Yanwu Xu

1996.03 02134, Boston, US +01 7249893729 yanwuxu@bu.edu

Personal Summary

Currently, I am a 5th-year Ph.D. student in the ECE department at Boston Univeristy land will be **graduating in the summer of 2024**. I am working on the most efficient large-scale generative models on both Images and language. My research works include UFOGen (One step large-scale text-to-image generative diffusion models), MobileDiffusion (0.2 second inference speed mobile model of large-scale text-to-image generative diffusion models), Diffusion-GANs hybrid generative models and so on.

Education

Boston University **Boston, MA, USA**
Ph.D, *Electrical and Computer Engineering.* *Jan. 2023-*
Advisor: Kayhan Batmanghelich.

The University of Pittsburgh **Pittsburgh, PA 15224, USA**
M.S.., Intelligent System Program. *Sep. 2019-2022*
Advisor: Kayhan Batmanghelich, Co-Advisor: Mingming Gong.

Central South University **Changsha, C.N.**
B.S., Electronic Mechanical Engineering. *Sep. 2013-2017*

Technical Skills

Programing Language: Python , Matlab, C++. Machine Learning Framework: Pytorch(master), Jax(master), Tensorflow2(Master).

Work Background

• Computer Vision: Text to Image Generation, Text image editing, Nerf 3D reconstruction, 3D image style transfer, video style transfer, video motion tracking, metric learning, few-shot learning, semantic Segmentation, object detection, image generation, image editing.
• Medical Image Analysis: brain tumor detection, lung lesion segmentation, medical scanning generation.
• Machine Learning: adversarial learning, domain transfer, weakly supervised learning, unsupervised learning.
• Familiar Model: Diffusion-GANs hybrid, Diffusion, GANs, VAE, F-RCNN, LSTM, etc.

Research Topics

My research topics mainly include the generative models (Diffusion-GAN, Diffusion, GAN, VAE) and their application on text-2-image generation, text image editing, weakly supervised learning, domain adaptation, and image-to-image translation. I study the stability of the training generative model on the large-scale dataset and training model with limited data and limited annotations. My current applications are mainly on image style transfer of 2D object, 3D object and videos and Nerf 3D resconstruction . My other research also covers the area of medical image analysis, which is weakly supervised brain tumor segmentation, domain adaptation, and domain generalization on abdominal segmentation tasks.

Work Experience

Sep 30.2023-: Student researcher, **Google**, Boston, MA, USA

Achiving the fastest on device large-scale text-to-image generative model (MobileDiffusion: Subsecond Text-to-Image Generation on Mobile Devices).

May 30.2023-Sep 1.2023: Research intern, **Google**, Sunnyvale, CA, USA

Modeling super fast one step large-scale text-to-image generative models (UFOGen: You Forward Once Large Scale Text-to-Image Generation via Diffusion GANs).

Sep.2022-Nov.2022: Student researcher, **Google**, Mountain View, CA, USA

During the research work, we target at boost the generation speed of Diffusion model and the text image editing with Diffusion model (Semi-Implicit Denoising Diffusion Models (SIDDMs)).

May.2022-Augst.2022: Research intern, **Google**, Mountain View, CA, USA

During the intern, we developed a on-device product for the general style transfer, which can be real-time for photos and videos. In this project, I built a pipeline for training the styleization models and the data preporcessing.

Augst.2022-November.2022: Student researcher, **Google**, Mountain View, CA, USA

In the extension work as student researcher, we are focusing on 3D object stylization and the high fidelity image stylization and editing. We propose two novel task and framework for 3D NeRF stylization and 2D image style transfer based on the recent diffusion generative model.

Research Experience

Diffusion GANs: In this work, collaborating with Google, we together build a fundamental work of hybrid Diffusion GANs (SIDDMs). We address the issue of instable training of GANs and the slowness of Diffusion Models. And then we propses a novel Diffusion GAN model (UFOGen) and enable the fastest one-step denoising diffusion model for large-scale text-to-image generation.

Conditional GANs: We found that AC-GAN fails on the large scale conditional generation on the image caused by the missing entropy term. To solve this issue, we propose the method TAC-GAN with mutual information maximization player to complement the missing term and achieve the SOTA on multiple datasets, CIFAR10, CIFAR100 and VGGFACE. Accepted by NeurIPS 2019.

Complementary Learning: In this task, we can only access the complimentary labels, which only indicate the negative category. Thus, to improve the previous discriminative method, we propose the generative-discriminative model to generate accurate annotated samples. The experiments show that our method improves this task by a large gap compared to the discriminative method. Accepted by AAAI 2020.

Image-to-Image Translation: The current popular I2I models, such as CycleGAN, GCGAN, and CUT, are too restrictive or weak for specific I2I tasks. For more general applications, we propose a Maximum Spatial Purterbation Consistency (MSPC) model, which designs an adversarial spatial perturbation against the translation network, and we achieve the best results over several popular image translation tasks and our designed face posing transfer dataset. Accepted by CVPR2022.

Domain Generalization: Different from the domain adaptation task, the DG does not have access to the images in the target domain. We can utilize a shallow CNN to generate the synthetic domain to generalize to the unseen domains. We propose an adversarial domain generalization against the main task, which can further improve the generalization ability of the classification and segmentation models. Under review of MICCAI2022.

Research Publication

Preprints:

•UFOGen: You Forward Once Large Scale Text-to-Image Generation via Diffusion GANs, **Yanwu Xu**, Yang Zhao, Zhisheng Xiao, Tingbo Hou.

•MobileDiffusion: Subsecond Text-to-Image Generation on Mobile Devices, Yang Zhao, **Yanwu Xu**, Zhisheng Xiao, Tingbo Hou.

•MedSyn: Text-guided Anatomy-aware Synthesis of High-Fidelity 3D CT Images, Arxiv preprint, Yanwu Xu, Li Sun, Wei Peng, Shyam Visweswaran, Kayhan Batmanghelich.

Accpeted Papers:

•Semi-Implicit Denoising Diffusion Models (SIDDMs), NeurIPS2023, **Yanwu Xu**, Mingming Gong, Shaoan Xie, Wei Wei, Matthias Grundmann, Kayhan Batmanghelich, Tingbo Hou.

•Unpaired Image-to-Image Translation With Shortest Path Regularization, CVPR2023, Shaoan Xie, **Yanwu Xu**, Mingming Gong, Kun Zhang.

•Hierarchical amortized gan for 3D high resolution medical image synthesis, IEEE journal of biomedical and health informatics, Li Sun, Junxiang Chen, **Yanwu Xu**, Mingming Gong, Ke Yu, Kayhan Batmanghelich.

•Maximum Spatial Perturbation Consistency for Unpaired Image-to-Image Translation, CVPR2022, **Yanwu Xu**, Shaoan Xie, Wenhao Wu, Kun Zhang, Mingming Gong*, Kayhan Batmanghelich* (* equal advising).

•Adversarial Consistency for Single Domain Generalization in Medical Image Segmentation, MICCAI2022, **Yanwu Xu**, Shaoan Xie, Maxwell Reynolds, Matthew Ragoza, Mingming Gong, Kayhan Batmanghelich.

• Generative-DiscriminativeComplementaryLearning, AAAI2020, **Yanwu Xu**, Mingming Gong, Junxiang Chen, Tongliang Liu, Kun Zhang, Kayhan Batmanghelich.

• Twin Auxiliary Classifiers GAN, NeurIPS2019 spotlight (0.2%), Mingming Gong*, **Yanwu Xu***, Chunyuan Li, Kun Zhang, Kayhan Batmanghelich (* equal contribution).

• CompressedSelf-AttentionforDeepMetricLearning,AAAI2020,

Ziye Chen, Mingming Gong, **Yanwu Xu**, Chaohui Wang, Kun Zhang, Bo Du.

• Robust Local Descriptor Learning, ACCV2018, **Yanwu Xu**, Mingming Gong, Tongliang Liu, Kayhan Batmanghelich, and Chaohui Wang.

• Multi-scale Masked 3-D U-Net for Brain Tumor Segementation, BraTS2018 LNCS, **Yanwu Xu**, Mingming Gong, Huan Fu, Dacheng Tao, Kun Zhang and Kayhan Batmanghelich.

• DSANet: Dynamic Segment Aggregation Network for Video-Level Representation Learning, ACMMM 2021, Wenhao Wu, Yuxiang Zhao, **Yanwu Xu**, Xiao Tan, Dongliang He, Zhikang Zou, Jin Ye, Yingying Li, Mingde Yao, Zichao Dong, Yifeng Shi

• Unaligned image-to-image translation by learning to reweight, ICCV 2021, Shaoan Xie, Mingming Gong, **Yanwu Xu**, Kun Zhang

• Multi-scale masked 3-D U-net for brain tumor segmentation, MICCAI Brainlesion Workshop 2019, **Yanwu Xu**, Mingming Gong, Huan Fu, Dacheng Tao, Kun Zhang, Kayhan Batmanghelich.

• 3D-BoxSup: Positive-Unlabeled Learning of Brain Tumor Segmentation Networks From 3D Bounding Boxes, Frontiers in neuroscience 14, 350, **Yanwu Xu**, Mingming Gong, Junxiang Chen, Ziye Chen, Kayhan Batmanghelich

Conference Reviewer

NeurIPS2023, CVPR2023, ICML2023, ICCV2023, MICCAI2023, NeurIPS2022, ECCV2022, MICCAI2022, ICPR2022, IJCAI2021, IJCAI2020, CDML2020(NeurIPS workshop), CVPR2021, AAAI2020 (intotal around 50 paper reviews)

Patent

With Google, Semi-Implicit Denoising Diffusion Models, GP-306001-00-PR (IDF-314475)