Supplementary Information for Modality Attention and Sampling Enables Deep Learning with Heterogeneous Marker Combinations in Fluorescence Microscopy

Supplementary Table 1: Summary of the annotated bone marrow dataset employed for training and evaluation of the proposed methods. There are 8 different samples composed of different annotated patches. The annotated classes are either sinusoids or arteries. The remaining pixels are considered background. There is a notable class imbalance, with most of the pixels being background and arteries being a minority class.

sample id (s^i)	**# patches** (J^i)	**sinusoids (%)**	**arteries (%)**
1	54	9.71	0.29
2	12	7.87	0.99
3	30	11.97	0.21
4	12	11.20	0.98
5	12	10.93	0.58
6	48	13.60	0.24
7	54	11.23	0.24
8	8	15.19	0.27
TOTAL	230	11.41	0.34

Supplementary Table 2: Comparison of bone marrow vasculature quantification as previously reported with segmentation based on Morphological Image Processing (MIP), and with the *MS-ME* model proposed in the current work. Arteries appear as not available (n/a) because their segmentation could not be achieved with the earlier MIP approach. Manual work herein refers to the user interaction required to apply a method on a sample.

	sinusoids		arteries	
	MIP [24]	MS-ME (proposed)	MIP [24]	MS-ME (proposed)
$F1$-score	61.9±15.7	91.2±3.9	n/a	71.2±4.4
# samples	12	47	n/a	29
manual work	30 min/sample	automatic	n/a	automatic

Supplementary Table 3: Summary of the annotated dataset employed for the segmentation of fetal liver vasculature, where the only class is named *vessels* and the remaining pixels are considered background.

sample id (s^i)	**# patches** (J^i)	**vessels (%)**
0	4	20.33
1	8	20.81
2	4	18.88
3	4	13.25
4	4	20.54
5	8	21.11
6	4	25.43
7	4	15.41
8	4	27.20
9	4	10.04
10	8	12.78
11	4	25.04
12	4	17.90
13	4	21.67
14	4	31.79
15	4	5.32
16	8	6.97
17	4	23.02
18	9	18.69
19	4	18.39
20	4	20.79
21	4	27.87
22	4	22.87
23	8	12.27
24	4	1.41
25	17	16.97
TOTAL	142	17.72

Supplementary Table 4: Computational complexity of the different models employed. Time is calculated as the median across batches. GPU memory footprint refers to the maximum usage recorded during training. Note that *UB* is an ensemble of 31 models, and hence the increase in parameters only affects the training time. A single model is employed at inference for the pertinent marker combination, making the inference time equivalent to that of a single model.

Models	**Training time (ms/batch)**	**Inference time (ms/batch)**	**GPU memory footprint (GB)**	**# parameters (x10^6)**
MZ	0.28	0.20	0.95	7.76
HeMIS	1.07	0.17	3.8	0.35
MS	0.32	0.19	0.95	7.76
HeMIS-MS	1.00	0.20	3.9	0.35
MS-DR	0.34	0.17	0.95	7.76
MS-VR	0.32	0.18	0.95	7.76
MS-SE	0.35	0.15	1.12	8.2
MS-ME	0.34	0.2	1.16	7.81
MS+	0.25	0.18	1.05	9.35
UB	7.57	0.2	0.95	240.56