

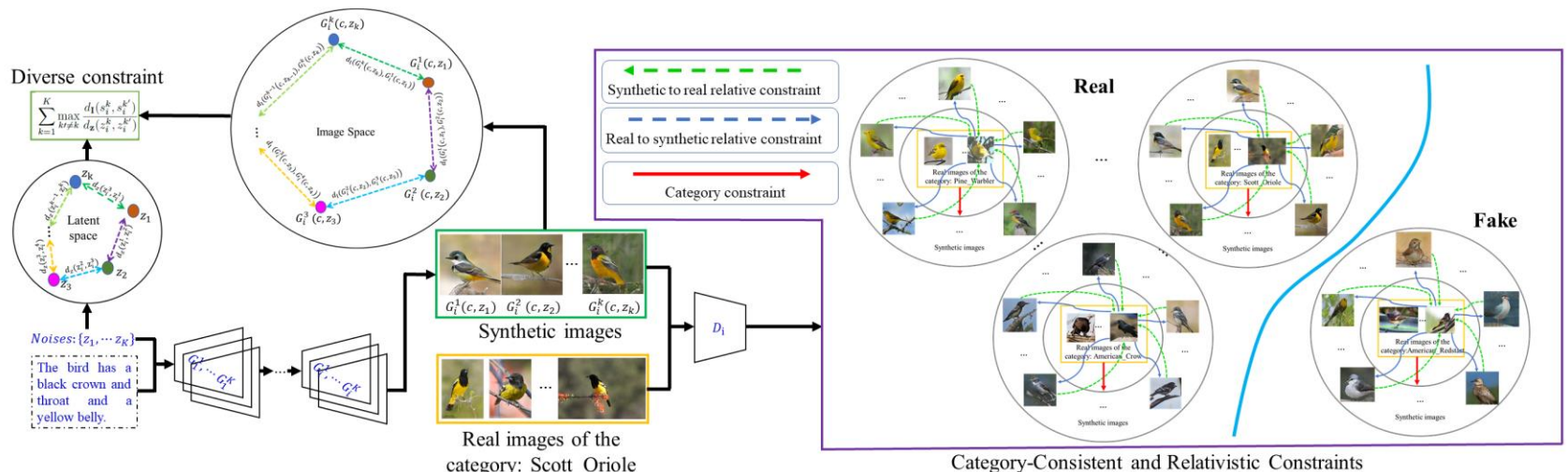
CRD-CGAN: Category-Consistent and Relativistic Constraints for Diverse Text-to-Image Generation

Tao HU, Chengjiang LONG, Chunxia XIAO

Frontiers of Computer Science, DOI: [10.1007/s11704-022-2385-x](https://doi.org/10.1007/s11704-022-2385-x)

Problems & Ideas

- Problems of Text-to-image generation based on GANs:
 - It is still challenging to simultaneously generate a set of photo-realistic as well as significantly diverse synthetic images conditional on text description.
 - Existing text-to-image GANs ignore the category attributes of the real image corresponding to the text description.
- Ideas: we focus on the category-consistent and relativistic diverse constraints to optimize the diversity of synthetic images.



Overview of our proposed framework CRD-CGAN. We use the K generators to generate K synthetic images. Based on the difference between image space and latent noise space, we use the diverse constraints to improve the diversity of K synthetic images. The proposed category-consistent and relativistic constraints are reflected in the discriminator at each stage.

Main Contributions

- Contributions:
 - We propose a new framework CRD-CGAN which contains K generators and a shared discriminator to improve the diversity of K high-quality synthetic images simultaneously.
 - We incorporate category and relativistic diverse conditional constraints, which effectively improve the diversity of synthetic images with the main visual feature of the corresponding of real images.
 - The proposed CRD-CGAN achieves the state-of-the-art performance on the Caltech-UCSD Birds-200-2011, the Oxford 102 Category Flower and the MS COCO 2014 datasets for text-to-image generation..

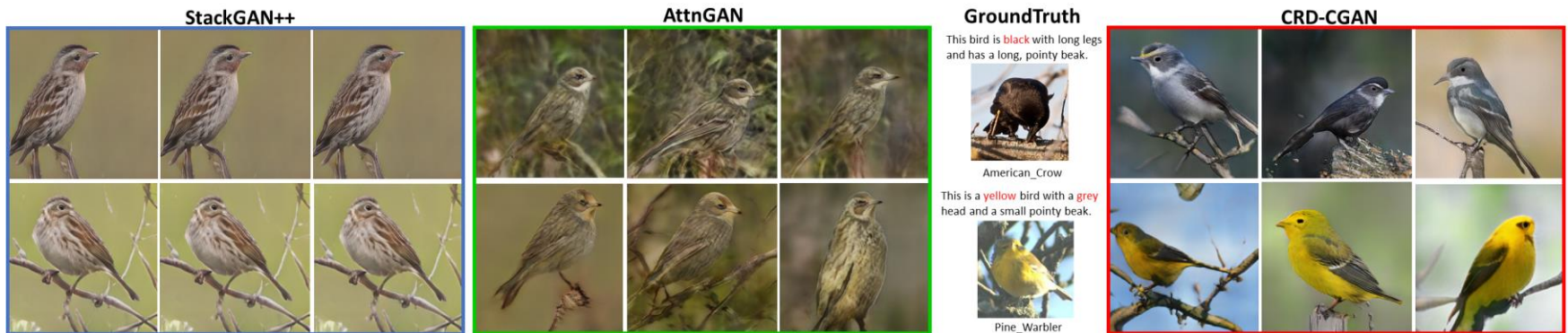


Illustration of three methods to generate synthetic images conditional on a text. There is a straightforward visual difference between synthetic and real images in StackGAN++. The synthetic images generated by AttnGAN also have not consistent of color with the real image. The synthetic images generated by our CRD-CGAN are as consistent as possible with the text description and consistent with the category visual feature of the real image category.