潘飞 深度学习与计算机视觉

性别：男　　出生年月：1993 年 6 月

电话：19357638601　　电子邮箱：feipanir@icloud.com

学历：博士　　政治身份：中共党员

具备扎实的计算机视觉与深度学习研发背景，长期从事语义分割、域自适应、自监督表征学习与图像生成相关研究，擅长在标注数据稀缺、跨域分布差异显著的场景下，构建合成数据与生成式增强方案，提升视觉模型的泛化能力与鲁棒性。拥有从合成数据设计与生成、算法训练与优化，到模型压缩、蒸馏与轻量化部署的完整研发经验，覆盖语义分割、异常检测、开放词汇视觉感知、生成式数据增强与向量检索系统等方向，并在 CVPR、ECCV、NeurIPS 等顶级学术会议发表相关工作。

教育背景

2018.03 - 2023.08　**博士 - 韩国科学技术院（QS 世界排名 65）**
电子信息工程 | 主攻方向: 机器学习 & 计算机视觉 | 导师: Dr. In So Kweon
博士论文: Geometric-guided Domain Adaptation for Semantic Segmentation

2016.02 - 2018.02　**硕士 - 韩国科学技术院**
电子信息工程 | 主攻方向: 机器学习 & 计算机视觉

2011.09 - 2015.07　**学士 - 西安电子科技大学**
通信工程 | 学院排名 5/825 (前 0.6%) | 成绩: 3.94/4.0

工作履历

2025.04 -2026.08　**杭州昼光科技有限公司 (Vibe Inc.)**, AI 算法研发 | 研发方向: 具身智能，VLA，端侧智能，模型压缩与量化

- **3D 机器视觉与机器人感知。**基于 Orbbec 与 Intel RealSense 3D 相机，在 Jetson 平台完成多种视觉模型的工程化部署（YOLO、MobileSAM、FastSAM、EdgeSAM、OWL-ViT 等），实现目标检测、语义分割与开放词汇视觉感知；构建基于 PointNet 的 GraspNet、AnyGrasp、Graspness 等抓取检测框架，并结合 MoveIt 实现机械臂抓取与运动控制，完成从视觉感知、姿态估计到机械臂执行的端到端闭环系统；成功部署 ORB-SLAM2 与 VGGT-SLAM，实现机器人实时定位与建图（SLAM）。
- **自采数据集与视觉-语言-动作（VLA）模型微调。**基于 Lerobot 机械臂平台，在自主定义的环境中完成抓取数据采集与数据集构建；基于 Physical Intelligence 的开源模型 π_0 对自采数据进行微调（fine-tuning），实现模型在自定义场景下的稳定抓取与泛化执行能力。
- **超低功耗端侧 Speaker Verification 算法研发。**面向可穿戴设备 AI 芯片的资源约束场景，设计并实现端侧 Speaker Verification 算法体系。针对算力、存储及功耗限制，设计轻量级声纹嵌入网络，并结合知识蒸馏、量化感知训练（QAT）及 ONNX 推理优化，实现模型在嵌入式芯片上的高精度部署；完成从模型训练、压缩优化到芯片端部署的完整算法闭环，单次推理功耗降低至 4mW。
- **多模态会议知识管理 Agent 系统。**设计面向会议协作场景的智能信息抽取系统，通过 ASR 将语音转换为文本，结合大语言模型（LLM）完成会议内容总结与关键信息提取；构建基于文本 Embedding 与向量数据库（Milvus/ChromaDB）的语义检索系统，通过优化文本 Chunking 策略提升信息召回效果；基于 Neo4J 融合图像与文本信息构建知识图谱，实现多模态知识管理与检索增强生成（RAG）。

2023.09 -2025.03　**密西根大学安娜堡分校（QS 排名：51）**, 博士后研究员 | 研究课题: 大语言视觉模型, 扩散模型, ViTs

- **使用扩散模型生成图像进行模型鲁棒性测试。**本项目开发了 ImageNet-D 作为一个新型视觉模型鲁棒性基准测试，主要通过扩散模型 (Diffusion Models) 合成的图像来测试视觉模型的鲁棒性。与以往的基准测试不同，ImageNet-D 在背景、纹理和材料上提供了高度多样化的变化，并具有卓越的合成质量。实验表明，ImageNet-D 导致了一系列视觉模型的准确性显著下降，从基本视觉模型 ResNet 到大语言模型如 CLIP 和 MiniGPT-4，准确性下降高达 60%。这突显了扩散模型在基准测试鲁棒性中的实用性。数据集和代码已公开提供 [5]。
- **使用大规模视觉-语言模型进行零样本学习。**本项目利用大规模视觉-语言模型（LVLMs）从街景和卫星图像中对建筑属性进行分类和分割，包括了房屋的高度，窗户，门，屋顶的位置，还有屋顶的类型。开发了一种基于 CLIP 模型和 SAM 模型的零样本模型 [6]，该模型无需人工标注，并能有效泛化到新领域。此方法在外墙分割任务上超越了当前的有监督方法，性能提升了 20%。
- **基于视觉 Transformer 的无监督分割。**本研究致力于改进自监督视觉 Transformer（ViT）特征，以实现对完整对象的分割。现有的 ViT 特征（例如 DINO）主要针对对象的局部区域进行分割；而我的工作通过在图割框架中引入吸引与排斥机制，进一步优化了这些 ViT 特征 [1]。这一改进使得 ViT 特征能够在图像和视频数据集中实现对完整对象的分割（Unsupervised Whole Object Segmentation），而无需任何关于对象的先验知识。

2016.03 -2023.08 **韩国科学技术院**, GSA | 研究课题：Segmentation, Detection, GANs，域内自适应，3D Object Motion

- **基于生成模型的合成数据构建与域自适应分割。**针对语义分割任务中真实标注数据稀缺的问题，构建了大规模合成数据集（Synthetic Dataset）用于弥补源域与目标域之间的分布差异。基于生成对抗网络（GAN）实现风格迁移，将合成图像的纹理、光照与色彩分布向真实场景对齐，缓解了域偏移（Domain Shift）问题 [5,9]；同时探索基于扩散模型（Diffusion Model）的图像生成方法，用于生成具有更高多样性与真实感的合成训练样本，提升分割模型在真实场景下的泛化能力。该方法在多种地理环境与天气条件下的语义分割任务中取得了当前最优（SOTA）性能。
- **基于 3D 姿态预测、单目视觉深度预测与自适应图像分割方法。**本研究提出了一种利用目标域中城市街景中物体的运动信息，进行域自适应的分割方，其中运动物体与时间线索是关键要素。针对目标域中缺乏标注数据的问题，我开发了一种基于几何约束的学习框架，能够从未标注的视频中同时学习 3D 物体运动 (3D Motion) 与深度图 (Monocular Depth Maps) [10]，并利用这些信息对目标运动物体进行分割与识别 (Detection and Segmentation) [4]。该方法通过引入 3D 运动引导机制，显著提升了在复杂场景中对关键物体（如行人、车辆和骑行者）进行分割。该方法不仅能够更准确地估计物体的深度信息，还能捕捉物体的运动轨迹，从而为分割任务提供更丰富的上下文信息。实验结果表明，该方法在目标域中的泛化能力显著优于传统方法，尤其是在处理动态物体时表现出色。这一成果为自动驾驶、智能监控等领域的视频分析任务提供了新的技术思路。
- **图像分割下的无监督、复合及多目标域自适应. 无监督、复合与多目标域自适应。**针对语义分割任务 (Semantic Segmentation)，开发了基于卷积神经网络（CNN）的创新自适应 (Domain Adaptation) 框架，适用于目标域为单一域 (Single Domain) [11,12] 或多个未知域的复合场景 (Compound Domains)（例如，不同地理环境 [7] 或天气条件 [9])。在方法上，提出了基于聚类 [9] 的域划分策略，利用基于生成对抗网络（GAN）的风格迁移 [9] 实现域适应，并通过掩码自编码器 (Masked Autoencoder) [7] 进行数据增强。此外，开发了一种基于 *softmax* 输出值的置信度度量方法 [12]，并结合自训练 (Self Training) [11] 策略，迭代优化目标域伪标签的质量。通过网络蒸馏与正则化 [9] 技术，训练了一个泛化能力强的学生模型，在多种地理环境（城市、乡村）和天气条件（晴天、雨天、多云、夜晚、阴天）下均取得了当前图像语义分割最优（SOTA）的性能表现。

2021.05 -2021.11 **Robert Bosch**, 研发实习 | 研究课题：Object Detection, Image Enhancement

- **基于 YOLO 的交通路口车辆监控。**本项目参与了利用 YOLO-v5 模型对多路交通监控摄像头视频进行车辆检测的开发工作。由于不同摄像头的拍摄角度、光照条件以及目标遮挡等因素导致的领域差异（domain discrepancies)，检测任务面临了显著挑战。为了解决这些问题，我们采用了多种创新的图像增强技术，包括基于生成对抗网络（GANs）的图像生成、LAB 色彩空间转换以及傅里叶变换，以生成合成数据。此外，还结合了多种数据增强方法，如裁剪、旋转、像素抖动和高斯噪声添加等。通过这些技术手段，我们成功将不同摄像头视角下的车辆检测鲁棒性提升了 20% 显著改善了模型在实际复杂场景中的表现。

软件著作

- **零样本识别（Zero-shot Classification and Segmentation）** [6]: 一种基于大规模视觉语言模型（LVLMs）的零样本识别模型，基于 Python 进行多种视觉任务，包括分类、检测、分割和图像描述，而无需任何人工标注。该技术特别适用于从街景和遥感图像中提取建筑物属性。通过利用预训练的视觉语言模型，零样本识别能够在没有特定任务训练数据的情况下，实现对图像内容的理解和分析，极大地提高了处理效率和灵活性。 github.com/feipanir/ZoBae
- **单目深度与运动估计（Monocular Depth & Motion [4]）**：MoDA 是一个自监督学习框架，旨在从自动驾驶车辆记录的无标签视频中，通过 Structure from Motion (SfM) 技术，学习单目深度、自车运动（ego-motion）以及物体运动（object motion)。该框架无需人工标注数据，能够有效利用大量未标注的视频数据，提升自动驾驶系统对环境的感知能力。MoDA 的开源代码已发布在 GitHub 上，供研究者和开发者参考和使用。 github.com/feipanir/MoDA
- **基于扩散模型的鲁棒性基准测试 (Robusntess Benchmark based on Diffusion Models)** [5]: 该项目利用 Python 构建了基于扩散模型生成的合成对象，旨在评估从标准的视觉模型 ResNet 到 Vision Transformer（ViT）基础模型（如 CLIP 和 MiniGPT-4）的深度神经网络的鲁棒性。通过生成多样化的合成数据，该项目为研究人员提供了一个强大的工具，以测试和改进模型在面对复杂和多样化输入时的表现。 github.com/chenshuang-zhang/imagenet_d
- **基于域自适应的语义分割模型 (Domain Adaptive Semantic Segmentation)** [12]: 基于 Python 的语义分割框架，具备领域自适应功能，旨在提升模型在真实世界驾驶场景图像上的泛化能力。该框架通过先进的领域自适应技术，有效减少了源域和目标域之间的分布差异，从而在复杂的现实环境中实现更精准的语义分割。 github.com/feipanir/IntraDA

技能

PyTorch, Tensorflow, Caffe, Python, C/C++, NumPy, OpenCV, Git, Docker, Scipy, CNNs, GANs, Diffusion Models, ControlNet, ViTs, LLMs, VLMs

获奖与荣誉

2024.06 最佳论文奖 在 CVPR（计算机视觉与模式识别会议）的“有限标注数据学习在图像与视频理解中的应用”研讨会上荣获最佳论文奖，表彰了我们在有限标注数据下通过创新方法提升图像与视频理解能力的研究成果。

2020.12 高通公司创新研发奖因在韩国攻读研究生期间的杰出研究成果而获得高通创新奖学金。该奖学金旨在表彰在通信与计算领域具有创新潜力的研究生，体现了我在相关领域的研究贡献与学术影响力。

2019.09 因在开发智能交通系统方面的创新研究而获得 Robert Bosch 博士奖学金（为期两年）。该奖学金支持我在智能交通领域的深入研究，助力我探索人工智能技术在交通管理与优化中的应用。

2015.05 汇顶科技奖学金，西安电子科技大学本科优秀毕业生。

2014.05 国家奖学金，西安电子科技大学本科期间的优异学术表现而获得国家奖学金。

发表著作

[1] **Fei Pan***, Yixing Wang*, Sangryul Jeon, Stella Yu. *Unsupervised Whole Object Discovery by Contextual Grouping with Repulsion. NeurIPS 2025 Workshop.*

[2] Sanghyun Jo, **Fei Pan**, In-Jae Yu, Kyungsu Kim. *DHR: Dual Features-Driven Hierarchical Rebalancing in Inter-and Intra-Class Regions for Weakly-Supervised Semantic Segmentation. ECCV, 2024.* Paper | Code

[3] Xu Yin, Woobin Im, Dongbo Min, Yuchi Huo, **Fei Pan**, Sung-Eui Yoon. *Fine-grained Background Representation for Weakly Supervised Semantic Segmentation. IEEE Trans. on Circuits and Systems for Video Technology, 2024.* Paper

[4] **Fei Pan**, Xu Yin, Seokju Lee, Axi Niu, Sungeui Yoon, In So Kweon. *MoDA: Leveraging Motion Priors from Videos for Advancing Unsupervised Domain Adaptative Segmentation. CVPR Workshop, 2024.* Best Paper Award. Paper

[5] Chengshuang Zhang, **Fei Pan**, Junmo Kim, In So Kweon, Chengzhi Mao. *ImageNet-D: Benchmarking Neural Network Robustness on Diffusion Synthetic Object. CVPR 2024.* Highlight (Acceptance Rate $< 2.0\%$). Paper | Code & Dataset

[6] **Fei Pan**, Sangryul Jeon, Brian Wang, Frank Mckenna, Stella Yu. *Zero-shot Building Attribute Extraction from Large-Scale Vision and Language Models. WACV 2024.* Paper | Code | Poster

[7] **Fei Pan***, Dong He*, Xu Yin, Chenshuang Zhang, Munchurl Kim. *Masking-augmented Collaborative Domain Congregation for Multi-target Domain Adaptation in Semantic Segmentation. IEEE IV, 2024.* Best Paper Nominated. Paper

[8] Francois Rameau, Jaesung Choe, **Fei Pan**, Seokju Lee, In So Kweon. *CCTV-Calib: A Toolbox to Calibrate Surveillance Cameras Around the Globe. Machine Vision and Applications, 2023.* Paper | Code

[9] **Fei Pan**, Sungsu Heo, Seokju Lee, In So Kweon. *ML-BPM: Multi-teacher Learning with Bidirectional Photometric Mixing for Open Compound Domain Adaptation in Semantic Segmentation. ECCV 2022.* Paper

[10] Seokju Lee, Francois Rameau, **Fei Pan**, In So Kweon. *Attentive and Contrastive Learning for Joint Depth and Motion Field Estimation. ICCV, 2021.* Paper | Code

[11] Inkyu Shin, Sanghyun Woo, **Fei Pan**, In So Kweon. *Two-phase Pseudo Label Densification for Self-training based Domain Adaptation. ECCV 2020.* Paper

[12] **Fei Pan**, Inkyu Shin, Francois Rameau, Seokju Lee, In So Kweon. *Unsupervised Intra-domain Adaptation for Semantic Segmentation through Self-Supervision. CVPR 2020.* Oral Presentation (Acceptance Rate $< 0.7\%$). Paper | Code

[13] Junsik Kim, Tae-Hyun Oh, Seokju Lee, **Fei Pan**, In So Kweon. *Variational Prototyping-Encoder: One-Shot Learning with Prototypical Images. CVPR, 2019.* Paper | Code

[14] Sanghyuk Park, **Fei Pan**, Sunghun Kang, and Chang D. Yoo. *Driver Drowsiness Detection System Based on Feature Representation Learning Using Various Deep Networks. ACCV, 2016.* Paper