Multiple Instance Captioning: Learning Representations from Histopathology Textbooks and Articles

Jevgenij Gamper
University of Warwick, UK
j.gamper@warwick.ac.uk

Nasir Rajpoot
University of Warwick, UK
n.m.rajpoot@warwick.ac.uk

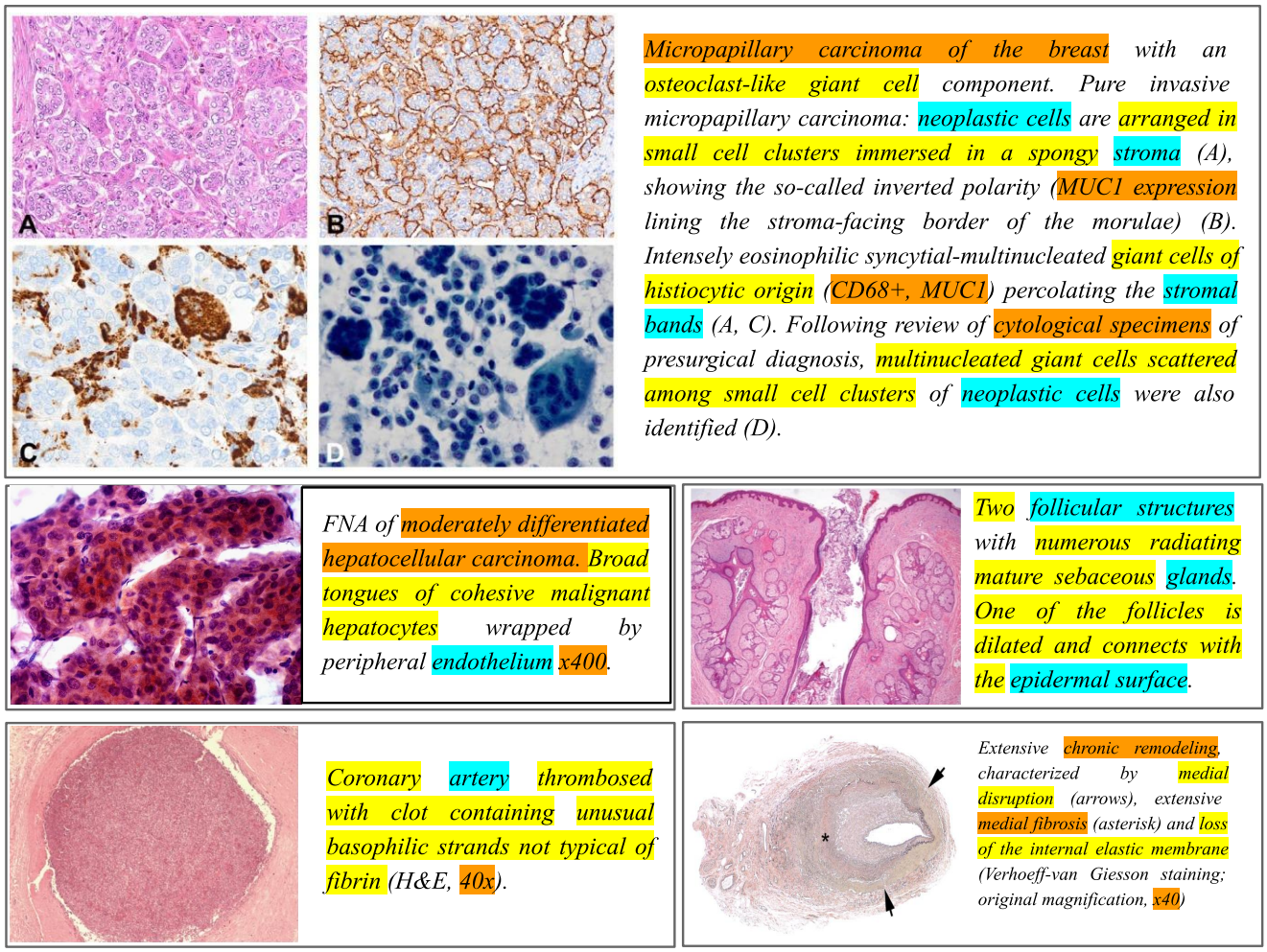
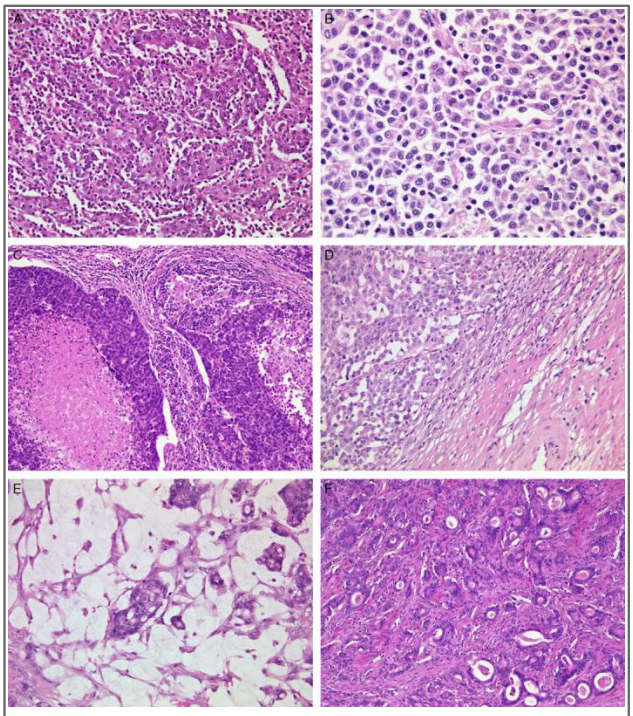


Figure 1: *Top row*: A bag of 4 image instances with a single caption. The bag of images contain three different stains and a cytology image. Labeled in color are examples of common tasks within computational pathology: diagnostic (orange); detection & classification (cyan); descriptive (yellow). *Second row*: Sample on the left is a cytology image with its corresponding caption, while sample on the right is a low power view of skin tissue. *Third row*: Image on the left is a low power view with an interesting label of a blood clot within a vessel. Image on the right is a tissue slide stained with Verhoeff-van Giesson staining.



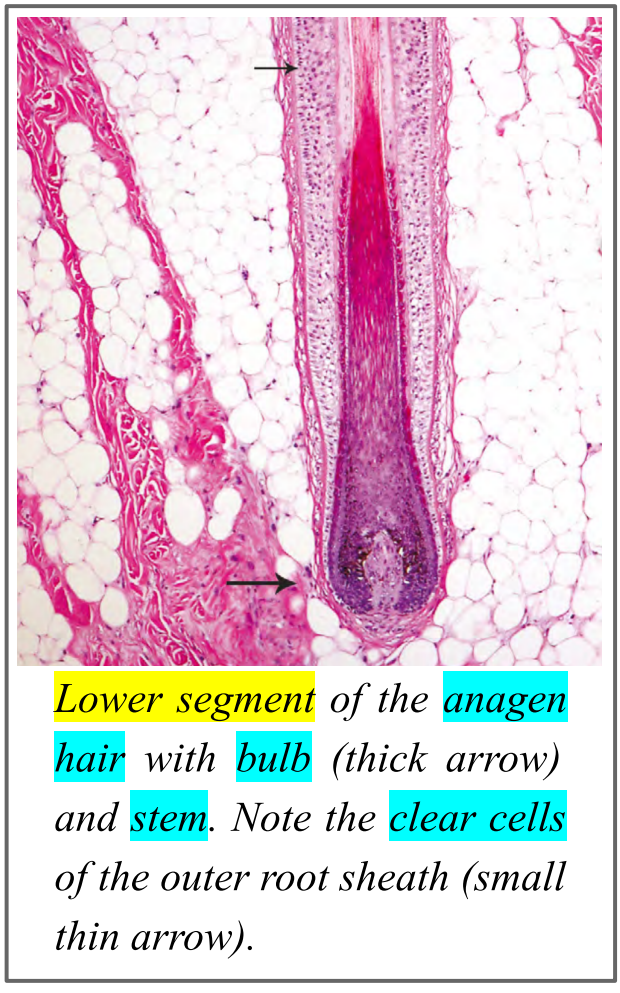
Histomorphology in MSI gastric cancers. Diffuse and solid sheets of tumor cells (A and D) with large amounts of tumor-infiltrating or surrounding lymphocytes (A, C, and D). Solid, nested growth with comedo necrosis (C), pushing margins with lymphocytes aggregating peritumorally (C and D). Some tumors demonstrated a more conventional morphology but then combined >1 pattern in 1 tumor, such as glandular and mucinous (E and F). Hematoxylin and eosin stained. Original magnification: x200.

Figure 2: A bag of 6 image instances with a single caption. Labeled in color are examples of common tasks within computational pathology: diagnostic (orange); detection & classification (cyan); descriptive (yellow); special cell detection (red). Note the diversity of task present within the caption, from identifying the MSI type, to evaluating the density of cells, to comparing the granular vs mucinous samples.

1. Supplementary Material

In the Supplementary material, we include additional samples from the ARCH dataset to demonstrate its diversity of images and supervision signal - Figures 1, 2, 3 and 4. This is to provide additional evidence that no other Computational Pathology (CP) dataset contains as diverse set of labels and images as ARCH. This becomes particularly important in screening tasks, such as anomaly detection, that relies on feature encoder being able to pick up subtle changes within images. As for example in the bottom row on the left in Figure 1 that contains a blood clot within a vessel. Figure 1 also includes samples of cytology images available within ARCH, and together with Figure 4 demonstrates a wide range of stains available within the dataset. Figure 2 shows one of the larger bags contained within ARCH.

We also include a full extension of the Figure 7 in the main text - see Figure 5.



Lower segment of the anagen hair with bulb (thick arrow) and stem. Note the clear cells of the outer root sheath (small thin arrow).

Figure 3: A sample of the pathology of skin tissue.

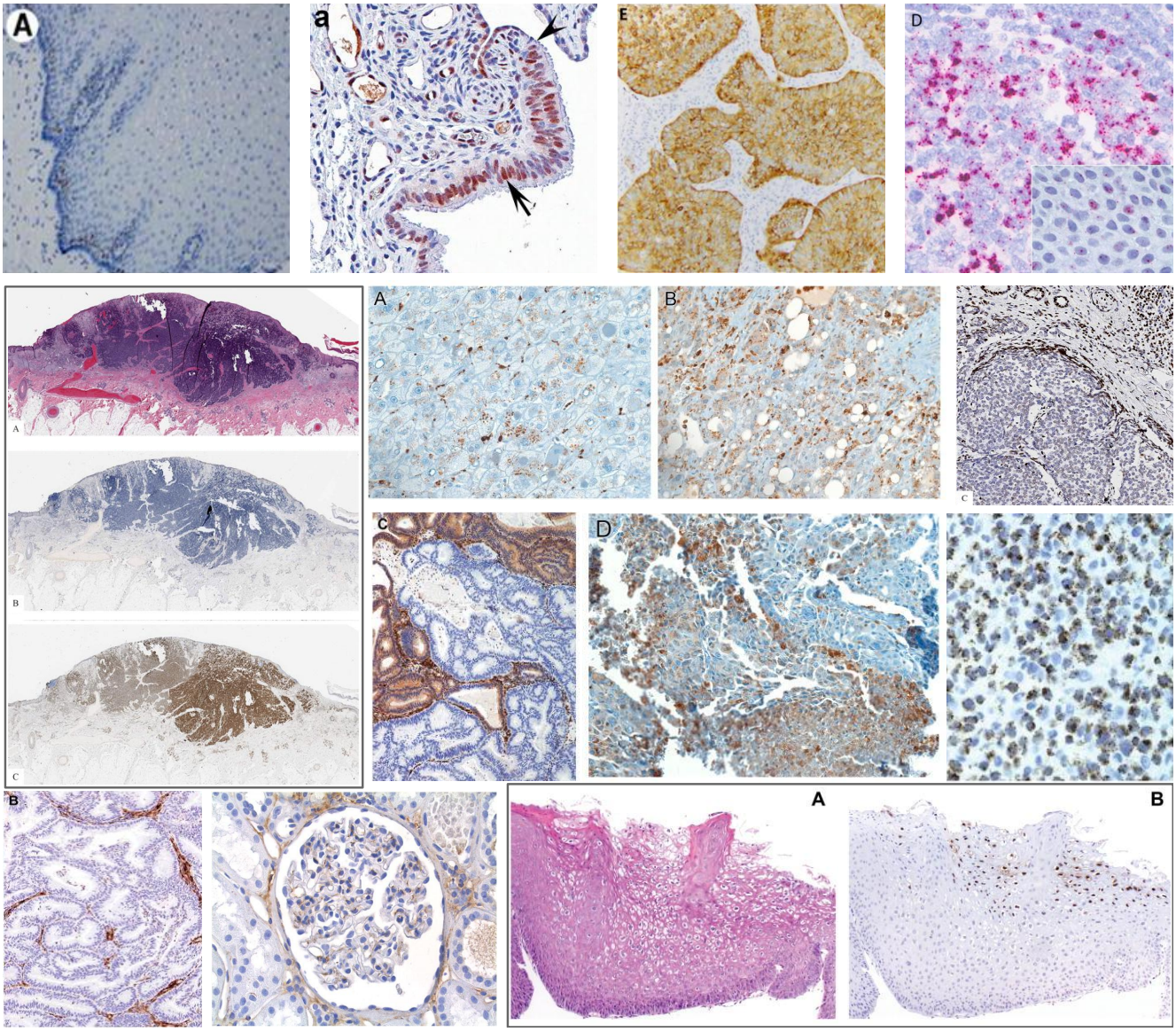


Figure 4: A collage of images to demonstrate the diversity of stain types present in the ARCH dataset. Note the pairs and triples of tissue images that were obtained under a different stain.

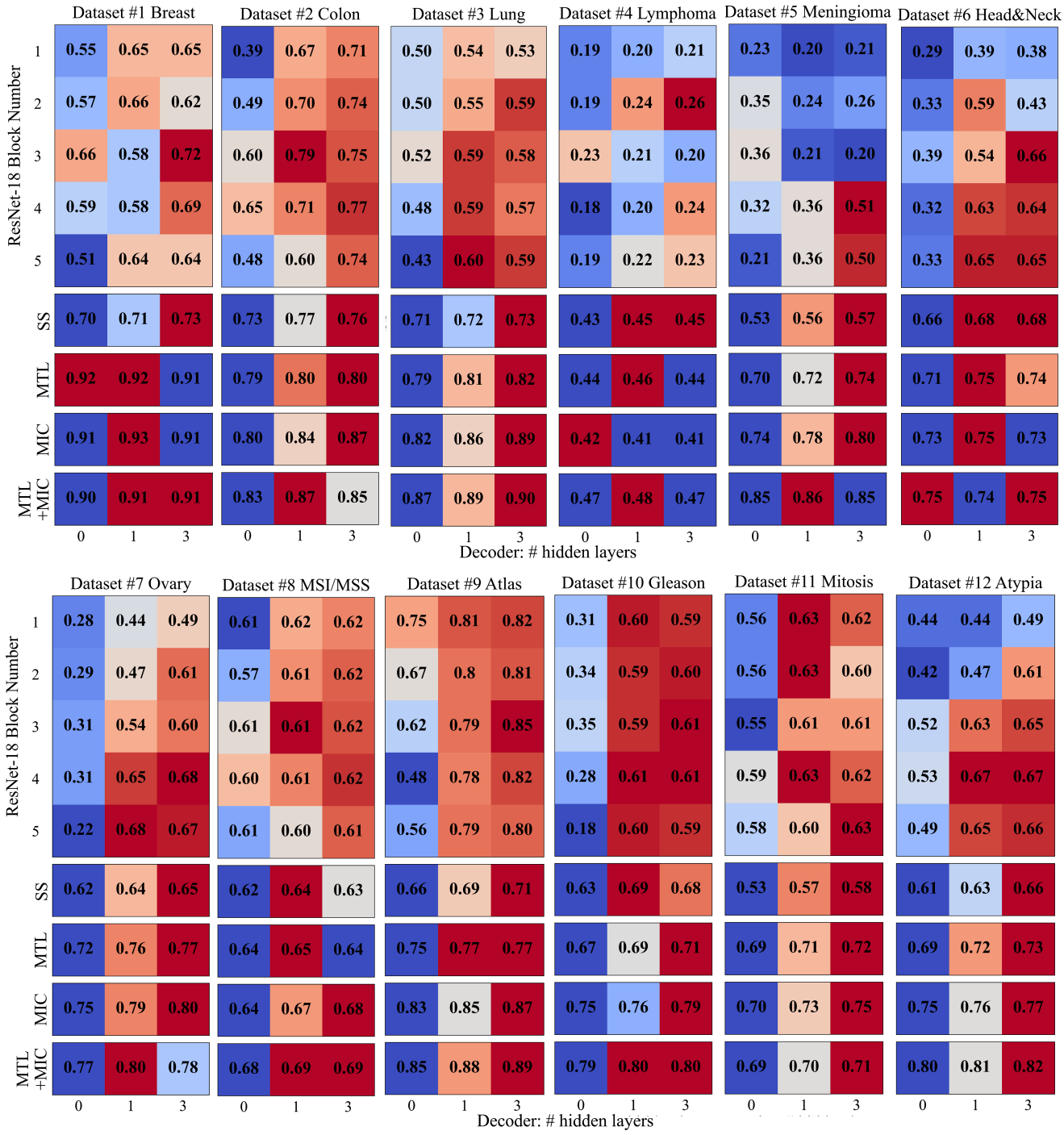


Figure 5: A study of pre-trained feature performance (y-axis) vs number of hidden layers in the decoder (x-axis). First five rows correspond to performance of features extracted from 5 residual blocks at a different depth in ResNet-18 trained on ImageNet. The remaining models correspond to: SS - self-supervised; MTL - multi-task learning model; MTL+MIC, a multi-task model trained along with ARCH dataset. All features are evaluated with a decoder with 0 (linear), 1 and 3 hidden layers with regularisation optimisation via grid-search.