Thesis Work Special – Top 7 Research Papers

1. AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks
2. Photographic Text-to-Image Synthesis with a Hierarchically-nested Adversarial Network
3. StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks
4. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks
5. Generative Adversarial Nets
6. StoryGAN: A Sequential Conditional GAN for Story Visualization
7. Language-Based Image Editing with Recurrent Attentive Models

1. AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks

Tao Xu 1, Pengchuan Zhang 2, Qiuyuan Huang 2, Han Zhang 3, Zhe Gan 4, Xiaolei Huang 1, Xiaodong He 5

Abstract - In this paper, we propose an Attentional Generative Adversarial Network (AttnGAN) that allows attention-driven, multi-stage refinement for fine-grained text-to-image generation. With a novel attentional generative network, the AttnGAN can synthesize fine-grained details at different sub-regions of the image by paying attentions to the relevant words in the natural language description. In addition, a deep attentional multimodal similarity model is proposed to compute a fine-grained image-text matching loss for training the generator. The proposed AttnGAN significantly out-performs the previous state of the art, boosting the best reported inception score by 14.14% on the CUB dataset and 170.25% on the more challenging COCO dataset. A detailed analysis is also performed by visualizing the attention layers of the AttnGAN. It for the first time shows that the layered attentional GAN is able to automatically select the condition at the word level for generating different parts of the image.

2. Photographic Text-to-Image Synthesis with a Hierarchically-nested Adversarial Network

Zizhao Zhang ∗, Yuanpu Xie ∗, Lin Yang

Abstract - This paper presents a novel method to deal with the challenging task of generating photographic images conditioned on semantic image descriptions. Our method introduces accompanying hierarchical-nested adversarial objectives inside the network hierarchies, which regularize mid-level representations and assist generator training to capture the complex image statistics. We present an extensile single-stream generator architecture to better adapt the jointed discriminators and push generated images up to high resolutions. We adopt a multi-purpose adversarial loss to encourage more effective image and text information usage in order to improve the semantic consistency and image fidelity simultaneously. Furthermore, we introduce a new visual-semantic similarity measure to evaluate the semantic consistency of generated images. With extensive experimental validation on three public datasets, our method significantly improves previous state of the arts on all datasets over different evaluation metrics.

3. StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks

Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Senior Member, IEEE, Xiaogang Wang, Member, IEEE, Xiaolei Huang, Member, IEEE, Dimitris N. Metaxas∗, Fellow, IEEE

Abstract - Although Generative Adversarial Networks (GANs) have shown remarkable success in various tasks, they still face challenges in generating high quality images. In this paper, we propose Stacked Generative Adversarial Networks (StackGANs) aimed at generating high-resolution photo-realistic images. First, we propose a two-stage generative adversarial network architecture, StackGAN-v1, for text-to-image synthesis. The Stage-I GAN sketches the primitive shape and colors of a scene based on a given text description, yielding low-resolution images. The Stage-II GAN takes Stage-I results and the text description as inputs, and generates high-resolution images with photo-realistic details. Second, an advanced multi-stage generative adversarial network architecture, StackGAN-v2, is proposed for both conditional and unconditional generative tasks. Our StackGAN-v2 consists of multiple generators and multiple discriminators arranged in a tree-like structure; images at multiple scales corresponding to the same scene are generated from different branches of the tree. StackGAN-v2 shows more stable training behavior than StackGAN-v1 by jointly approximating multiple distributions. Extensive experiments demonstrate that the proposed stacked generative adversarial networks significantly outperform other state-of-the-art methods in generating photo-realistic images.

4, 7. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Jun-Yan Zhu∗, Taesung Park∗, Phillip Isola, Alexei A. Efros

Abstract – Image-to-image translation is a class of vision and graphics problems where the goal is to learn the mapping between an input image and an output image using a training set of aligned image pairs. However, for many tasks, paired training data will not be available. We present an approach for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. Our goal is to learn a mapping $G : X \rightarrow Y$ such that the distribution of images from G(X) is indistinguishable from the distribution Y using an adversarial loss. Because this mapping is highly under-constrained, we couple it with an inverse mapping $F : Y \rightarrow X$ and introduce a cycle consistency loss to enforce $F(G(X)) \approx X$ (and vice versa). Qualitative results are presented on several tasks where paired training data does not exist, including collection style transfer, object transfiguration, season transfer, photo enhancement, etc. Quantitative comparisons against several prior methods demonstrate the superiority of our approach.

5, 12. Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie∗, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair†, Aaron Courville, Yoshua Bengio‡

Abstract – We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player

game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to ½ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

6, 14. StoryGAN: A Sequential Conditional GAN for Story Visualization

Yitong Li∗1, Zhe Gan 2, Yelong Shen 4, Jingjing Liu 2, Yu Cheng 2, Yuexin Wu 5, Lawrence Carin 1, David Carlson 1 and Jianfeng Gao 3

Abstract – In this work we propose a new task called Story Visualization. Given a multi-sentence paragraph, the story is visualized by generating a sequence of images, one for each sentence. In contrast to video generation, story visualization focuses less on the continuity in generated images (frames), but more on the global consistency across dynamic scenes and characters – a challenge that has not been addressed by any single-image or video generation methods. Therefore, we propose a new story-to-image-sequence generation model, StoryGAN, based on the sequential conditional GAN framework. Our model is unique in that it consists of a deep Context Encoder that dynamically tracks the story flow, and two discriminators at the story and image levels, respectively, to enhance the image quality and the consistency of the generated sequences. To evaluate the model, we modified existing datasets to create the CLEVR-SV and Pororo-SV datasets. Empirically, StoryGAN outperformed state-of-the-art models in image quality, contextual consistency metrics, and human evaluation.

7, 21. Language-Based Image Editing with Recurrent Attentive Models

Jianbo Chen∗, Yelong Shen†, Jianfeng Gao†, Jingjing Liu†, Xiaodong Liu†

Abstract – We investigate the problem of Language-Based Image Editing (LBIE). Given a source image and a natural language description, we want to generate a target image by editing the source image based on the description. We propose a generic modeling framework for two sub-tasks of LBIE: language-based image segmentation and image colorization. The framework uses recurrent attentive models to fuse image and language features. Instead of using a fixed step size, we introduce for each region of the image a termination gate to dynamically determine after each inference step whether to continue extrapolating additional information from the textual description. The effectiveness of the framework is validated on three datasets. First, we introduce a synthetic dataset, called CoSaL, to evaluate the end-to-end performance of our LBIE system. Second, we show that the framework leads to state-of-the-art performance on image segmentation on the ReferIt dataset. Third, we present the first language-based colorization result on the Oxford-102 Flowers dataset.