HistGen: Histopathology Report Generation via Local-Global Feature Encoding and Cross-modal Context Interaction Supplementary Materials

Zhengrui Guo, Jiabo Ma, Yingxue Xu, Yihui Wang, Liansheng Wang, and Hao Chen

Prompt for TCGA Diagnostic Report Cleaning using GPT-4
For report data preprocessing with GPT-4, this prompt is used: Help me check the formatting and spelling of the supplied pathology report, including incorrect use of punctuation like misusing of 'x' and 'X', and capitalization as well as deletion of some words of unclear meaning. It checks formatting and spelling while preserving the report's meaning, ensuring clinical correctness in the final report.

Fig. 1: Prompt for Report Data Preprocessing using GPT-4. This prompt checks formatting and spelling while preserving the report's diagnostic meaning, ensuring clinical correctness in the final report.

Table 1: Composition of the curated WSI-Report dataset. The dataset structure is detailed as follows, with each subset corresponding to a distinct disease type. Raw reports and more information on different subsets could be sourced from the TCGA platform. Notably, to guarantee each WSI corresponds to the report description, cases where one report matches multiple WSIs are excluded.

Cancer Type	BRCA	UCEC	KIRC	THCA	LGG	LUAD	HNSC	LUSC
# of Case	999	503	501	486	444	444	442	442
Cancer Type	COAD	PRAD	BLCA	STAD	LIHC	KIRP	CESC	GBM
# of Case	412	368	353	337	329	265	251	245
Cancer Type	SARC	PAAD	PCPG	READ	ESCA	THYM	KICH	SKCM
# of Case	243	174	171	156	117	113	108	97
Cancer Type	TGCT	MESO	UVM	OV	UCS	ACC	DLBC	CHOL
# of Case	87	67	65	62	56	54	43	37

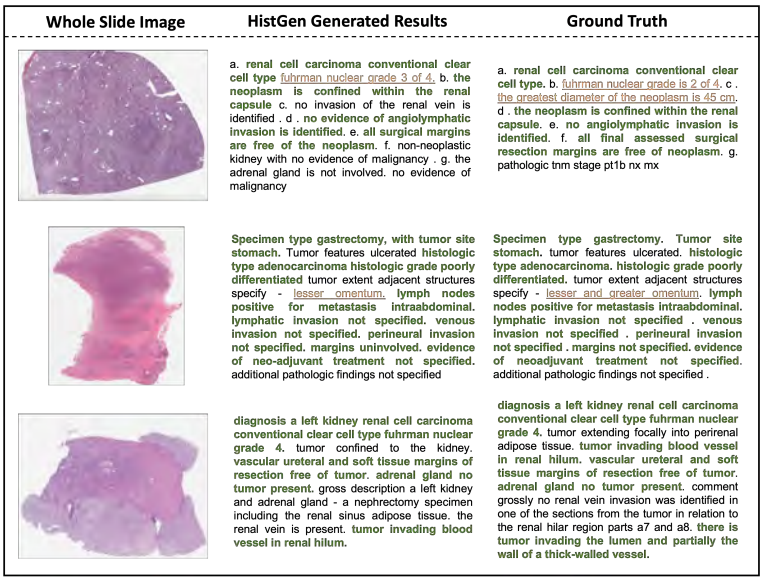
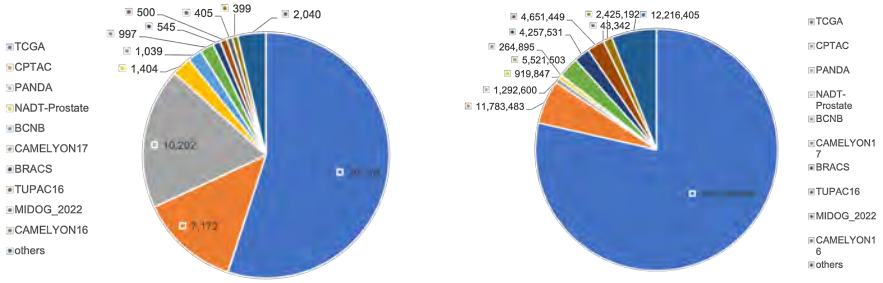


Fig. 2: Qualitative analysis of the proposed HistGen model: Words in bold green indicate alignment between our model's generated results and the ground truth. Words underlined in orange represent diagnostic details our model misses. The first two examples demonstrate the model's superior captioning capability, accurately diagnosing provided WSIs with minor, non-critical differences from the ground truths. In the third example, the model correctly predicts the diagnosis despite lacking detailed context present in the ground truth.



(a) WSI distribution for DINOv2 ViT-L feature extractor pre-training

(b) Patch distribution for DINOv2 ViT-L feature extractor pre-training

Fig. 3: Data distribution for DINOv2 ViT-L feature extractor pre-training. We have collected over 30 different pathology datasets containing over 60 primary sites. Patches are extracted from whole slide images at level 0, with dimensions of 512×512. These patches are subsequently resized to 224×224 for pre-training the feature extractor. Subfigure (a) shows the details of our collected WSIs and subfigure (b) denotes the details of patches (tiled from the above WSIs) used for pre-training.