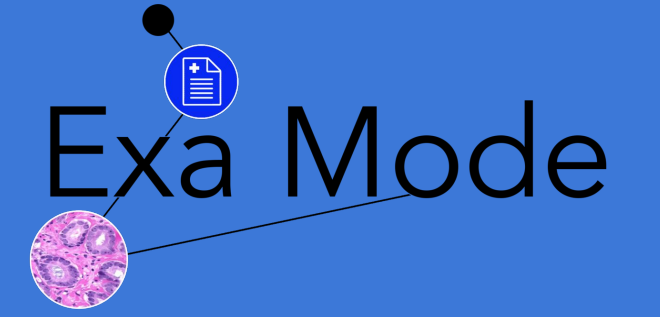


Gigapixel Patch Semantic Segmentation for Histopathology



Ruben Hekster, Damian Podareanu, Joris Mollinga, Jieyi Li



1. Introduction

- Whole Slide Images (WSIs) have very large resolution image files (100,000 x 200,000 pixels) and depict human tissue. They are used for cancer diagnosis in clinical practice.
- These WSIs are annotated by medical experts, creating a mask representing the cancers location. The mask has similarly large dimensionality as the WSI."
- This annotation can also be accomplished using **Deep Learning (DL)** techniques.
- Digital evaluation of entire WSIs is not possible due to **memory constraints**. The common approach is to create patches (squares) which can be evaluated digitally.

This poster focuses on **stepwise increasing the patch size (field of view)** and discusses scalability issues when training a semantic segmentation model with large patch sizes on a distributed system.

2. Positioning with respect to related work

- The Diagnostic Image Analysis Group of the Department of Pathology at the Radboud University Medical Center in Nijmegen, the Netherlands, organised the CAMELYON16 - and the CAMELYON17 challenges (CAncer METastases in LYmph nOdes challenge).
- [1], which won the CAMELYON16 challenge using patches of 256 times 256 pixels and the GoogLeNet architecture, indicated they got better results on the non-downsampled images compared to downsampled.
- Analysis of entire WSI has only been achieved through compression, like in [2].

3. Histopathology Datasets

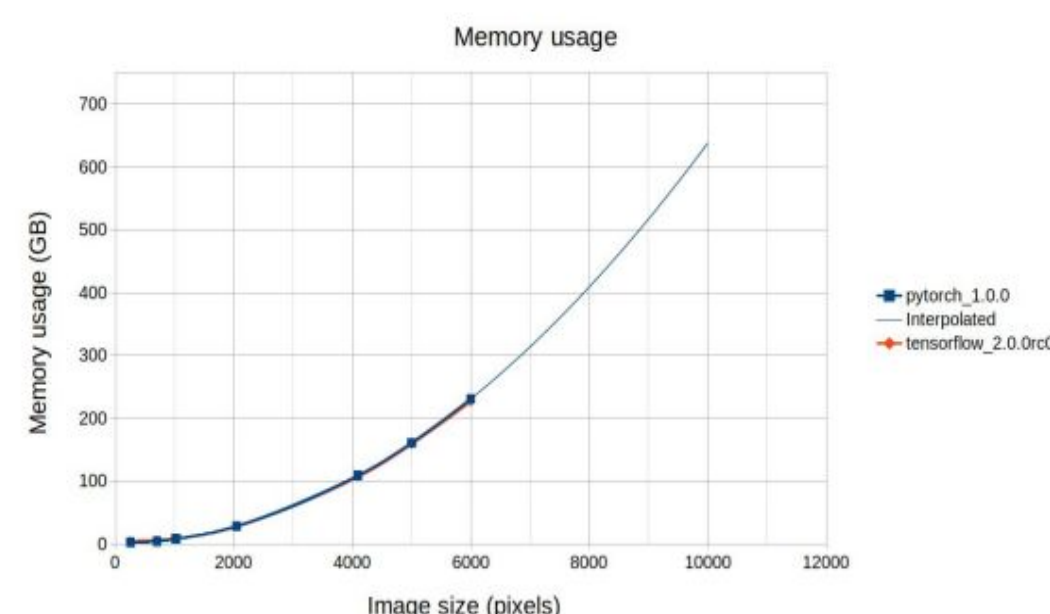
For this research, two common pathology datasets were used, for TIFF format images:

CAMELYON16

- 270 pixel - annotated WSIs.
- Data from two medical centers, average memory footprint of 1.9GB per WSI.

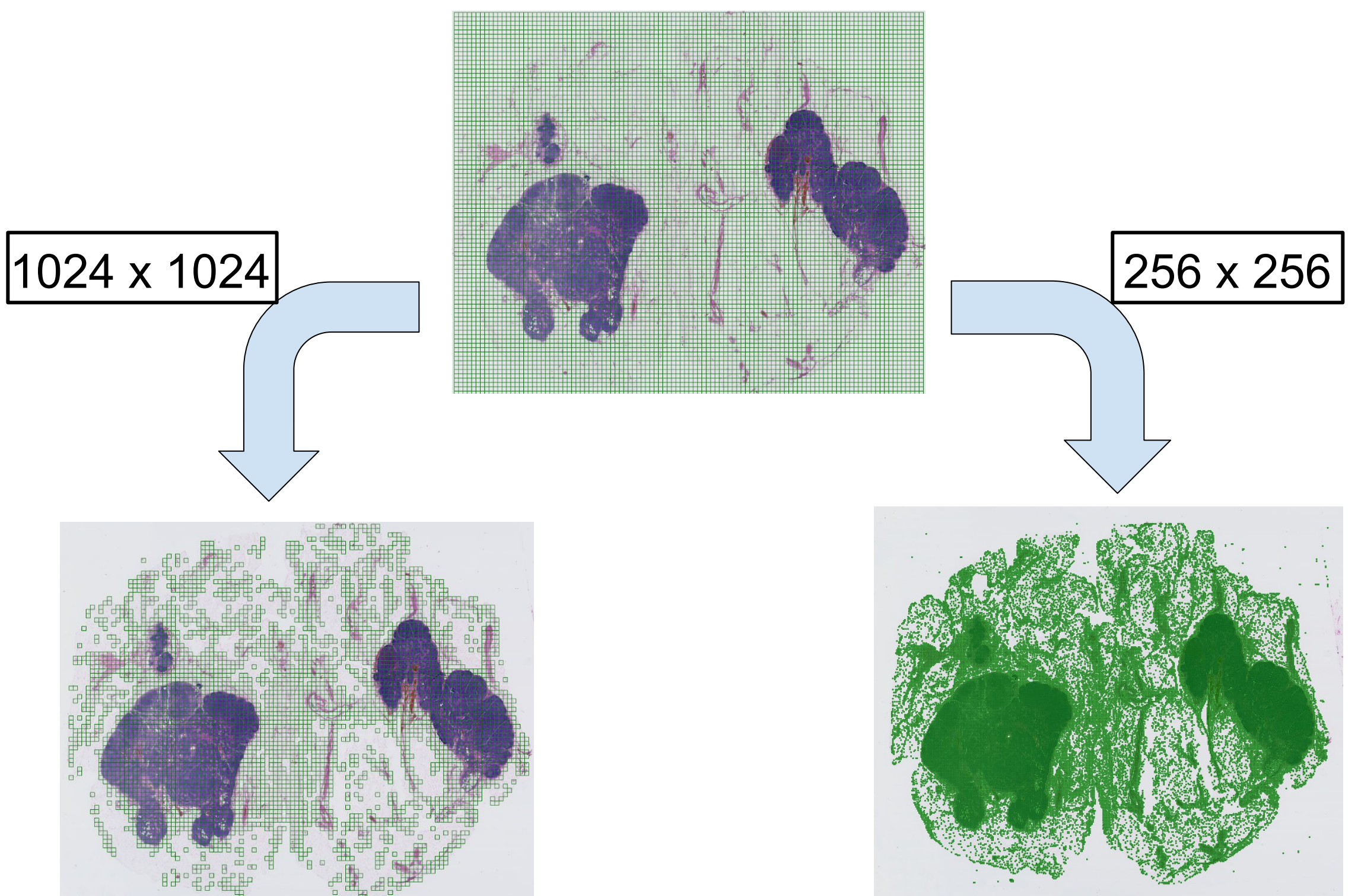
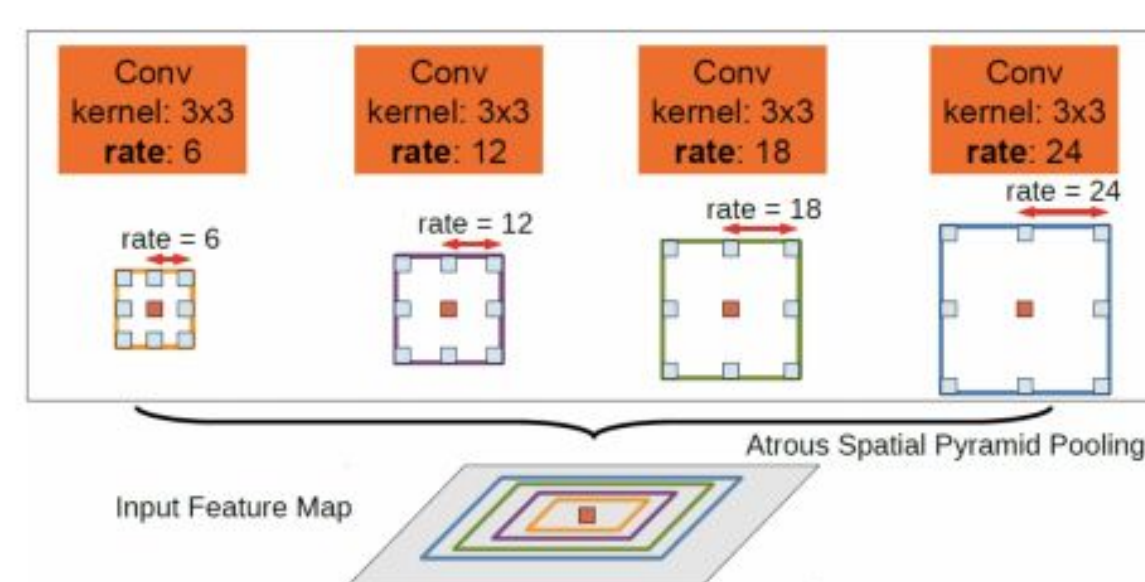
CAMELYON17

- 500 pixel - annotated WSIs.
- Data from five medical centers, average memory footprint of 3.4GB per WSI.
- On the right we can see the relation between the patch size and the memory footprint of model training. Due to memory constraint, at this point we cannot process entire WSI's.



4. Model and training system

- The architecture used is based on the deep neural semantic segmentation network Deeplab V3+, which has proven effective using **Atrous Spatial Pyramid Pooling (ASPP)**. This means that pooling operations exist in the encoder - decoder structure at different scales of receptive fields.
- This reduces computational complexity, while at the same time enables multi - scale analysis of **contextual information**.
- This also makes use of **depthwise separable convolutions**, further reducing the computational requirements.
- The model is trained using four NVIDIA® Titan™ RTX GPU's with 24 GB of GDDR6 memory. The memory consumption for a batch size of 1 at the patch size of 2048x2048 pixels is ~25GB. Hence, for the larger experiments we perform model training on the **Endeavour Supercomputer**.



Example of a Whole - Slide -Image that is divided in smaller patches, representing different field - of - views. On the left side, the patches are 1024 x 1024, and on the right side the patch size is 256 x 256. This enables the evaluation of different types of contextual information, contributing to the pathology classification. As is to be seen, we discard a large part of the patches, due to the patch belonging to the background, and thus not containing any tissue of interest.

5. Results on Memory vs. Accuracy

- Main insight: Enlarging the patch size increases accuracy.
- This shows a **trade - off between memory and accuracy**.

Table 2. CAMELYON16 results for size 256, 704, 1024, 2048

	256	704	1024	2048
Val. Accuracy	0.7756	0.9313	0.9453	0.9410
Val. mIoU	0.6086	0.8647	0.9167	0.9231
Val. AUC	0.7459	0.9259	0.9537	0.9578
Val. dataset size	560	160	92	70

Table 3. CAMELYON17 results for size 256, 704, 1024, 2048, from different medical centers

	256	704	1024	2048
Training centers: all				
Val. accuracy	0.7982	0.9364	0.9566	0.9502
Val. mIoU	0.6489	0.8734	0.9102	0.9268
Val. center: all				
Val. AUC	0.7812	0.9344	0.9560	0.9643
Val. dataset size	552	180	104	76

6. Distributed training on Endeavour

- Our largest training run featured 256 dual-socket Intel Xeon Platinum 8260 nodes, each with 192 GB of memory, interconnected by a low-latency 100GB/s Intel OmniPath network. We used Intel MKL and Intel MKL-DNN for efficient execution for the CPU runs.
- In order to maximize per-node throughput, we used 2 training processes per node, each pinned to a separate socket, effectively performing distributed training within a node. We note weak scaling characteristics for scale-out experiments to 256 nodes. We used Horovod [3] all - reduce, using floating-point 16 compression.

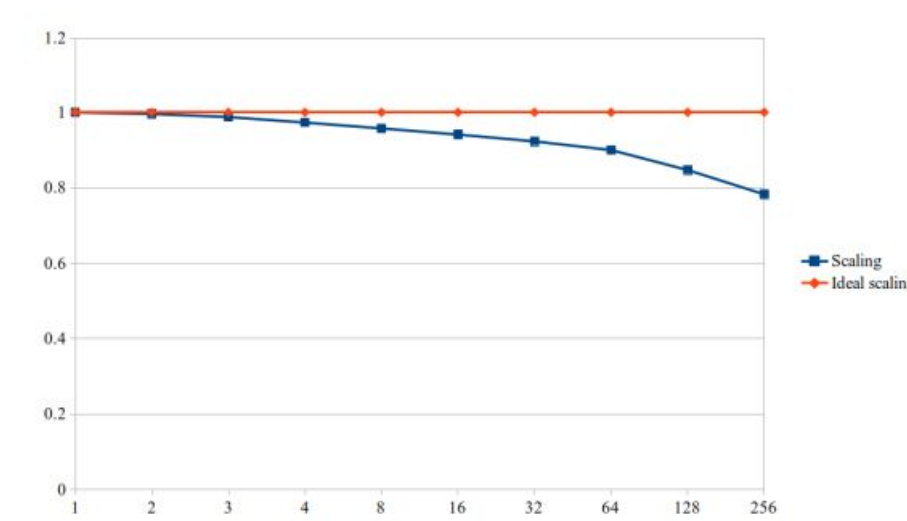


Fig. 5. Weak Scaling on 2048 patch size

7. Conclusion

- Multi - scale contextual information is beneficial for detecting small region-of-interests in large Whole - Slide - Images.
- Patch size (Field of View) can be translated to a **memory footprint of both input data in combination with model activation maps and gradients**. This can be **traded for accuracy on a pixel level**, and on a macro level measuring evaluation metrics as mIoU and AUC.
- Scale-out experiments show that **multi-worker model-parallelism** is needed to process gigapixel size medical imaging from high quality sensing equipment.

References

- Wang, D., Khosla, A., Gargeya, R., Irshad, H., Beck, A.H.: Deep learning for identifying metastatic breast cancer (2016)
- Tellez, D., Litjens, G., van der Laak, J., Ciompi, F.: Neural image compression for gigapixel histopathology image analysis (2018)
- Sergeev, A., Balso, M.D.: Horovod: fast and easy distributed deep learning in TensorFlow. (2018)