MICCAI 2022 Singapore

Whiterabbit.ai

Deep is a Luxury We Don't Have

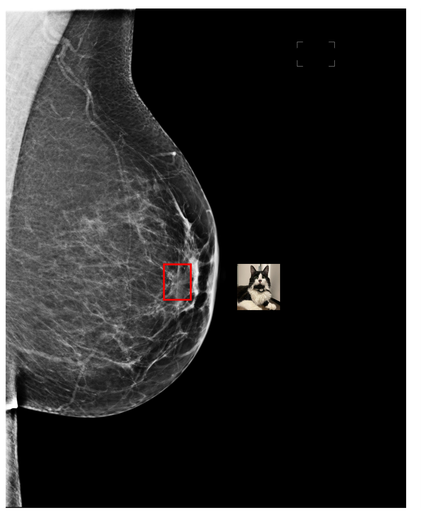
Ahmed Taha, Yen Nhi Truong Vu, Brent Mombourquette, Thomas Paul Matthews, Jason Su, Sadanand Singh

Motivation:

- Medical images have high spatial dimensions, making it difficult to model long-range spatial dependencies.
- To model these dependencies, one can build a deeper network, yet this solution only helps improve the effective receptive field (ERF) sublinearly while significantly increasing the computational cost.

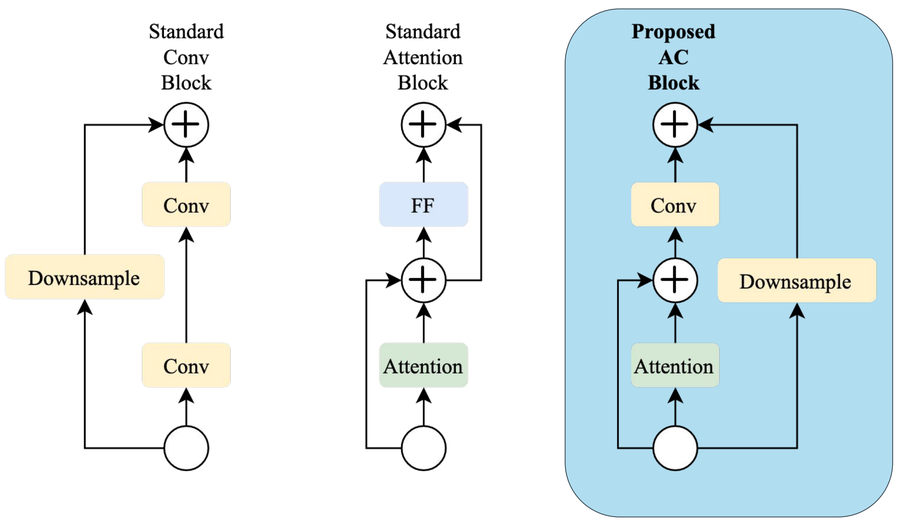
Contributions:

- We introduced **H**igh-resolution **C**onvolution **T**ransformer (HCT), an efficient model that brings long-range reasoning capability to improve classification performance on high resolution medical images.
- We demonstrated HCT's fitness for medical images through its dynamic effective receptive field.

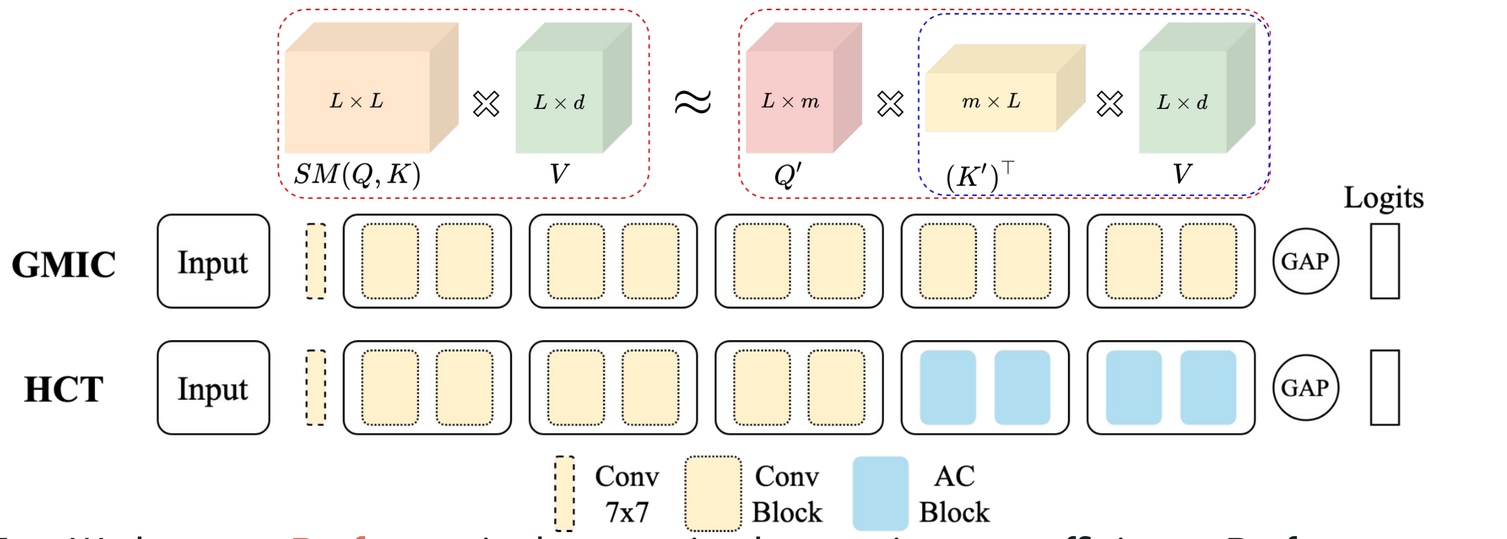


Approximate size difference between a 2D mammogram (3328x2560) and an ImageNet size cat image (224x224). The suspicious finding in the mammogram is as big as the natural image.

HCT: High Resolution Convolution Transformer



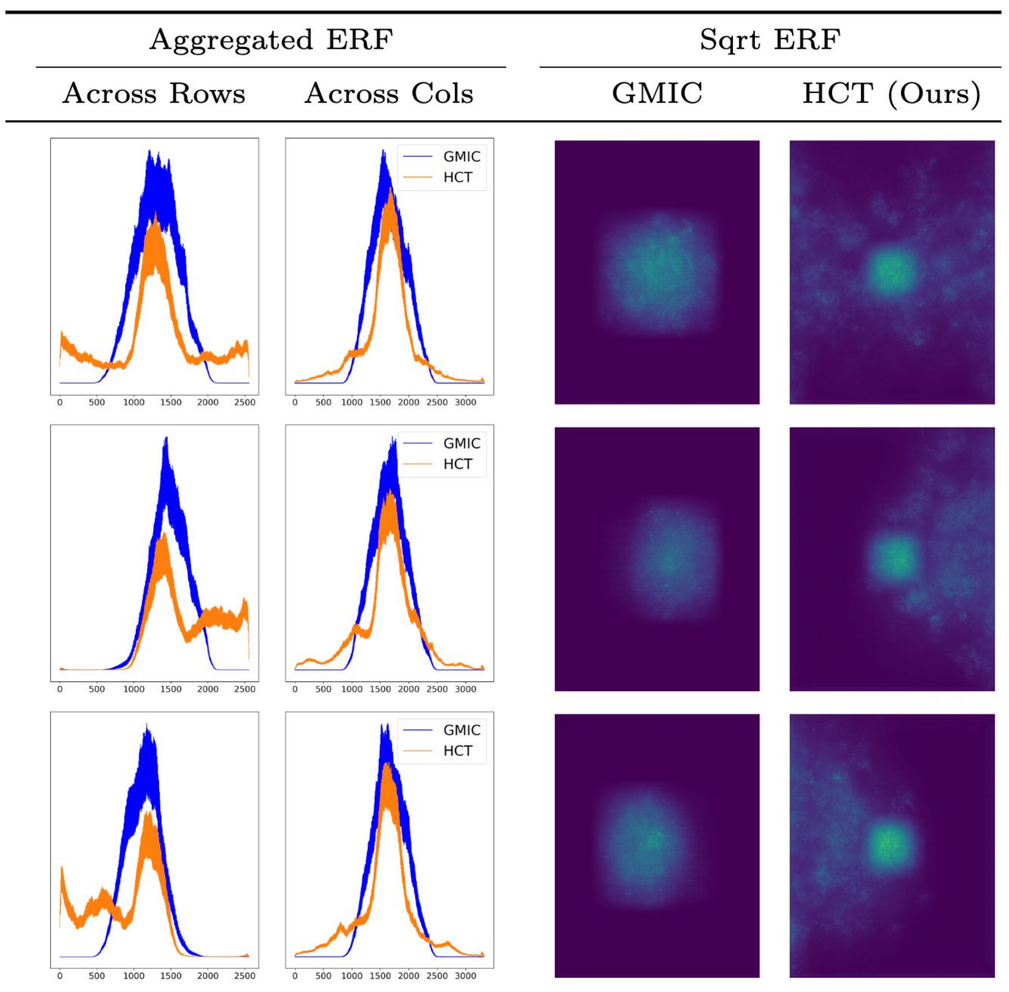
Attention-Convolution (AC) block brings both long-range attention and spatial downsampling capability. AC block consumes and produces feature maps. Thus, this layer integrates seamlessly in various vision architectures designed for different tasks such as classification or detection.



Top: We leverage **Performer** in the attention layer to improve efficiency. Performer uses the kernalization trick to approximate the softmax operation. This allows us to change the order of matrix multiplication and compute self-attention with linear complexity.

Bottom: We integrate **AC block** into GMIC, a ResNet-22 variant developed for mammograms. The result, **HCT**, is an efficient convolutional-transformer architecture.

Effective Receptive Field and Performance Evaluation

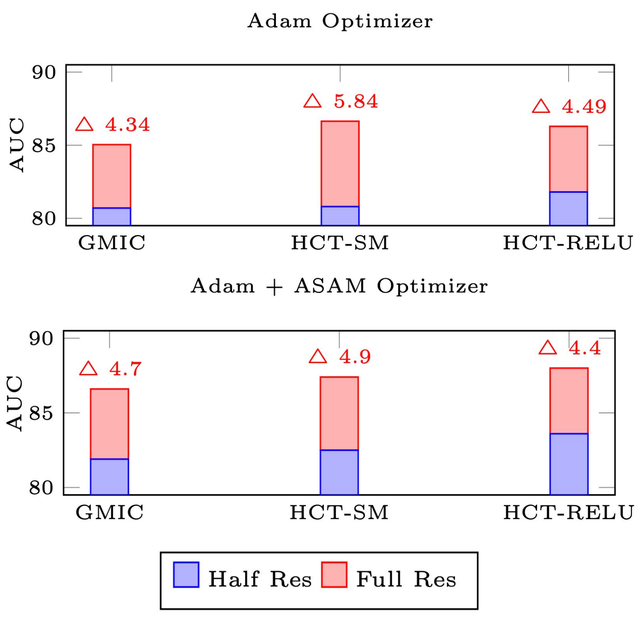


Left: We qualitatively evaluate the ERF of GMIC and HCT 3 times (rows) using (1) 100 random breasts (left and right), (2) 100 random right breasts, and (3) 100 random left breasts. To highlight the ERF difference, we aggregate the ERF across images' rows and columns. GMIC's ERF is rigid and follows a Gaussian distribution. In contrast, **HCT's ERF spreads dynamically within the breast**.

Bottom: Quantitatively, we evaluate HCT using OPTIMAM, a high resolution mammography dataset with 11,215 malignant images and 61,785 negative images. **HCT with Performer attention approximation consistently outperforms GMIC** across different training settings. HCT with Performer-RELU approximation and ASAM optimizer achieves the highest performance of **88.0 AUC.**

Architecture	#Params	Linear Approximation	Patch	Image
			Adam Optimizer	
GMIC	2.80	–	96.13 [95.43, 96.78]	85.04 [83.74, 86.36]
HCT (**ours**)	1.73	Nyströmformer	**96.41 [95.76, 97.01]**	84.83 [83.49, 86.14]
HCT (**ours**)	1.73	Performer-Softmax	96.35 [95.68, 96.97]	**86.64 [85.38, 87.86]**
HCT (**ours**)	1.73	Performer-RELU	96.34 [95.66, 96.97]	86.29 [85.02, 87.54]
			Adam + ASAM Optimizer	
GMIC	2.80	–	96.29 [95.62, 96.92]	86.58 [85.34, 87.80]
HCT (**ours**)	1.73	Nyströmformer	96.65 [96.02, 97.23]	86.73 [85.49, 87.95]
HCT (**ours**)	1.73	Performer-Softmax	96.68 [96.05, 97.26]	87.39 [86.14, 88.59]
HCT (**ours**)	1.73	Performer-RELU	**96.73 [96.09, 97.32]**	**88.00 [86.80, 89.18]**

Image Resolution and Finding Size Studies



Left: We emphasize the **importance of processing medical images at high resolution**. We train GMIC and HCT using half (1664x1280) and full (3256x2560) resolution inputs. Across all models and optimizers, models trained with full-resolution outperform their half-resolution counterparts significantly (+4% AUC).

Right: We evaluate GMIC and HCT on three subsets of malignant images against all negative images. The small, medium, and large subsets contain positive images with respectively the bottom, middle and top ⅓ in terms of malignant finding size. **HCT's long range modeling capability boosts performance on the medium and large finding subset.** HCT achieves a significant improvement of 3.1% on the medium subset. Even on the large subset where GMIC achieves its best performance, HCT-RELU still boosts performance by 1.1%.