Xin Ma

189-8048-1629 xin.ma1@monash.edu
Doctor of Philosophy Monash University

Highly motivated Ph.D. student at Monash University. My research interests lie in video and image generation and editing, multimodal models, low-level vision, and among others. Currently, several papers have been published in international conferences or journals. In terms of practical application, several national invention patents have been applied. Here is my personal website. The current total citation on Google Scholar is **1,514**.

Education

Present Sep 2022	Faculty of Information Technology, **Monash University** Ph.D. Student in Computer Science
Jun 2021 Sep 2018	School of Artificial Intelligence, **University of Chinese Academy of Sciences** Master's Degree in Computer Technology
Jun 2018 Sep 2014	School of Electrical and Information Engineering, **JiangSu University** Bachelor's Degree in Applied Electronic Information Engineering

Work Experiences

Present Dec 2022	Researcher @ Shanghai Artificial Intelligence Laboratory (**Inter**) › Research on generative models. We propose LaVie, a large-scale text-to-video framework that produces high-quality and temporally coherent videos. A general Transformer-based latent video diffusion model, referred to as Latte, is introduced. And an image animation framework Cinemo based on motion diffusion models is introduced.
Sep 2022 Jun 2021	Algorithm engineer @ Meituan •Vision Intelligence Department (**Full-time Work)** › Research on model compression: A model compression tool has been developed to assist developers in rapidly deploying models to edge devices without compromising model accuracy while simultaneously enhancing model inference speed. This tool has been extensively implemented across various businesses at Meituan. One paper was accepted by CVPR 2022 during this period.
Aug 2020 Apr 2020	Agorithm engineer @ Meituan •Vision Intelligence Center (**Inter**) › Research on Image Dewatermarking Algorithm: An image dewatermarking algorithm was proposed based on attention mechanism and self-supervised learning. The algorithm is now launched on Meituan and used in Meituan takeaway dewatermarking system. Related work was accepted by ICPR 2020 and selected as an oral presentation.

Publications

- Latte: Latent Diffusion Transformer for Video Generation
  Xin Ma, Yaohui Wang, Gengyun Jia, Xinyuan Chen, Ziwei Liu, Yuan-Fang Li, Cunjian Chen, Yu Qiao
  Transactions on Machine Learning Research (TMLR), 2025
- Consistent and Controllable Image Animation with Motion Diffusion Models
  Xin Ma, Yaohui Wang, Gengyun Jia, Xinyuan Chen, Tien-Tsin Wong, Yuan-Fang Li, Cunjian Chen
  Computer Vision and Pattern Recognition (CVPR), 2025, CCF-A
- LaVie: High-Quality Video Generation with Cascaded Latent Diffusion Models
  Yaohui Wang*, Xinyuan Cheng*, **Xin Ma***, Shangchen Zhou, Ziqi Huang, Yi Wang, Ceyuan Yang, Yinan He, Jiashuo Yu, Peiqing Yang, Yuwei Guo, Tianxing Wu, Chenyang Si, Yuming Jiang, Cunjian Chen, Chen Change Loy, Bo Dai, Dahua Lin, Yu Qiao, Ziwei Liu

International Journal of Computer Vision (IJCV), 2024, JCR Q1 & CCF-A

- LEO: Generative Latent Image Animator for Human Video Synthesis
  Yaohui Wang, **Xin Ma**, Xinyuan Chen, Antitza Dantcheva, Bo Dai, Yu Qiao
  International Journal of Computer Vision (IJCV), 2024, JCR Q1 & CCF-A
- Internvid: A large-scale video-text dataset for multimodal understanding and generation
  Yi Wang, Yinan He, Yizhuo Li, Kunchang Li, Jiashuo Yu, **Xin Ma**, Xinyuan Chen, Yaohui Wang, Ping Luo, Ziwei Liu, Yali Wang, Limin Wang, Yu Qiao
  International Conference on Learning Representations (ICLR), 2024
- SEINE: Short-to-Long Video Diffusion Model for Generative Transition and Prediction
  Xinyuan Chen, Yaohui Wang, Lingjun Zhang, Shaobin Zhuang, **Xin Ma**, Jiashuo Yu, Yali Wang, Dahua Lin, Yu Qiao, Ziwei Liu
  International Conference on Learning Representations (ICLR), 2024
- Uncertainty-Aware Image Inpainting with Adaptive Feedback Network
  Xin Ma, Xiaoqiang Zhou, Huaibo Huang, Gengyun Jia, Yaohui Wang, Xinyuan Chen, Cunjian Chen
  Expert Systems with Applications (ESWA), 2023, JCR Q1 & CCF-C
- Compressing Models with Few Samples: Mimicking then Replacing
  Huanyu Wang, Junjie Liu, **Xin Ma**, Yang Yong, Zhenhua Chai, Jianxin Wu
  Computer Vision and Pattern Recognition (CVPR), 2022, CCF-A
- Style-based Attentive Network for Real-World Face Hallucination
  Mandi Luo*, **Xin Ma***, Huaibo Huang, Yi Li, Ran He
  Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2022, CCF-C
- Contrastive Attention Network with Dense Field Estimation for Face Completion
  Xin Ma*, Xiaoqiang Zhou*, Huaibo Huang, Gengyun Jia, Zhenhua Chai, Xiaolin Wei
  Pattern Recognition (PR), 2021, JCR Q1 & CCF-B
- Partial NIR-VIS Heterogeneous Face Recognition with Automatic Saliency Search
  Mandi Luo, **Xin Ma**, Zhihang Li, Jie Cao, Ran He
  IEEE Transactions on Information Forensics and Security (T-IFS), 2021, JCR Q1 & CCF-A
- FA-GAN: Face Augmentation GAN for deformation-invariant face recognition
  Mandi Luo, Jie Cao, **Xin Ma**, Xiaoyu Zhang, Ran He
  IEEE Transactions on Information Forensics and Security (T-IFS), 2021, JCR Q1 & CCF-A
- Inconsistency-aware Wavelet Dual-branch Network for Face Forgery Detection
  Gengyun Jia, Meisong Zheng, Chuanrui Hu, **Xin Ma**, Yuting Xu, Luoqi Liu, Yafeng Deng, Ran He
  IEEE Transactions on Biometrics, Behavior, and Identity Science (T-BIOM), 2021
- Free-Form Image Inpainting via Contrastive Attention Network
  Xin Ma, Xiaoqiang Zhou, Huaibo Huang, Zhenhua Chai, Xiaolin Wei, Ran He
  International Conference on Pattern Recognition (ICPR), 2020 (**oral** 5%), CCF-C
- Unsupervised Contrastive Photo-to-Caricature Translation based on Auto-distortion
  Yuhe Ding*, **Xin Ma***, Mandi Luo, Aihua Zheng, Ran He
  International Conference on Pattern Recognition (ICPR), 2020, CCF-C

—— Preprints ——

- Training-free Stylized Text-to-Image Generation with Fast Inference
  Xin Ma, Yaohui Wang, Xinyuan Chen, Tien-Tsin Wong, Cunjian Chen
- Consistent and Controllable Image Animation with Motion Linear Diffusion Transformers
  Xin Ma, Yaohui Wang, Gengyun Jia, Xinyuan Chen, Tien-Tsin Wong, Cunjian Chen

Granted Patents

- Model training method, map building method and device, CN114972909B
- Attention-mechanism-based image completion method and device, CN112184582B
- Image completion method based on uncertainty estimation, CN112686817B
- Human face image super-resolution method based on attention mechanism, CN111080513B
- Image super-resolution method of adversarial generative network based on fusion mutual information, CN110660020B

- Image super-resolution method of deep neural network fusing mutual information, CN110211035B
- Cartoon style image conversion model training method, image generation method and device, CN112232485B

Awards & Certificates

2025 Outstanding reviewer for CVPR 2025
2025 CVPR 2025 Travel Support Award
2022 PRCV2022 Aerial-Ground Intelligent Unmanned System Environment Perception Challenge, the second place
2021 ICCV2021 Low Power Computer Vision Challenge: the second place
2021 Merit Student of University of Chinese Academy of Sciences

Competences & Languages

Programming Python, C/C++, Matlab, Latex, Pytorch, Linux
Languages **English** —CET-6; PTE 62

马鑫

189-8048-1629 xin.ma1@monash.edu
计算机科学 • 博士 蒙纳士大学

蒙纳士大学在读博士。研究兴趣为机器学习和模式识别，主要关注于视频和图像生成与编辑、多模态模型、底层视觉生成等。目前已向国际会议或期刊发表多篇论文。在实际应用方面，目前已申请多项国家发明专利。个人学术网页请见这里。目前谷歌学术总引用量为 **1,514**。

教育背景

现在 2022.09 | **蒙纳士大学** • 信息技术系
计算机科学 • 博士

2021.06 2018.09 | **中国科学院大学** • 人工智能学院
计算机技术 • 硕士

2018.06 2014.09 | **江苏大学** • 电气信息工程学院
电子信息技术 • 学士

工作经历

现在 2022.12 | （实习）见习研究员 @ 上海人工智能实验室 • 通用视觉组

- **生成模型:** 主要研究生成模型，设计了一个大规模 text-to-video 框架LaVie, 可以生成高质量的视频。提出了一种基于 Transformer 的视频扩散模型，称为 Latte。同时，也设计了一种基于运动残差的图像动画化方法Cinemo。

2022.09 2021.07 | （全职）算法工程师 @ 美团 • 视觉智能中心

- **模型压缩**: 在此期间开发了一个模型压缩工具，可以帮助开发人员在不影响模型精度的情况下快速地将模型部署到边缘设备上，并提高模型推理速度。该工具已广泛应用于美团的各种业务中。同时在 CVPR2022 上合作发表了一篇小样本的模型压缩方法。

2020.08 2020.04 | （实习）深度学习算法研发 @ 美团 • 基础视觉组

- **图像去水印算法研究:** 基于注意力机制和自监督学习，设计了一种高质量的图像去水印算法。该算法现已上线美团网，用于美团外卖去水印系统。相关工作已被 ICPR2021 接受，并选为 oral。

论文发表

- Latte: Latent Diffusion Transformer for Video Generation
  Xin Ma, Yaohui Wang, Gengyun Jia, Xinyuan Chen, Ziwei Liu, Yuan-Fang Li, Cunjian Chen, Yu Qiao
  Transactions on Machine Learning Research (TMLR), 2025
- Consistent and Controllable Image Animation with Motion Diffusion Models
  Xin Ma, Yaohui Wang, Gengyun Jia, Xinyuan Chen, Tien-Tsin Wong, Yuan-Fang Li, Cunjian Chen
  Computer Vision and Pattern Recognition (CVPR), 2025, CCF-A
- LaVie: High-Quality Video Generation with Cascaded Latent Diffusion Models
  Yaohui Wang*, Xinyuan Cheng*, **Xin Ma***, Shangchen Zhou, Ziqi Huang, Yi Wang, Ceyuan Yang, Yinan He, Jiashuo Yu, Peiqing Yang, Yuwei Guo, Tianxing Wu, Chenyang Si, Yuming Jiang, Cunjian Chen, Chen Change Loy, Bo Dai, Dahua Lin, Yu Qiao, Ziwei Liu
  International Journal of Computer Vision (IJCV), 2024, JCR Q1 & CCF-A
- LEO: Generative Latent Image Animator for Human Video Synthesis
  Yaohui Wang, **Xin Ma**, Xinyuan Chen, Antitza Dantcheva, Bo Dai, Yu Qiao

International Journal of Computer Vision (IJCV), 2024, JCR Q1 & CCF-A

- Internvid: A large-scale video-text dataset for multimodal understanding and generation
  Yi Wang, Yinan He, Yizhuo Li, Kunchang Li, Jiashuo Yu, **Xin Ma**, Xinyuan Chen, Yaohui Wang, Ping Luo, Ziwei Liu, Yali Wang, Limin Wang, Yu Qiao
  International Conference on Learning Representations (ICLR), 2024
- SEINE: Short-to-Long Video Diffusion Model for Generative Transition and Prediction
  Xinyuan Chen, Yaohui Wang, Lingjun Zhang, Shaobin Zhuang, **Xin Ma**, Jiashuo Yu, Yali Wang, Dahua Lin, Yu Qiao, Ziwei Liu
  International Conference on Learning Representations (ICLR), 2024
- Uncertainty-Aware Image Inpainting with Adaptive Feedback Network
  Xin Ma, Xiaoqiang Zhou, Huaibo Huang, Gengyun Jia, Yaohui Wang, Xinyuan Chen, Cunjian Chen
  Expert Systems with Applications (ESWA), 2023, JCR Q1 & CCF-C
- Compressing Models with Few Samples: Mimicking then Replacing
  Huanyu Wang, Junjie Liu, **Xin Ma**, Yang Yong, Zhenhua Chai, Jianxin Wu
  Computer Vision and Pattern Recognition (CVPR), 2022, CCF-A
- Style-based Attentive Network for Real-World Face Hallucination
  Mandi Luo*, **Xin Ma***, Huaibo Huang, Yi Li, Ran He
  Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2022, CCF-C
- Contrastive Attention Network with Dense Field Estimation for Face Completion
  Xin Ma*, Xiaoqiang Zhou*, Huaibo Huang, Gengyun Jia, Zhenhua Chai, Xiaolin Wei
  Pattern Recognition (PR), 2021, JCR Q1 & CCF-B
- Partial NIR-VIS Heterogeneous Face Recognition with Automatic Saliency Search
  Mandi Luo, **Xin Ma**, Zhihang Li, Jie Cao, Ran He
  IEEE Transactions on Information Forensics and Security (T-IFS), 2021, JCR Q1 & CCF-A
- FA-GAN: Face Augmentation GAN for deformation-invariant face recognition
  Mandi Luo, Jie Cao, **Xin Ma**, Xiaoyu Zhang, Ran He
  IEEE Transactions on Information Forensics and Security (T-IFS), 2021, JCR Q1 & CCF-A
- Inconsistency-aware Wavelet Dual-branch Network for Face Forgery Detection
  Gengyun Jia, Meisong Zheng, Chuanrui Hu, **Xin Ma**, Yuting Xu, Luoqi Liu, Yafeng Deng, Ran He
  IEEE Transactions on Biometrics, Behavior, and Identity Science (T-BIOM), 2021
- Free-Form Image Inpainting via Contrastive Attention Network
  Xin Ma, Xiaoqiang Zhou, Huaibo Huang, Zhenhua Chai, Xiaolin Wei, Ran He
  International Conference on Pattern Recognition (ICPR), 2020 (**oral** 5%), CCF-C
- Unsupervised Contrastive Photo-to-Caricature Translation based on Auto-distortion
  Yuhe Ding*, **Xin Ma***, Mandi Luo, Aihua Zheng, Ran He
  International Conference on Pattern Recognition (ICPR), 2020, CCF-C

—— Preprints ——

- Training-free Stylized Text-to-Image Generation with Fast Inference
  Xin Ma, Yaohui Wang, Xinyuan Chen, Tien-Tsin Wong, Cunjian Chen
- Consistent and Controllable Image Animation with Motion Linear Diffusion Transformers
  Xin Ma, Yaohui Wang, Gengyun Jia, Xinyuan Chen, Tien-Tsin Wong, Cunjian Chen

专利授权

- 一种模型训练的方法、构建地图的方法及装置，CN114972909B
- 一种基于注意力机制的图像补全方法及装置，CN112184582B
- 一种基于不确定性估计的图像补全方法，CN112686817B
- 一种基于注意力机制的人脸图像超分辨率方法，CN111080513B
- 一种基于融合互信息的对抗生成网络的图像超分辨率方法，CN110660020B

- 融合互信息的深度神经网络的图像超分辨率方法，CN110211035B
- 漫画风格图像转换模型的训练方法、图像生成方法及装置，CN112232485B

获奖及证书

2025 CVPR2025 杰出审稿人
2025 CVPR2025 差旅支援奖
2022 PRCV2022 Aerial-Ground Intelligent Unmanned System Environment Perception Challenge, 第二名
2021 ICCV2021 Low Power Computer Vision Challenge，第二名
2021 中国科学院大学三好学生

技能和语言

编程 Python, Matlab, C++, Latex, Pytorch
语言 **英语** — 大学英语六级，PTE 62