# Supplementary Material for: Cross-Modal Translation and Alignment for Survival Analysis

Fengtao Zhou[1] Hao Chen[1,2,*]
[1]Department of Computer Science and Engineering
[2]Department of Chemical and Biological Engineering
The Hong Kong University of Science and Technology
fzhouaf@connect.ust.hk jhc@cse.ust.hk

## 1. Implementation Details

Our proposed Cross-Modal Translation and Alignment (CMTA) framework and all compared methods are implemented in PyTorch 1.12.0 and Python 3.9.12. The source code has been released[†]. These methods are trained and validated on a high-performance workstation with 8 NVIDIA RTX A6000 GPUs. To avoid the experimental results fluctuation caused by randomness, random seeds of PyTorch and NumPy are set to 1.

As for our proposed model, all patch features are fed into the 1024-256-256 fully connected layers to obtain 256-dimension embeddings. Meanwhile, genetic features are also fed into the 1024-256 fully connected layers to obtain 256-dimension embeddings. There is one manually tunable hyper-parameter $\alpha$ which is used for controlling the contribution of alignment constrains. Due to the intrinsic differences among different datasets, we need to set different $\alpha$ for different datasets. $\alpha$ is set to 1.0 for BLCA, GBMLGG and UCEC datasets, $10^2$ for BRCA dataset and $10^{-4}$ for LUAD dataset. During training, we utilized SGD optimization with a learning rate of 0.001, weight decay of $1 \times 10^{-5}$ and momentum of 0.9. This network is trained for 30 epochs in total.

## 2. Impacts of Patch Size

In WSI analysis, 256×256 patch at 20× magnification (equivalent to 512×512 patch at 40× magnification) is the most common setting. In this section, we further conduct some experiments to explore the impacts of magnification levels for different cancers, as shown in Table 1. The overall performance at 20× magnification level is superior to results at other magnification levels. However, some specific cancers are not sensitive to magnification levels, such as GBMLGG. That is because the five-years survival and ten-years survival are strongly related to some specific genes. In contrast, some specific cancers are obviously sensitive to magnification levels, such as BLCA and BRCA. That means the survival prediction for these kinds of cancers mainly depend on the phenotypes in pathological images.

| Size | Datasets | | | | |
|---|---|---|---|---|---|
| | BLCA | BRCA | GBMLGG | LUAD | UCEC |
| 512 | 0.6910 | 0.6679 | 0.8531 | 0.6864 | 0.6975 |
| 1024 | 0.6673 | 0.6717 | 0.8530 | 0.6724 | 0.6892 |
| 2048 | 0.6719 | 0.6741 | 0.8532 | 0.6891 | 0.6949 |

Table 1. Impacts of patch size cropped from whole slide image.

## 3. Visualization

There are two kinds of crucial feature representations in our model, *i.e.*, intra-modal representation (*i.e.,* $p$ and $g$) and cross-modal representation (*i.e.,* $\hat{p}$ and $\hat{g}$). The former is learned and aggregated form single modality. The latter is learned and translated from modal-related information. In order to illustrate the correlations intra-modal representation and cross-modal representation, we pick some representative samples from each dataset and draw the heatmap for each sample. The visualization results are shown in Figure 1-5. For genomic profiles, cross-modal representation is calculated from genomic-related information in pathomic features. As we can see from all figures, the regions with high attention scores are not completely overlapping between intra-modal representation and cross-modal representation, even have an obvious clear gap in some specific datasets. That means, not all patches in pathological images are related to genomic profiles. This phenomenon also demonstrated our claim in introduction section that leveraging genomic profiles as guidance to aggregate features in pathological images would discard some important information irrelevant to gene expression.

*Corresponding author
[†]https://github.com/FT-ZHOU-ZZZ/CMTA

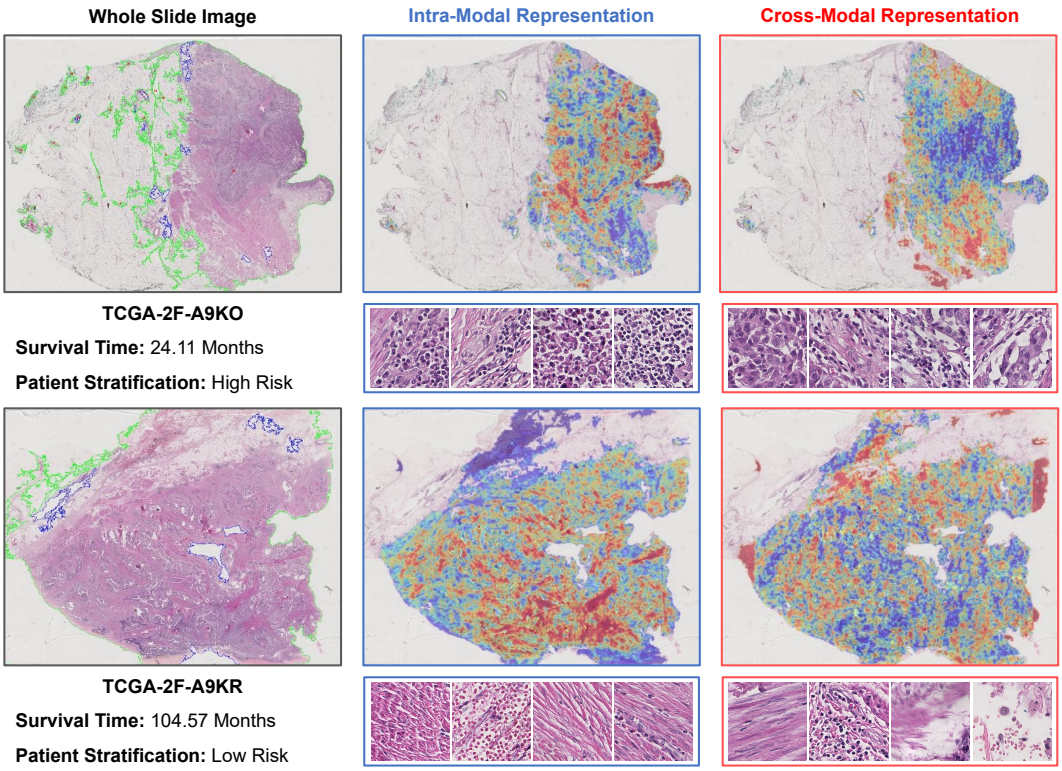


Figure 1. Visualization Results of Low Risk High Risk Samples in BLCA Dataset.

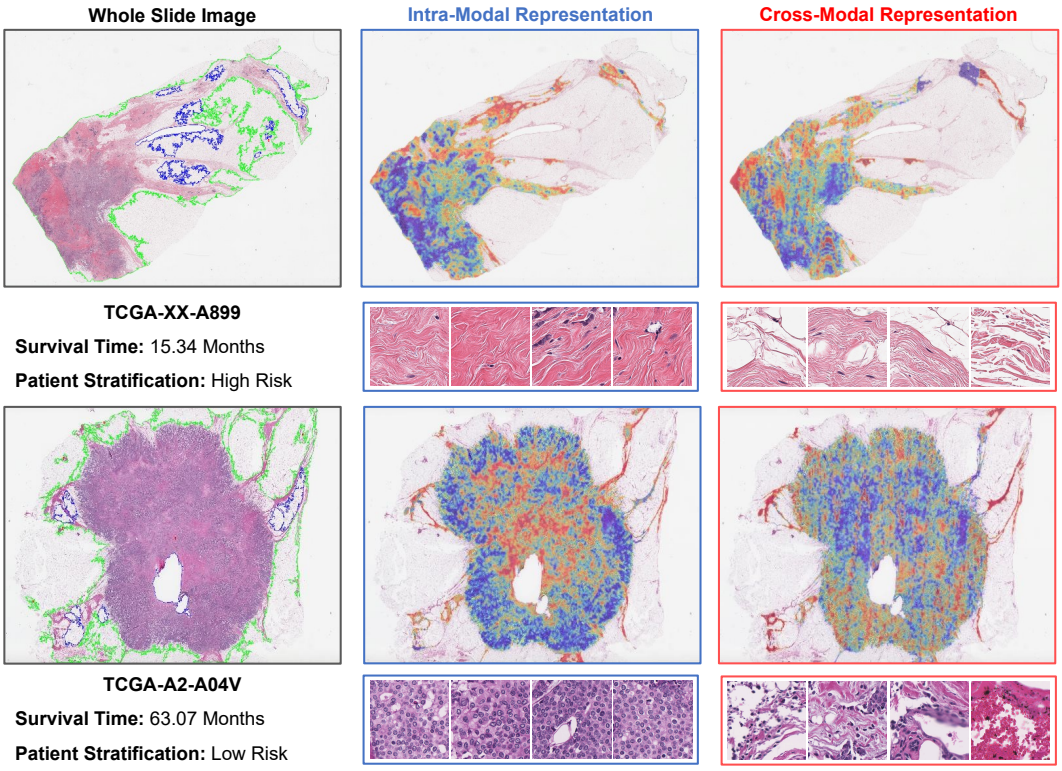


Figure 2. Visualization Results of Low Risk High Risk Samples in BRCA Dataset.

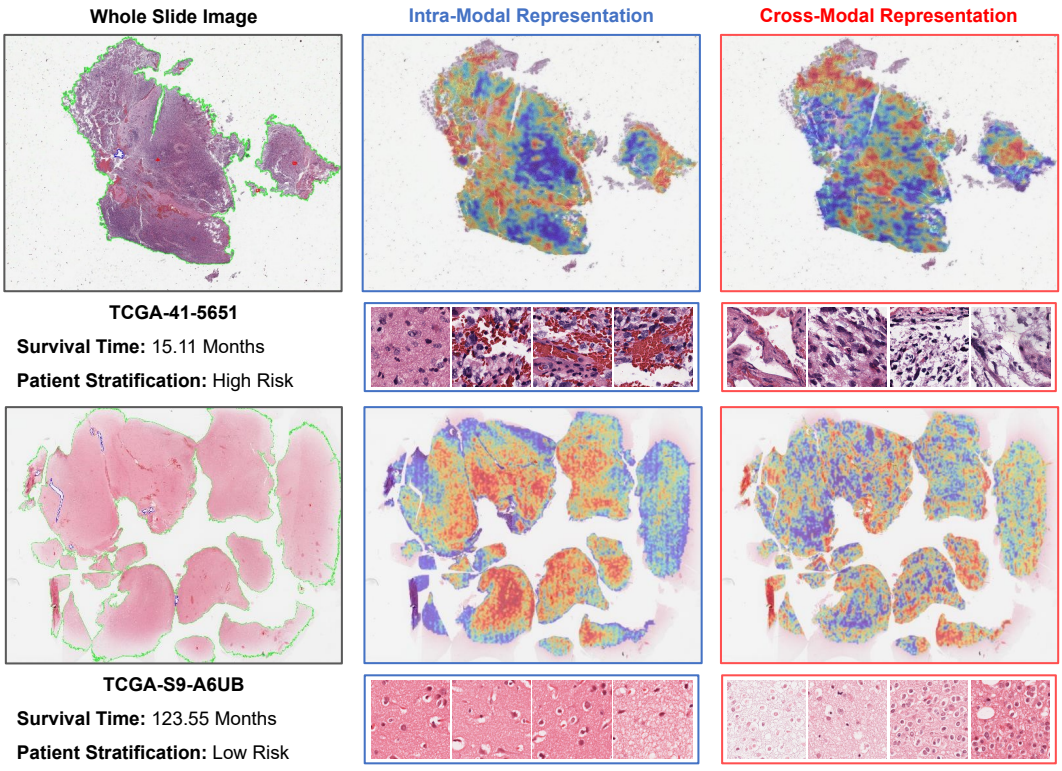


Figure 3. Visualization Results of Low Risk High Risk Samples in GBMLGG Dataset.

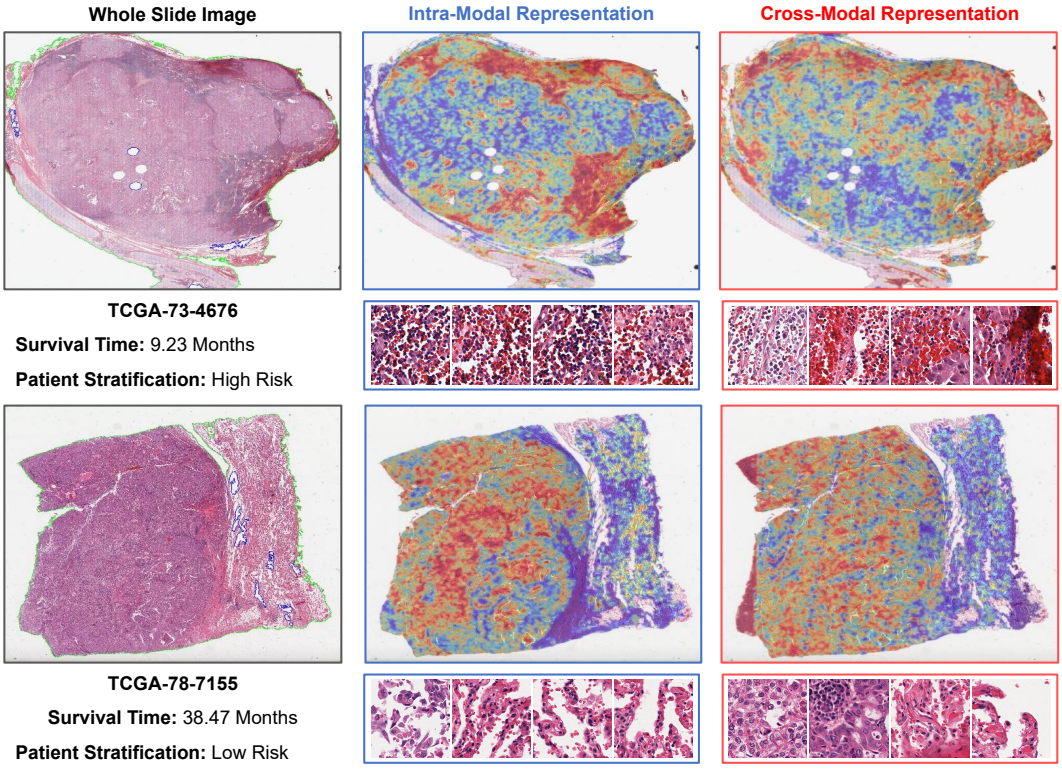


Figure 4. Visualization Results of Low Risk High Risk Samples in LUAD Dataset.

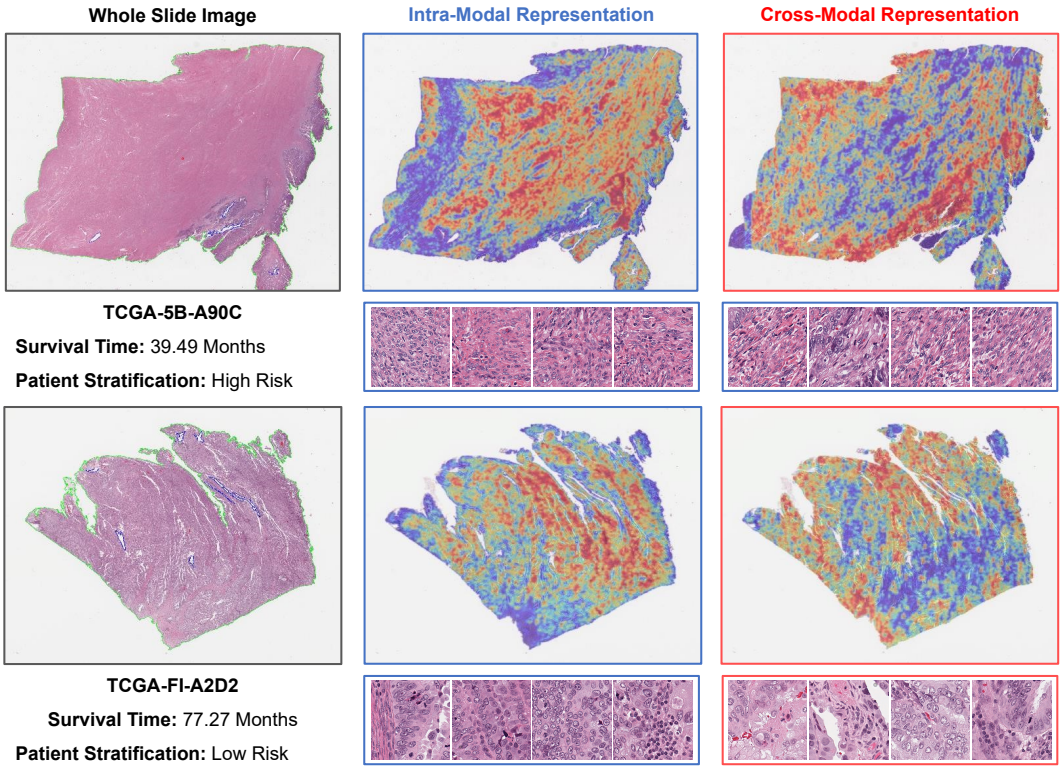


Figure 5. Visualization Results of Low Risk High Risk Samples in UCEC Dataset.