Multimodal Prototyping for Cancer Survival Prediction

Andrew H. Song[1], Richard J. Chen[1], Guillaume Jaume[1], Anurag J. Vaidya[1], Alexander S. Baras[2], Faisal Mahmood[1]
[1]Mass General Brigham [2]Johns Hopkins University
asong@bwh.harvard.edu, faisal.mahmood@bwh.harvard.edu

Mass General Brigham · HARVARD MEDICAL SCHOOL · Mahmood Lab AI for Pathology

ICML International Conference On Machine Learning

Motivation

Multimodal Pathology [1]

- Patient-level outcome prediction from fusion of complementary biological information – e.g.) Tissue **Morphology** + **Transcriptomics** expression

Morphology

- **Digitized tissue sections** (whole-slide images, WSIs), of up to ***100,000 x 100,000*** pixels (at 0.5μm/pixel)
- Typically tokenized into > 10,000 patch tokens (256 x 256 pixels)

Transcriptomics (Genes)

- Whole-Transcriptome RNA-sequencing provides expressions for > 20,000 genes (tokens)

Limitations of current multimodal approaches

- Multimodal fusion of large sets of token embeddings often leads to
  → Computationally-infeasible training
  → Unstable training dynamics for survival prediction

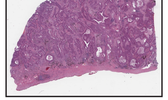
Can we create a *token-efficient multimodal framework*?

Multimodal patient representation

Morphology

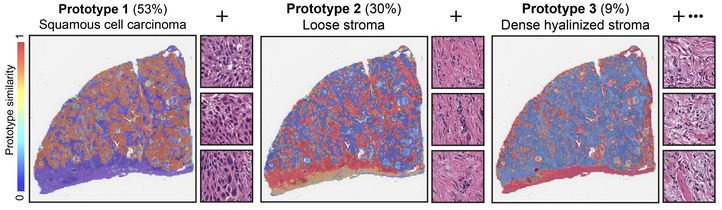
Redundant morphological information in WSI

- Handful of morphological patterns repeated throughout the tissue (e.g., cancer cells, stroma, adipose tissue)



Prototype-based summarization of WSI

- **WSI** ≅ Distribution of ***morphological concepts***
- Summarization of WSI based on two important conditions
  - Feature representation of each concept
  - Cardinality (proportion) of each concept in WSI
- Huge compression
  - Prototypes ($C = 8 \sim 32$) ≪ patches per WSI ($N \sim 10^4$)
- Optimal transport or Gaussian mixture models [2]



Transcriptomics

Only sparse set of genes are relevant for cancer

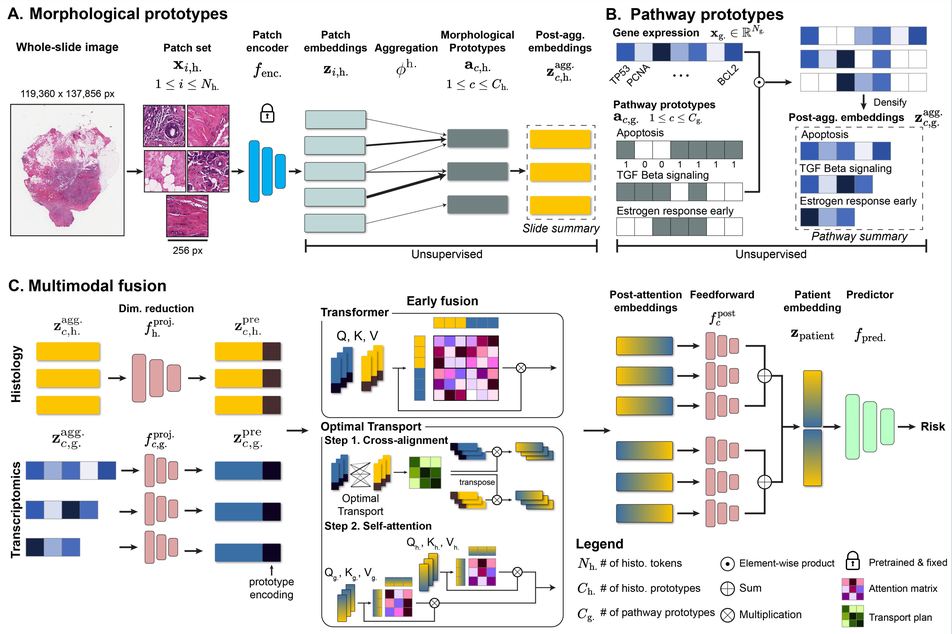
↓

Prototype-based summarization (pathway) of transcriptomics

- **Pathways** provide natural functional grouping of genes
  → **Example Pathway:** HALLMARK_APOTOSIS
  → **Associated genes** (161 genes): [ADD1, AIFM3, …, WEE1, XIAP]
- **Pathways** can be effectively treated as ***prototypes***
  → **Semantic group**: Apoptosis = Programmed cell death
  → **Compression**: > 20,000 genes ⇒ A functional group of 200 genes

⟹ **MMP**: **M**ulti**M**odal **P**rototyping for survival prediction

MMP for multimodal cancer survival prediction



Morphology: Patch embedding distribution (Gaussian Mixture Model)

- $p(\boldsymbol{z}_n;\theta) = \sum_{c=1}^{C} \boldsymbol{\pi}_c \cdot N(\boldsymbol{z}_n;\boldsymbol{\mu}_c,\Sigma_c) \Longrightarrow$ Each component: a prototype and its distribution
- **Morphology prototype set**: $\boldsymbol{z}^{\text{pre}}_{\text{histo}} = \{\boldsymbol{z}^{\text{pre}}_{\text{c, histo}}\}_{c=1}^{C} \Longrightarrow \boldsymbol{z}^{\text{pre}}_{\text{c, histo}} = \boldsymbol{f}([\widehat{\boldsymbol{\pi}}_c, \widehat{\boldsymbol{\mu}}_c, \widehat{\Sigma}_c]) \in \mathbb{R}^{\mathbf{d}}$

Transcriptomics: 50 Hallmark functional gene sets (Pathways)

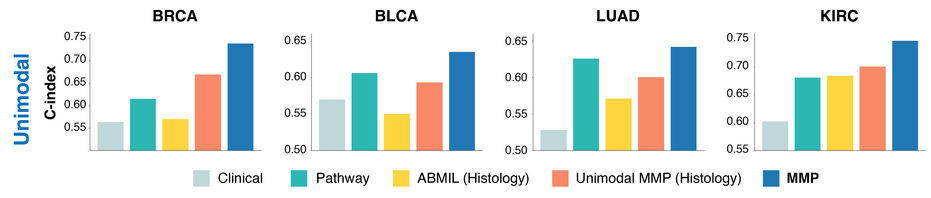
- **Pathway prototype set**: $\boldsymbol{z}^{\text{pre}}_{\text{gene}} = \{\boldsymbol{z}^{\text{pre}}_{\text{c, gene}}\}_{c=1}^{50} \Longrightarrow \boldsymbol{z}^{\text{pre}}_{\text{c, gene}} \in \mathbb{R}^{\mathbf{d}}$

Multimodal early fusion (Morphology prototypes + Pathway prototypes)

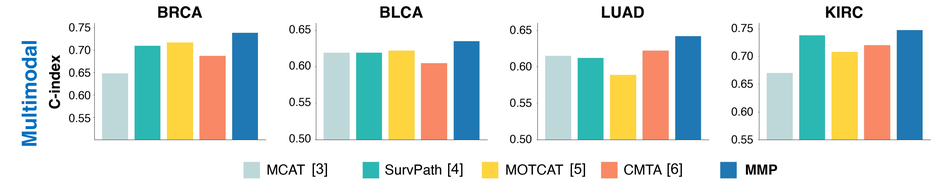
- Transformer-based fusion or Optimal Transport-based fusion

MMP for slide-level survival prediction

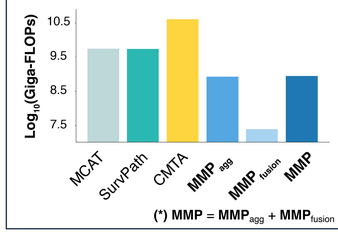
- **Multimodal** > **Unimodal**: Complementary information benefits prognosis



- **Prototype** > **non-prototype multimodal baselines**: Smaller token set helps performance



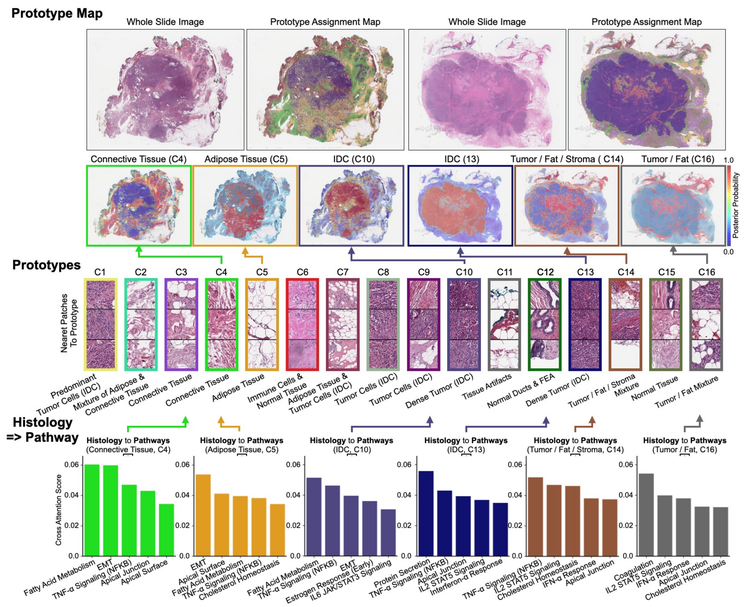
MMP computational complexity



Huge reduction in # of morphology tokens
⟹ 5 ~ 10x less computation overall

Mini-batch of patients for survival training
⟹ More stable training vs. non-prototype baselines with mini-batch of single patient
⟹ Allows for *Cox loss* function

MMP for Interpretability



Prototype-oriented interpretability

- Visualization of the most similar prototype on WSI (Cluster map)
- Each prototype represents different morphological concepts

Cross-modal interpretability

- Bidirectional (Histology → Pathway, Pathway → Histology)
- Intuitive interpretability based on prototypes in both domains

References

- [1] Song AH et al., Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering*, 2023
- [2] Song AH et al., Morphological prototyping for unsupervised slide representation learning in computational pathology. *CVPR*, 2024
- [3] Chen RJ et al., Multimodal co-attention transformer for survival prediction in gigapixel whole slide images, *ICCV*, 2021
- [4] Jaume G et al., Modeling dense multimodal interactions between biological pathways and histology for survival prediction, *CVPR*, 2024
- [5] Xu Y et al., Multimodal optimal transport-based co-attention transformer with global consistency for survival prediction, *ICCV*, 2023
- [6] Zhou F et al., Cross-modal translation and alignment for survival analysis, *ICCV*, 2023

PAPER

CODE