PRCV 2026 中国·哈尔滨
The 9th Chinese Conference on Pattern Recognition and Computer Vision
22-25, August, 2026, Harbin

Learning Whole Slide Image Representations from Few High-Resolution Patches via Cascaded Dual-Scale Reconstruction

Yujian Liu[1], Yuechuan Lin[1], Dongxu Shen[2], Haoran Li[1], Yutong Wang[1], Qingquan Wang[1], Xiaoli Liu[3]*, Shidang Xu[1]*

[1] South China University of Technology, SCUT-XU Group
[2] Hong Kong University of Science and Technology (Guangzhou)
[3] AiShiWeiLai AI Research

~9 high-resolution patches / WSI

4.5% of the original training patches

+6.3% ACC, **+4.8%** AUC vs full-data baselines

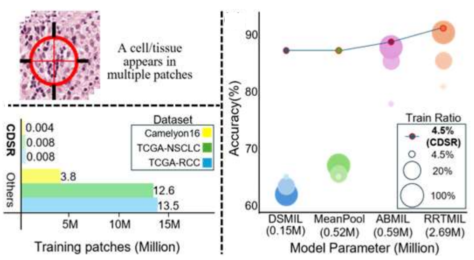
7.4 average training time (vs 41.1 days for SSL)

Whole slide image classification with selective high-resolution patch sampling and local-to-global reconstruction.

1. Motivation

Dense small-patch MIL pipelines for WSI analysis suffer from:

- Domain gap from natural-image pretraining
- Fragmented tissue morphology
- Redundant patches with low information
- Heavy computation and long training time



Why CDSR ?

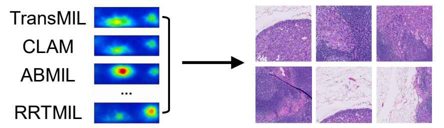
- **Preserve morphology with high-resolution patches**
- **Reduce redundancy by selecting informative and diverse regions**
- **Learn efficient WSI representations compatible with MIL aggregators**

Contributions

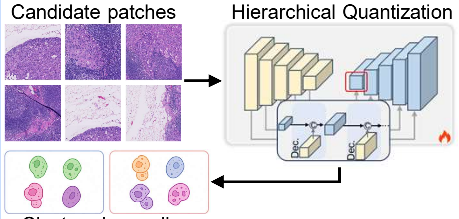
- **Attention-guided candidate selection + QHVAE-based semantic refinement**
- **L2G-Net fuses local morphology and global context through reconstruction**
- **Efficient high-resolution feature learning for slide-level prediction**

2. Method Overview

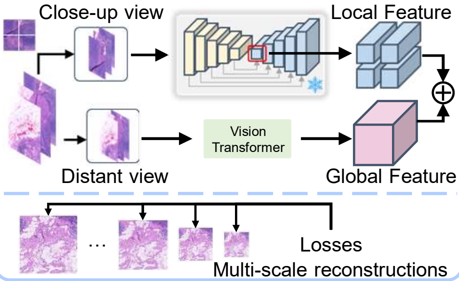
1) Selective high-resolution patch sampling



2) QHVAE semantic refinement



3) Local-to-global reconstruction



Training & Inference

- QHVAE is trained on retained medium-resolution patches (1024×1024).
- L2G-Net is trained on selected high-resolution patches (2048×2048).
- During inference, the decoder is discarded and fused features are aggregated by MIL.

3. Experimental Setup

Datasets	Camelyon-16, TCGA-NSCLC,TCGA-RCC
Metrics	ACC, F1-Score, AUC PSNR, SSIM, LPIPS
Patch Scales	X_s: 256×256; X_m: 1024×1024; X_h: 2048×2048

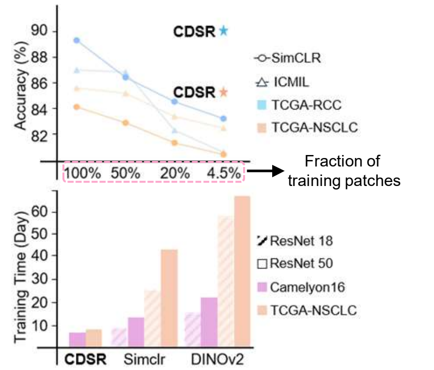
4. Main Results

Slide-level classification performance (best CDSR)

Datasets	ACC (%)	F1-Score (%)	AUC (%)
Camelyon-16	91.5	91.0	96.2
TCGA-NSCLC	85.4	85.3	93.5
TCGA-RCC	90.5	86.9	98.4

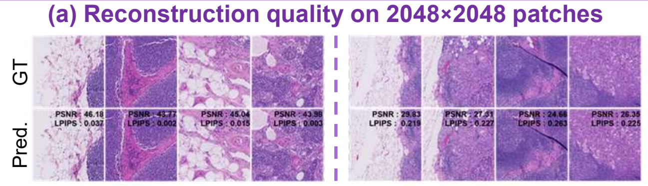
Replacing Res.IN with CDSR yields average gains of **+6.3% ACC, +7.7% F1,** and **+4.8% AUC** across the three datasets.

CDSR uses far less data and trains much faster



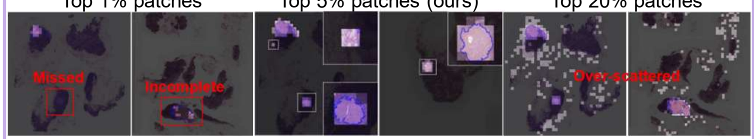
5. Qualitative & Effectiveness Analysis

(a) Reconstruction quality on 2048×2048 patches



QHVAE and L2G-Net preserve visual structure; L2G-Net achieves average PSNR 29.82 dB on 2048×2048 patches (left).

(b) Top-patch visualization by attention ranking



Top 5% attention-ranked patches provide better tumor coverage than top 1% and less noise than top 20%.

Conclusion

- CDSR learns compact high-resolution WSI representations from a small set of informative patches.
- The method preserves morphology, reduces redundancy, and cuts training cost substantially.

SCUT-XU Group | Join Us
Scan to explore our research and open positions.

Using only 4.5% of the training patches and 18% of the training time on average, CDSR remains competitive with strong WSI baselines.