
MICCAI Doctoral Consortium 2025
MSB Doctoral Symposium MICCAI 2025
Toward Next-generation Medical Vision Backbones:
Modeling Finer-grained Long-range Visual Dependency
Mingyuan Meng / 蒙明远 (PhD)
School of Computer Science, The University of Sydney, Australia
Email: mmen2292@uni.sydney.edu.au; Homepage: https://mungomeng.github.io/
PhD Supervisors: Jinman Kim, Dagan Feng, and Lei Bi
MICCAI 2025
THE UNIVERSITY OF SYDNEY
Background
Medical images depict standardized anatomical regions across patients, resulting in high structural similarity with subtle anatomical and pathological characteristics discernible only in high resolutions containing tissue-level image textural details.
Research Gap
❖ CNN limited by intrinsic locality, while modeling long-range visual dependency is crucial for global comprehension of image contexts.
❖ Transformer cannot process high-resolution image features with enriched medical image details.
❖ MLP is a potential alternative to transformer - yet to be recognized in medical imaging community.
Overview of Doctoral Thesis (Too Long? Just Look at Here!)
Research Problem & Motivation:
Innovate transformer/MLP-based vision models to effectively capture long-range visual dependency among medical imaging characteristics to enhance diverse medical image computing tasks.
Study I: Transformer for Registration
Research Aim: (i) Innovate transformer-based methods for medical image registration; (ii) Explore insights into long-range visual dependency modeling in the scenario of registration.
Related Publication: MICCAI 2022, MICCAI 2023 Pattern Recognition 2024
Study II: Transformer for Seg & Prognosis
Research Aim: (i) Innovate transformer-based methods for multi-modal imaging analysis; (ii) Explore insights into the modeling of complementary cross-modal visual dependency.
Related Publication: IEEE JBHI 2022, MICCAI 2023 npj Precision Oncology 2024
Study III: MLP for Registration
Research Aim: (i) Introduce MLP-based coarse-to-fine network for medical registration; (ii) Uncover the potential of MLPs in modeling fine-grained long-range visual dependency.
Related Publication: CVPR 2024 (Best Paper Candidate/Finalist, Top 24)
Study IV: Empirical Investigation of MLP
Research Aim: (i) Investigate the employment of MLP in a variety of medical vision tasks; (ii) Validate the superiority of MLPs over CNNs and transformers in modeling medical dependency.
Related Publication: MICCAI 2024 MedIA (Under Review)
Key Finding & Conclusion:
MLP-based models provide feasibility in modeling finer-grained long-range dependency among subtle medical visual details, establishing it as a superior paradigm over transformers/CNNs for medical vision.
Study I – Transformer for Registration
Feature Learning Encoder
Relevance Modeling Decoder
❖ A transformer-based model (named NICE-Trans) for medical image registration enhanced by long-range spatial dependency modeling.
❖ Embed transformers into a non-iterative coarse-to-fine architecture to progressively model coarse-to-fine spatial dependency in a single iteration, marking the first adopt of transformers into non-iterative coarse-to-fine registration.
❖ Joint affine and deformable image registration in a single network.
❖ Reveal key insights into how long-range dependency modeling enhances medical image registration.
Study II – Transformer for Seg & Prognosis
PET-CT Merging Encoder
PT-MLN Diverging Decoder
❖ A transformer-based model (name XSurv) for joint survival prediction and tumor segmentation from multi-modality PET-CT images.
❖ A new paradigm for multi-modal survival modeling with cross-modal feature interaction enhanced by long-range dependency modeling.
❖ A Hybrid Parallel Cross-Attention (HPCA) block was introduced to explicitly model complementary cross-modal dependency via cross-attention transformers.
❖ Simultaneously extract modality-specific prognostic features while sharing contextual knowledge across modalities.
Study III – MLP for Registration
❖ A correlation-aware coarse-to-fine MLP-based framework (named CorrMLP) for deformable medical image registration.
❖ First MLP-based coarse-to-fine registration network unlocking the potential of MLPs for capturing pixel-wise spatial dependency.
❖ First MLP block designed to model correlation-aware multi-range visual dependency for medical image registration.
❖ A novel correlation-aware registration architecture leveraging both image-level and step-level correlations.
❖ Preliminarily establish MLP as promising backbones for medical vision tasks due to its capability to capture fine-grained pixel-wise long-range dependency.
Study IV – Empirical Investigation of MLP
❖ A comprehensive empirical study evaluating MLP on medical image reconstruction, registration, and segmentation.
❖ Regardless of the specific MLP variants, employing MLP blocks at full resolution consistently enabled superior performance over CNN- and transformer-based methods across all evaluation tasks, even outperforming task-optimized specialist models.
❖ Uncover the under-explored capability of MLP to capture fine-grained long-range dependency among subtle medical visual details.
❖ Step toward next-generation medical vision backbones equipped with finer-grained long-range medical vision dependency modeling.