Supplementary Materials: EndoNet: A Model for the Automatic Calculation of H-Score on Histological Slides

Egor Ushakov [1,*], Anton Naumov [1], Vladislav Fomberg [1], Polina Vishnyakova [2,3], Aleksandra Asaturova [2], Alina Badlaeva [2], Anna Tregubova [2], Evgeny Karpulevich [1], Gennady Sukhikh [2], Timur Fatkhudinov [3,4]

Figure S1. General pipeline of pre-training process.

	UNet (ResNet34)	UNet (ResNet50)	LinkNet (ResNet34)	FPN (ResNet34)	UNet++ (ResNet50)
Stroma AP	0.8524	0.8283	0.8175	0.8478	0.8522
Epithelium AP	0.6823	0.6507	0.7007	0.6626	0.6871
mAP	0.7674	0.7395	0.7591	0.7552	0.7696

Table S1. Results computed on test sample for all trained models.

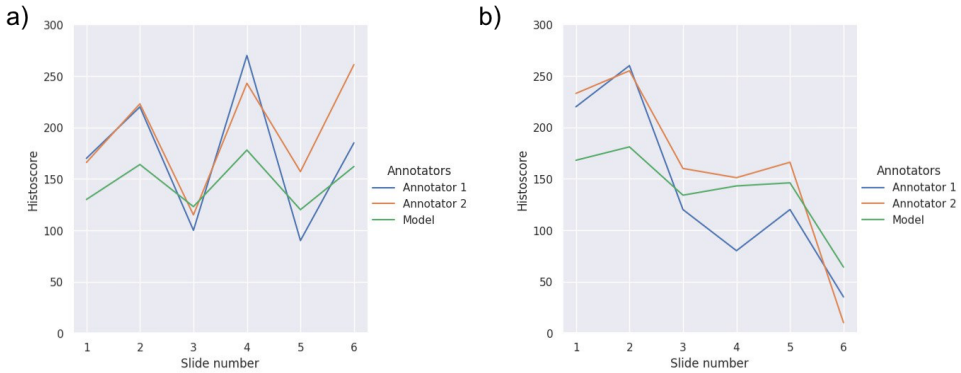

Figure S2. H-scores by pathologists and EndoNet model for 6 slides in a) stroma and b) epithelium.

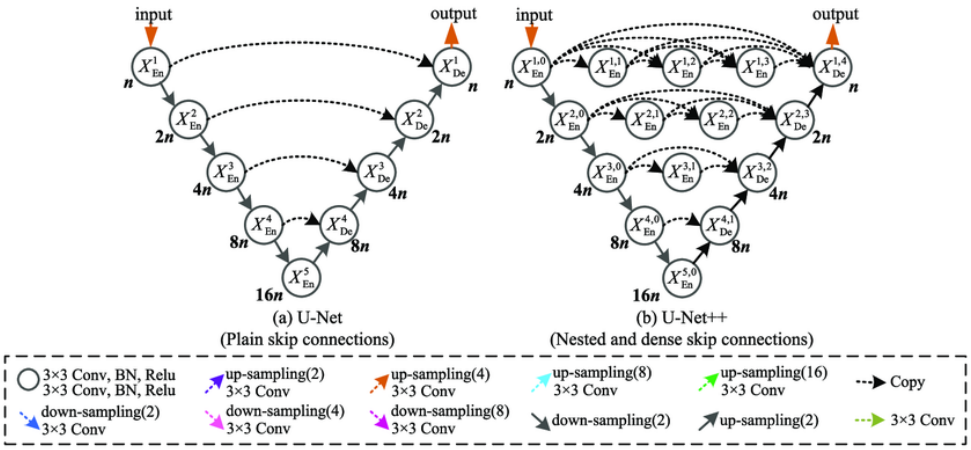

Figure S3. Schematic view of a) U-Net and b) UNet++ architecture (images from Yunjiao Deng "ELU-Net: An Efficient and Lightweight U-Net for Medical Image Segmentation" https://www.researchgate.net/figure/Comparison-of-U-Net-a-U-Net-b-U-Net-3-c-and-the-proposed-ELU-Net-d-The_fig2_359639742 by CC BY 4.0 licence)

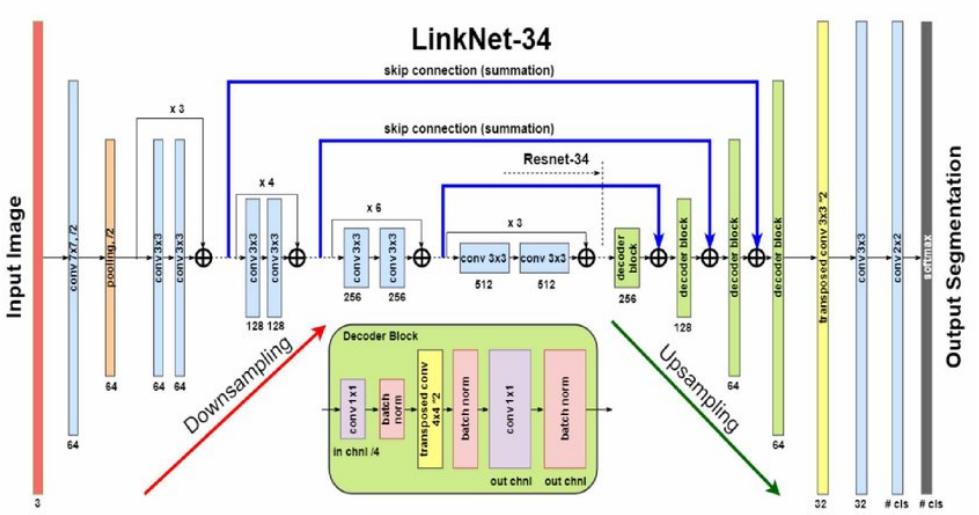

Figure S4. Schematic view of LinkNET architecture (image from Alexey A. Shvets and Alexander Rakhlin and Alexandr A. Kalinin and Vladimir I. Iglovikov "Automatic Instrument Segmentation in Robot-Assisted Surgery Using Deep Learning" https://arxiv.org/abs/1803.01207 by CC BY 4.0 license)

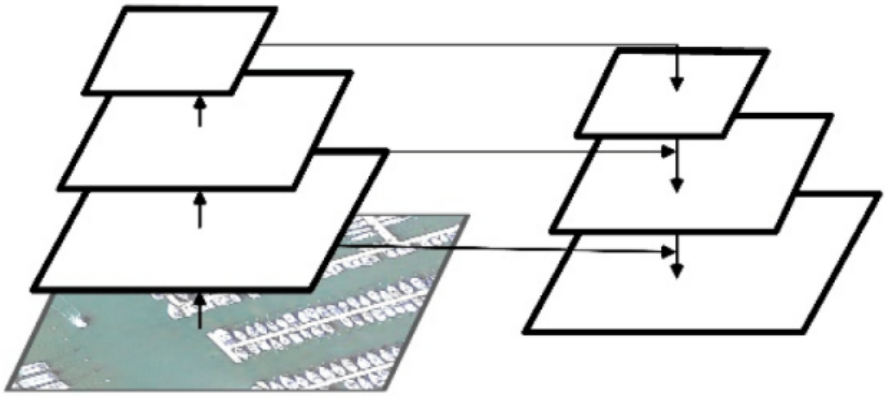

Figure S5. Schematic view of FPNet architecture (images from Wei Guo "Extended Feature Pyramid Network with Adaptive Scale Training Strategy and Anchors for Object Detection in Aerial Images" https://doi.org/10.3390/rs12050784 by CB BY 4.0 license)

UNet (Figure S3), FPNet (Figure S5), LinkNet (Figure S4), and UNet++(Figure S3) are all deep learning models used for semantic segmentation computer vision task. They share some similarities but also have differences in their architectures and performance. Below is an overview of each model and their differences:

1. UNet:
    - UNet consists of an encoder-decoder structure with skip connections between corresponding layers in the encoder and decoder.
    - The encoder downsamples the input image to capture low-level features, and the decoder gradually upsamples the features to produce a segmentation mask.
2. FPNet (Feature Pyramid Network):
    - FPNet incorporates a feature pyramid with multiple scales, allowing the network to capture features at different resolutions.
    - FPNet uses lateral connections to merge features from different scales to generate a single, unified feature map.
    - FPNet architecture is particularly useful for handling objects of varying sizes.
3. LinkNet:
    - LinkNet uses a series of encoder blocks and decoder blocks with skip connections.
    - LinkNet employs a decoder structure that is based on a "link" mechanism, which simplifies the information flow between encoder and decoder. This design choice helps reduce the computational complexity and memory requirements compared to UNet.
4. UNet++ (UNet Plus Plus):
    - UNet++ is an extension of the UNet architecture aimed at improving segmentation performance.
    - It builds upon the skip connections in UNet by introducing nested skip pathways in the encoder-decoder structure.
    - UNet++ captures features at multiple scales more effectively by nesting multiple UNet-style subnetworks.
    - The nested structure helps to model complex object boundaries and fine-grained details.

    Differences between these models:

    - UNet and UNet++ both use a UNet-style architecture with skip connections, but UNet++ introduces a nested structure to capture features at more scales.

- FPNet focuses on feature pyramid networks, making it suitable for object detection tasks, whereas UNet, LinkNet, and UNet++ are primarily designed for image segmentation.
- LinkNet emphasizes efficient training and inference by simplifying the information flow between encoder and decoder, reducing computational complexity.