Text-To-Image Synthesis Using KTGAN

[1]Kavitha Chaduvula*, [2]Baburao Markapudi, [3]N. L. Jahnavi, [4]V. Manoj, [5]S. Sirisha, [6]S. K. Azeez
Seshadri Rao Gudlavalleru Engineering College,
Gudlavalleru-521356, Krishna District, AP, India.
Correspondingauthor: kavithachaduvula12@gmail.com

Article Info
Page Number: 862 – 869
Publication Issue:
Vol. 71 No. 3 (2022)

Article History
Article Received: *12 January 2022*
Revised: *25 February 2022*
Accepted: *20 April 2022*
Publication: *09 June 2022*

Abstract
This paper proposes a new model, Knowledge Transfer Generative Adversarial Network (KTGAN), for text-to-image synthesis. A couple of new methods are used such as an Alternate Attention Transfer Mechanism (AATM) and a Semantic Distillation Mechanism (SDM), to assist generator higher connect inter functional hole in linking textual content and picture. AATM modifies phrase interest weights and interest weights of picture segment one after other, to gradually spotlight necessary phrase statistics and enrich small print of generated images. The semantic distillation mechanism makes use of picture encoder skilled in the Image-to-Image project to information coaching of textual content encoder in Text-to-Image process, to produce higher textual content points and greater excellent pictures. By significant investigational testing on two public data sets, KTGAN surpass the existing approach notably, and additionally attains the comparative effects over one-of-a-kind comparison metrics.

Keywords:- KTGAN, text-to-speech synthesis, AATM, SDM, interest weights

Introduction

Text-to-Image generation objectives to create practical picture which is significantly constant with a given textual content, with the aid of studying a bounding between the semantic textual content area (M. Bharti et al., 2019) and complicated RGB picture area. A key undertaking in generating sensible objects with significant small print is the assorted hole between powerful ideas in textual content illustration and component level contents of artificial pictures (A. RoyChowdhury et al., 2019). Developing a wonderful generator to fill the area hole is tough. Large no of processes based totally on Generative Adversarial Networks (GANs) fill the area hole by means of making use of a discriminator to compare generated pair and real pair. Anyhow, such a discriminator on my own is normally inadequate to mannequin hidden semantic persistency between textual content and photo, and accordingly, consequences in structural blunders in generated pix (view Fig 1, the "Direct T2I" row). Currently, the interest method has been used to solve this problem and this courses the generator to higher in shape positive visible phrases with corresponding picture sub regions. But the usage of word-level interest by myself does now not make sure world semantic persistency because of the variety between the textual content and photograph methods. Hence, the MirrorGAN fashions Text2Image and Image2Text collectively to beautify world inter functional semantic persistency. However, the Image2Text in MirrorGAN is nonetheless a inter functional creation, which is now not simpler than comparable technology assignment such as Image to Image function. Hence, the hassle of semantic disparity in middle of diverse records nonetheless be same. SEGAN develops a

novel contrastive loss and a Semantic Consistency Module (SCM) to higher line up generated photo and floor reality in function space.
Nevertheless because of the diversified semantic disparity, SEGAN can't pull out fine textual content aspects that can information the synthesis of practical and targeted images.

Literature Review

The goal of this research is to unsupervised adapt an present object detector to a novel target domain. ourself think that there are numerous unlabeled videos in this field. We get labels on the target data automatically through combining more confidence detections from current detector with rigid examples obtained by leveraging temporal cues with a tracker. These labels are then used to retrain real model usingthe automatically generated labels. We propose a updated knowledge distillation loss and enquire various methods for alloting soft-labels to target domain training examples (A. RoyChowdhury et al., 2019).Our method has been empirically tested on difficult face and pedestrian detection works: a face detector trained on wider-Face, which has more efficient pictures dragged from the net, is altered to more amount of scrutiny data set; a pedestrian detector trained on transparent, dayhours pictures from the BDD-100K driving data set is altered to remaining scenes, which includes showery, misty, and dark. These findings show importance of using real-world examples gleaned via tracing, the benefit of using soft-labels via distillation loss vs hard-labels, and also promising efficiency as a simple process for unsupervised domain adaptation of object detectors with little reliance on hyper parameters.
We show how to execute an end2end person search using knowledge distillation. In person search, end2end approaches are present state of the art. because they determine both detection and re-identification problems at the same time. Because of a poor detector, these joint optimization approaches exhibit the greatest drop in performance. In a teacher-student scenario, we offer two separate methods for further control of end2end person search strategies. The first one is based on cutting-edge object detection knowledge distillation. We utilise this to use a specialised detector to oversee the detector of our person search model at multiple stages. The next approach is fresh, simple, and significantly better efficient. Using look-up table of ID attributes which is computed previously, this collects information from a teacher re-identification process (M. Bharti et al., 2019). It allows the student to relax while learning identification traits, allowing him or her to have eye on detection assignment. This technique not just supports in correction of improper detector training in combined optimization while also enhancing search for people. However, in this scenario, model compression reduces the performance difference between the teacher and the student. On two benchmark data sets, We show that two current state of art approaches can be significantly updated utilising our developed knowledge distillation approach. Furthermore, our application compares the accomplishment of minor and major models in the model compression challenge.
We introduce a new controlled text2image generative adversarial network and that can efficiently generate excellent standard images while also controlling components of the image creation based on natural language descriptions. We offer a spatial and channel-wise attention-driven generator that can untangle several visual properties at the word level and allowing the model to concentrate on creating and modifying subregions for the most important terms. By connecting words with picture regions, a word-level discriminator is also presented to provide fine grained controllable report (L. Bowen et al., 2019). allowing for the guiding of an efficient generator capable of manipulating particular features while not disturbing the creation of additional content. In addition to that, perceptual loss is used to decrease the uncertainity in image creation and to inspire the generator to alter certain properties needed in the updated text. Large scale testing on standard data sets show that our

model exceed the current technology and can manage fabricated images effectively making use of natural language descriptions.

Proposed System

Within this paper author is using GAN model (generative Adversarial Network) to convert text to images. In propose paper author modifying GAN with transfer learning to accommodate text with images so generator model get trained on TEXT and discriminator model get trained on images with embed text and when we give any text then generator GAN model will predict equivalent image for given text.

In propose paper Alternate Attention Transfer Mechanism (AATM) and Semantic Distillation Mechanism (SDM), in assisting the generator in reducing inter functional chasm separating text and picture. The AATM alternately modifies the attention weights of phrases and Attention weights for image divisions to constantly focus useful phrase information and upgrade the quality of generated pictures. The SDM takes help of image encoder which is trained in the image2image function to assist the training of text encoder which is used in the text2image function for developing more exceptional images.

Methodology

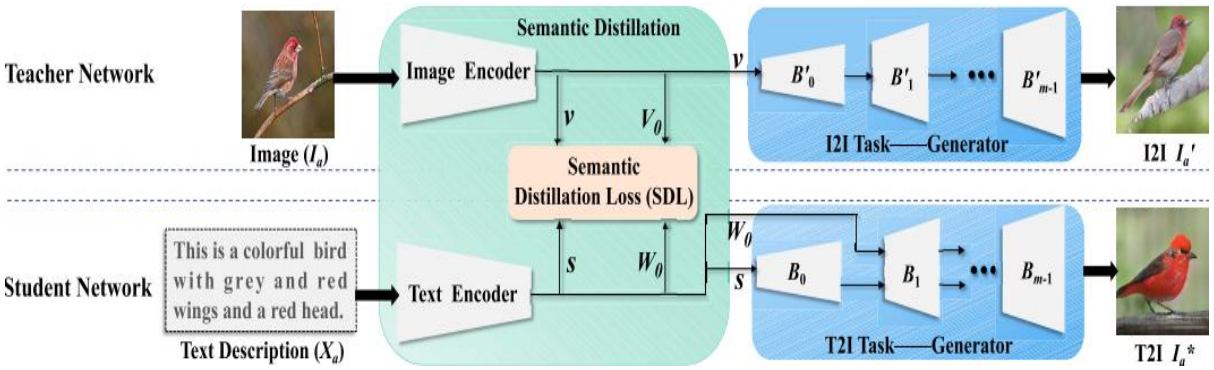


Fig 1: Work Flow

We suggest fresh Knowledge Transfer Generative Adversarial Network (KTGAN) with two new mechanisms for Text to Image synthesis: a Semantic Distillation Mechanism which includes usage of image encoder to guide text encoder for better image quality. An Alternate Attention Transfer Mechanism to discover necessary phrases in text.

The following are the paper's main contributions:

(1) We developed a Semantic Distillation Mechanism which has new Semantic distillation loss function (SDL) and using this we are able to guide the text to image task using image to image task for better results.

(2) We introduced an Alternate Attention Transfer Mechanism to frequently update the attention weights and increase the image quality.

(3) We tested KTGAN on couple of different data sets CUB-Bird and large scale MS-COCO. Experimental effects and study shows the importance of KT-GAN and extensively expanded overall effect in contrast towards preceding modern strategies on all 4 comparison metrics.

We created the following modules to help us carry out this project.

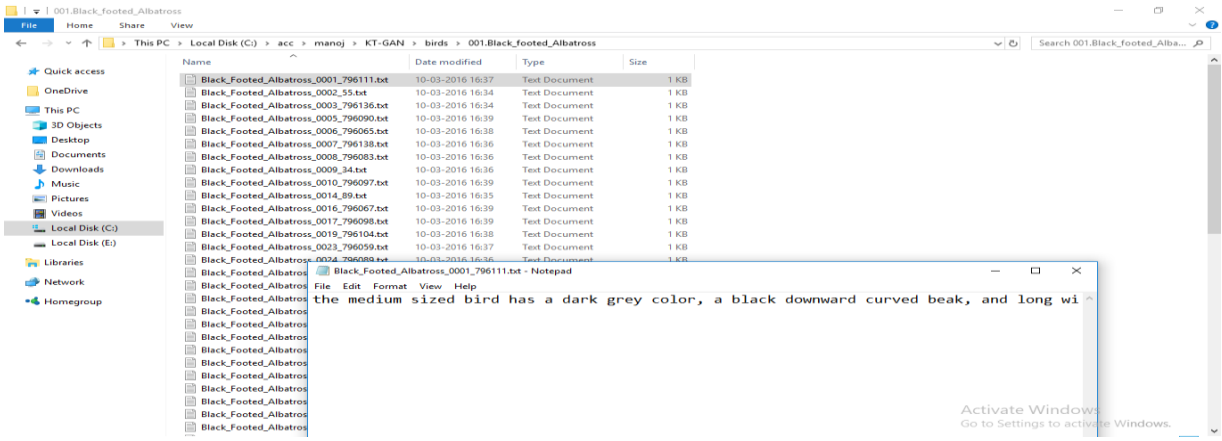
1) Upload CUB-Bird Data set: using this module we will upload dataset folders to application
2) Generate & Load KT-GAN Model: using this module we will read all images and then generate KT-GAN model

3) Generate Image from Text: using this module we will input TEXT and then KT-GAN will generate image from that text

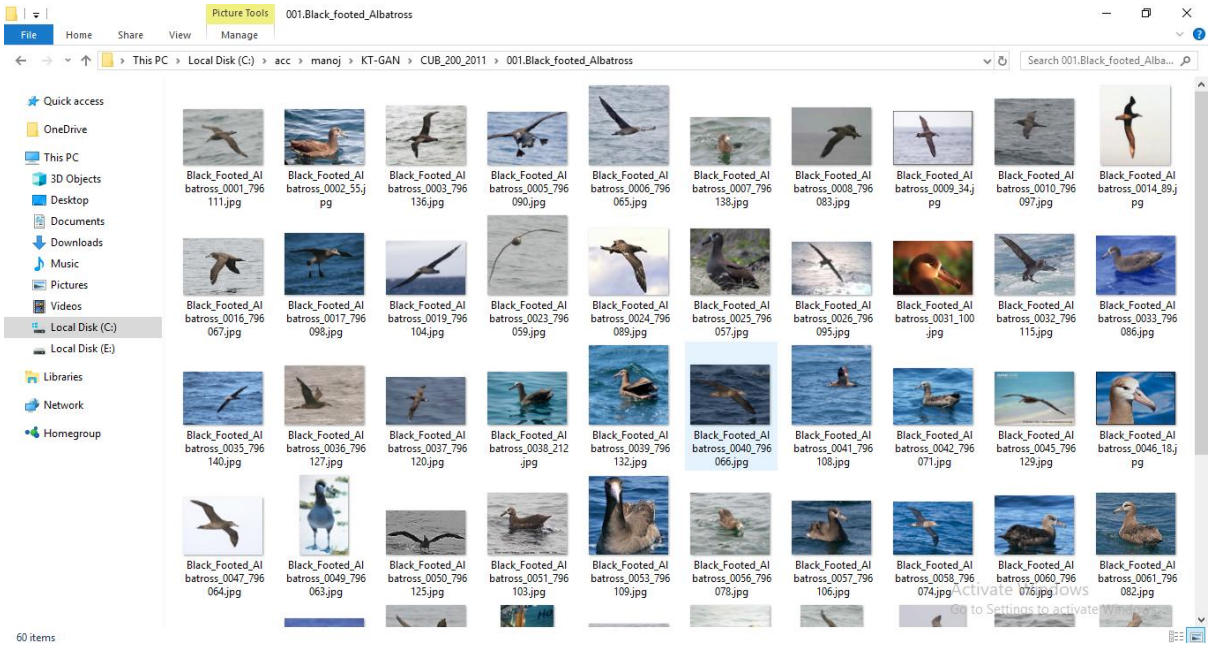
About Data Set

To implement this project author has used CUB-Bird dataset which contains TEXT and images and by using both TEXT and images we will train KT-GAN model and in below screen we are showing dataset details

In 'birds/birdname/.txt' file contains text for each bird and this you can see in below screen



In above screen we can see bird description text for each bird and in below screen we can see images of all those birds and this images you can find inside 'CUB_200_2011/bird_name/' folder like below screen.



So by using above TEXT and images we will train KT-GAN model.

Results and Discussion

Fig 2: In above screen, model is generated and in text field enter some bird description to get image

Fig 3: In above screen in text field I entered bird description as 'bird with a white breast and a black crown and black webbed feet' and then press 'Generate Image from Text' button to get below output

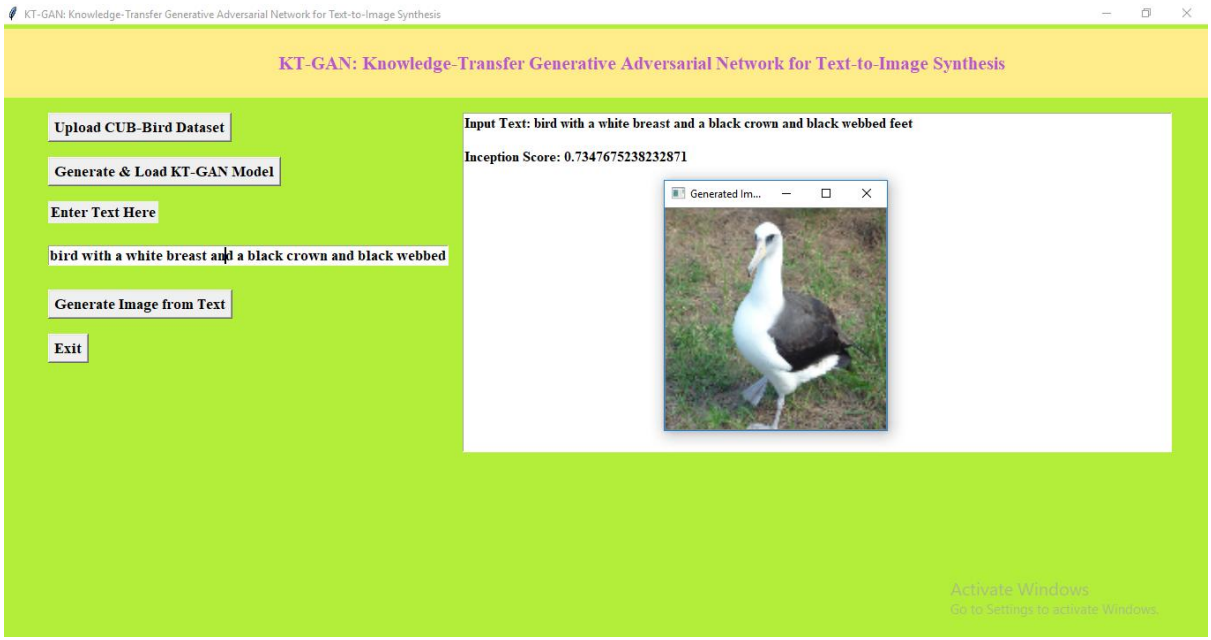


Fig 4: For given bird description we got above image and now try another description for bird

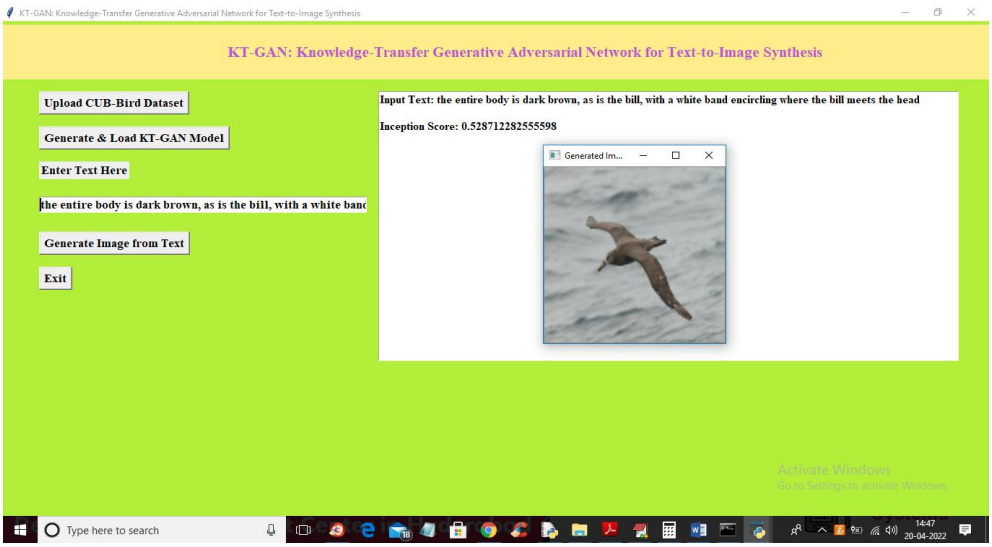
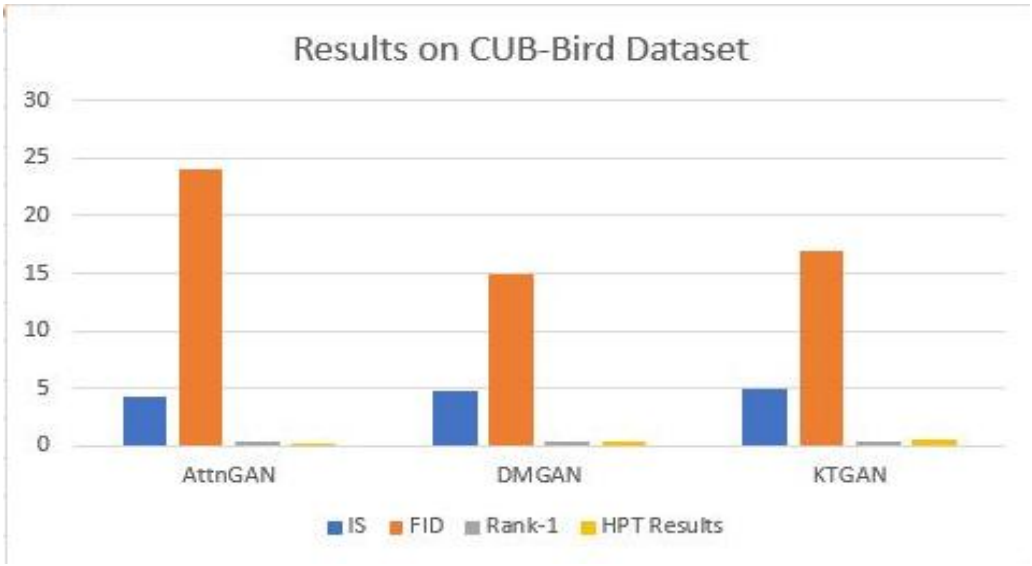
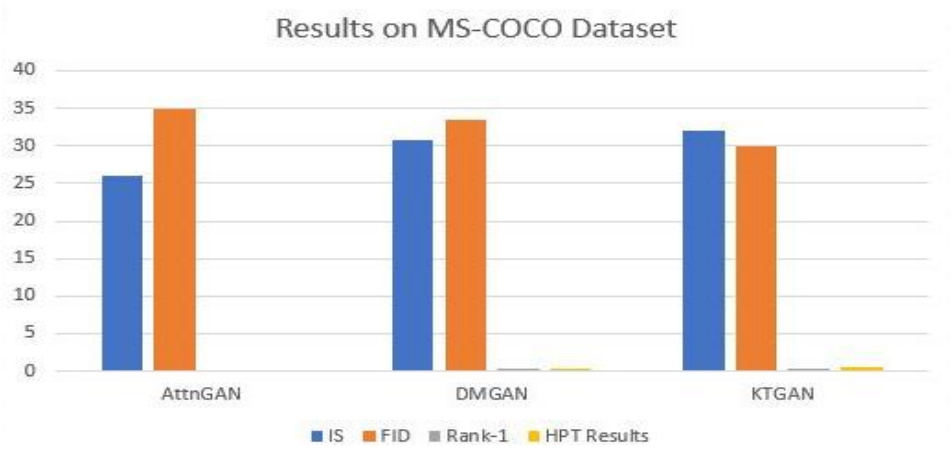


Fig 5: For above description we got the bird image as shown in the above figure

Here the results are shown in graph on two datasets named CUB-Bird and MS-COCO datasets and the performance is measured using four measures named inception score(IS), Rank-1 score, frechet inception distance(FID), Human perceptual test results.
Rank-1 score gives us the relation between the image generated and text entered. FID compares the image generated and realistic images. Human perceptual test calculates how well the humans are able to understand the image generated and IS calculates the power of generating different types of images.



	IS	FID	Rank-1	HPT Results
AttnGAN	4.24	24	28%	22.01%
DMGAN	4.81	15	32%	32.99%
KTGAN	4.91	17	33%	45.01%



	IS	FID	Rank-1	HPT Results
AttnGAN	26.01	34.99	22%	20.00%
DMGAN	30.67	33.54	24%	35.00%
KTGAN	31.99	29.99	25%	46.00%

Conclusion

Here, we developed modern method for generating images from text and this is done using the two main mechanisms namely Alternate Attention Transfer Mechanism and Semantic Distillation Mechanism. Using these two mechanisms we had completed this KTGAN for Text2Image generation. The SDM uses image encoder instructed in image2image function to assist text encoder for text2image function. This involves the knowledge flow from image encoder to text encoder which leads to the improvement in the quality of images generated. The AATM is used to assign weights to the words and through this we are able to identify the important words. Using these two mechanisms we are able to decrease the heterogeneous gap

and able to generate better quality images. The results of KTGAN shows that the performance of this is better compared to the previously used methods.

References

[1] A. RoyChowdhury et al., "Automatic adaptation of object detectors to new domains using self-training," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 780–790.
[2] M. Bharti, G. Fabio, and A. Sikandar, "Knowledge distillation for endto-end person search," in Proc. BMVC, 2019, pp. 1–16.
[3] L. Bowen, Q. Xiaojuan, L. Thomas, and H. S. T. Philip, "Controllable text-to-image generation," in Proc. NeurIPS, 2019, pp. 2065–2075.
[4] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4681–4690.
[5] F. Faghri, J. David Fleet, J. R. Kiros, and S. Fidler, "Vse++: Improved visual-semantic embeddings," in Proc. BMVC, 2018, pp. 1–9.
[6] M. NicolásGuil Francisco Castro and J. Manuel Marín-Jiménez, "Endto-end incremental learning," in Proc. ECCV, Sep. 2018, pp. 233–248.
[7] F. Tung and G. Mori, "Knowledge distillation based on similarity," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 1365–1374.
[8] H. Geoffrey, V. Oriol, and D. Jeff, "Distilling the knowledge in a neural network," in Proc. NeurIPS Workshop, 2015, pp. 1–9.
[9] J. Ian Goodfellow et al., "Generative adversarial nets," in Proc. NeurIPS, 2014, pp. 2672–2680.
[10] G. Yin, B. Liu, L. Sheng, N. Yu, X. Wang, and J. Shao, "Semantics disentangling for text-to-image generation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 2327–2336.
[11] H. Zhang et al., "StackGAN++: Synthesis of realistic images using stacked generative adversarial networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 8, pp. 1947–1962, Aug. 2019.
[12] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in Proc. NeurIPS, 2017, pp. 6626–6637.
[13] H. Tan, X. Liu, X. Li, Y. Zhang, and B. Yin, "Adversarial nets with semantic enhancements for text-to-image synthesis," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 10501–10510.
[14] W. Huang, Y. Xu, and I. Oppermann, "Realistic image generation using region-phrase attention," 2019, arXiv:1902.05395. [Online]. Available: http://arxiv.org/abs/1902.05395
[15] J. Li, K. Fu, S. Zhao, and S. Ge, "Spatiotemporal knowledge distillation for efficient estimation of aerial video saliency," IEEE Trans. Image Process., vol. 29, pp. 1902–1914, 2020