilla architectures, DCGANs, WGANs, and EBGANs, as a way to assess our results.

References

[1] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.

[2] Arjovsky, Martin, and Lon Bottou. "Towards principled methods for training generative adversarial networks." NIPS 2016 Workshop on Adversarial Training. In review for ICLR. Vol. 2016. 2017.

[3] Arjovsky, Martin, Soumith Chintala, and Lon Bottou. "Wasserstein GAN." arXiv preprint arXiv:1701.07875 (2017).

[4] J. Zhao, M. Mathieu, and Y. LeCun. Energy-based Generative Adversarial Network. ArXiv e-prints, September 2016.

[5] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." arXiv preprint arXiv:1611.07004 (2016).

[6] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

[7] Huang, Gao, et al. "Densely connected convolutional networks." arXiv preprint arXiv:1608.06993 (2016).

[8] Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." European Conference on Computer Vision. Springer International Publishing, 2016.