CHIN-YUAN YEH

(+886)972311499 ⋄ marrch30@gmail.com
11088 5F., No. 22, Aly. 5, Ln. 423, Sec. 5,
Zhongxiao E. Rd., Xinyi Dist., Taipei, Taiwan

SUMMARY

Pursuing PhD degree in deep learning research, with a focus on preventing AI from personal harm.
Recent works: white-box and black-box adversarial attack against image-to-image translation GANs, with potential application on tackling the problem of DeepFake/DeepNude.
Bilingual: English and Mandarin Chinese.
Personal website: `jimmy-academia.github.io`

CURRENT POSITION

PhD Student. Advisor: Professor Ming-Syan Chen *September 2020 - present*

Data Science Program, Graduate Institute of Communication Science, National Taiwan University

EDUCATION

National Taiwan University *January 2018 - April 2020*
Master of Science
Department of Electrical Engineering Graduate GPA: 3.74

National Taiwan University *September 2013 - June 2017*
Bachelor of Science
Department of Physics Graduate GPA: 3.86

SELECTED RESEARCH WORKS

Attack as the Best Defense: Nullifying Image-to-image Translation GANs via Limit-aware Adversarial Attack (paper accepted to ICCV 2021)

In order to progress towards a practical method for tackling DeepFake/DeepNude algorithm, we propose a query-based black-box adversarial attack, the *Limit-Aware Self-Guiding Gradient Sliding Attack (LaS-GSA)*, which is able to effectively nullify image translation function and force the model output to become similar to the input. Optimization techniques utilized in *LaS-GSA* includes: 1) limit-aware sampling, which samples query from hyper-ellipsoid stretched from a hypersphere adhering to the adversarial limit; 2) gradient sliding step to prolong the procedure along the constraint boundary; 3) self-guiding prior, which is an approximation of the true adversarial gradient extracted solely from the threat model and the target image. The prior is obtained by exploiting the fact that the Jacobian of the image-to-image translation function is sufficiently diagonalized due to the semantic consistent constraint. Source code: https://github.com/jimmy-academia/LaSGSA

Disrupting Image-Translation-Based DeepFake Algorithms with Adversarial Attacks (paper accepted to WACV2020 DeepPAB Workshop)

This work addresses the serious challenge presented by DeepNude, a Deepfake application that undresses photography of any human being, by introducing the novel aspect of adversarially attacking the DeepFake model. Projected Gradient Descent (PGD) Attack is utilized with modifications on image translation GANs including CycleGAN, pix2pix, and pix2pixHD models trained with CelebA dataset. Two attacks are proposed, including the Nullifying Attack, which causes the attacked model to output an image similar to the input, and the Distorting Attack, which causes the model to output a distorted figure. Source code: https://github.com/jimmy-academia/Adversarial-Attack-CycleGAN-and-pix2pix

TECHNICAL STRENGTHS

Programming Language	Python & Pytorch
Software & Tools	Unix System, Bash scripts, Latex, Blender

WORK EXPERIENCE

Taiwan AI Academy *October 2021 - present*
Research and Teaching Assistant

· I organize and introduce cutting-edge research development to an Taiwanese-industry facing audience. Topics include: "Introduction and Application of Graph Neural Network," "Adversarial Attacks: where to find them and how to defend."

Institute of Information Science, Academia Sinica *April - August 2020*
Graduate Research Assistant

· I studied Graph Neural Network and optimization techniques to further advance the research on black-box adversarial attacks. Provided assistant in hosting the International Conference on Database Systems for Advanced Applications, DASFAA.

Department of Electrical Engineering, National Taiwan University *August 2018 - April 2020*
Research Assistant

· I studied relevant topics of Generative Adversarial Networks and adversarial attacks, to conduct the research work "Disrupting Image-Translation-Based DeepFake Algorithms with Adversarial Attacks." Advised by Professor Sheng-De Wang.

Shalom Inc. 旅安資訊 *March - August 2018*
Software Engineer Intern

· On-site internship under this leading tech company dedicated to providing information infrastructures for hotel management. Completed various tasks including setting up AWS cloud server, writing scripts for automatic pricing update script via custom APIs, and fetching customer information from records.

Institute of Physics, Academia Sinica *February - July 2016*
Undergraduate Research Assistant

· Conducted material science research on "Growth and Analysis of Flexible Oxide Electronic Materials" (可撓性微波氧化物材料之成長與分析) under advisors, Professor Chang, Chia-Seng and Professor Chu, Ying-Hao. The project is funded by Taiwan Ministry of Science and Technology for College Student Research Project 105-2815-C-001-011-M.

OTHER EXPERIENCE

Mandatory military training in the Republic of China (Taiwan) Marine Corps. *June - October 2017*

One year exchange student at the University of British Columbia. *September 2016 - May 2017*

Chief editor of the 34^{th} issue of "SpaceTime," the department journal of the Department of Physics, National Taiwan University. *January 2013*