Supplementary Material - Magnification Prior: A Self-Supervised Method for Learning Representations on Breast Cancer Histopathological Images

Prakash Chandra Chhipa[1], Richa Upadhyay[1], Gustav Grund Pihlgren[1], Rajkumar Saini[1], Seiichi Uchida[2] and Marcus Liwicki[1]
[1]Machine Learning Group, EISLAB, Luleå Tekniska Universitet, Luleå, Sweden
[2]Human Interface Laboratory, Kyushu University, Fukuoka, Japan
[1]first.middle.last@ltu.se, [2]uchida@ait.kyushu-u.ac.jp

A. Self-supervised methods learn magnification invariant representations

The MPCS methods not only outperform for magnification-specific tasks but representations learned through the proposed method also demonstrate a consistent edge in classification performance in cross-magnification evaluation over the ImageNet model. These experiments were conducted on the Efficient-net b2 encoder only to prevent additional computation usage. However, the ResNet-50 encoder can be benchmarked, if needed. Table 1 shows (type-1 mean cross magnification evaluation) mean cross-magnification accuracy where the model is evaluated on other magnifications except the magnification on which the model was trained. The MPCS-Order Pair methods outperform the ImageNet model and other methods with a mean cross-magnification ILA of 80.99% and PLA of 81.49% when the model was trained on 40x magnification and evaluated on 100x, 200x, and 400x. Whereas MPCS-Random Pair outperforms with mean cross-magnification ILA 84.84% and PLA 82.97% when trained on 200x and ILA 84.83% and PLA 83.76% when trained on 400x. For 100x, MPCS-Ordered Pair obtains ILA 80.99%, and MPCS-Random Pair obtains PLA 85.05%. Further, Table 2 evaluates the mean performance of models trained on other magnifications except on which evaluation is performed (type-2 mean cross magnification evaluation). Interestingly, type-2 cross magnification evaluation also shows similar trends except in 400x, in which the ImageNet model obtained high PLA performance. Empirical analysis on type-1 and type-2 cross magnification suggests that MPCS self-supervised pre-trained models perform better than the ImageNet model by learning magnification invariant representations.

Table 1: Type 1 cross magnification performance comparison of proposed methods (Leave one magnification out on which model trained). The values represent mean performance of remaining magnifications (e.g. train on 40x and evaluated (mean) on 100x, 200x, and 400x.)

Method	Trained on 40X	Trained on 100X	Trained on 200X	Trained on 400X
	Mean Cross-Magnification Image Level Accuracy			
ImageNet	79.56±11.74	79.56±11.74	82.97±6.77	84.16±4.98
MPCS-Ordered Pair	**80.99±8.91**	**80.99±8.91**	84.40±3.81	84.20±5.58
MPCS-Random Pair	80.13±9.60	80.13±9.60	**84.84±5.30**	**84.83±5.30**
	Mean Cross-Magnification Patient Level Accuracy			
ImageNet	81.01±9.59	84.38±5.78	81.45±6.89	83.31±7.31
MPCS-Ordered Pair	**81.49±6.96**	84.02±5.78	82.19±3.83	83.10±7.44
MPCS-Random Pair	81.20±7.09	**85.05±7.15**	**82.97±5.84**	**83.76±6.63**

Table 2: Type 2 cross magnification performance comparison of proposed methods (select one magnification in on which model was not trained). The values represent mean performance of a magnification whereas trained on other magnifications (e.g. evaluated on 40x and trained on 100x, 200x, and 400x.

Method	Evaluated on 40X	Evaluated on 100X	Evaluated on 200X	Evaluated on 400X
	Mean Cross-Magnification Image Level Accuracy			
ImageNet	84.33 ± 4.32	83.65 ± 6.07	84.31 ± 8.93	78.86 ± 10.62
MPCS-Ordered Pair	85.35 ± 4.29	84.56 ± 6.06	**85.28 ± 8.20**	78.98 ± 7.82
MPCS-Random Pair	**86.55 ± 5.10**	**84.82 ± 5.54**	84.99 ± 7.79	**79.17 ± 9.60**
	Mean Cross-Magnification Patient Level Accuracy			
ImageNet	81.71 ± 6.25	83.76 ± 6.82	84.35 ± 8.40	**80.33 ± 8.10**
MPCS-Ordered Pair	81.97 ± 6.25	84.54 ± 6.52	84.63 ± 7.80	79.51 ± 5.20
MPCS-Random Pair	**83.24 ± 6.89**	**84.98 ± 6.12**	**84.84 ± 7.04**	79.92 ± 6.53

B. Additional ablation results on using 80%, 60%, and 40% of labels of the training set on BreakHis dataset

This section describes the extended ablation study on using labels in an incremental manner. The main section of the results describes and compares all three variants of the MPCS method with ImageNet pre-trained models when fine-tuned on 20% and 100% labels of the training set are used. To continue the trend for completeness of analysis,

Table 3: Performance evaluation of the proposed methods in limited labelled data setting when fine-tuning on 80% labels of train set.

Method	Patient Level Accuracy (RR)				Mean	Image Level Accuracy				Mean
	40X	100X	200X	400X		40X	100X	200X	400X	
ImageNet (Eff-net b2)	90.48±3.45	90.53±3.60	90.36±4.67	87.30±3.88	89.66±3.90	91.32±3.48	91.46±3.40	90.34±3.55	87.32±4.76	90.11±3.79
MPCS-FP (Eff-net b2)	91.26±3.64	91.54±3.68	90.84±2.66	88.32±3.23	90.49±3.30	91.83±3.48	92.34±3.20	91.33±2.45	88.30±3.50	90.95±3.15
MPCS-OP (Eff-net b2)	92.05±3.17	92.73±2.45	91.54±2.80	88.98±3.42	91.33±2.96	91.67±3.53	92.45±2.80	91.45±3.60	88.55±3.40	91.03±3.33
MPCS-RP (Eff-net b2)	92.16±4.18	91.51± 3.18	91.13±2.81	89.16±2.43	91.00±3.15	92.25±3.61	92.48±3.20	92.01±3.49	88.71±3.15	91.36±3.79
ImageNet (RN-50)	90.16±3.40	90.01±4.1	89.32±3.54	87.00±4.45	89.12±3.87	90.53±5.12	91.03±3.23	90.31±3.59	86.48±4.92	89.58±4.21
MPCS-FP (RN-50)	90.44±3.43	91.32±2.54	90.43±3.32	88.95±2.10	90.29±2.85	90.14±3.48	91.00±3.88	90.71±3.24	88.30±2.46	90.04±3.27
MPCS-OP (RN-50)	92.00±2.42	92.16±3.08	91.15±2.88	88.30±3.30	90.90±2.92	92.16±2.80	92.15±2.89	91.05±2.43	88.50±2.45	90.00±2.64
MPCS-RP (RN-50)	91.02±3.56	91.20± 3.28	91.10±3.88	88.20±3.17	90.38±2.92	91.02±2.44	91.22±3.13	90.02±3.65	87.06±3.84	89.80±3.27

Table 4: Performance evaluation of the proposed methods in limited labelled data setting when fine-tuning on 60% labels of train set.

Method	Patient Level Accuracy (RR)				Mean	Image Level Accuracy				Mean
	40X	100X	200X	400X		40X	100X	200X	400X	
ImageNet (Eff-net b2)	89.28±3.56	89.03±3.40	89.22±3.12	86.10±3.32	88.40±3.35	90.02±3.54	90.16±3.53	89.11±3.21	86.00±3.52	88.82±3.45
MPCS-FP (Eff-net b2)	90.22±3.87	90.45±3.20	89.72±2.70	88.32±3.21	89.67±3.25	90.56±3.32	91.28±3.43	90.20±2.55	88.20±3.40	90.06±3.30
MPCS-OP (Eff-net b2)	91.10±3.21	91.54±2.66	90.44±2.65	88.96±3.55	90.51±3.02	90.50±3.22	91.74±2.66	90.63±3.41	88.49±3.91	90.34±3.30
MPCS-RP (Eff-net b2)	91.00±4.18	90.30± 3.22	90.10±2.61	89.00±2.55	90.10±3.14	91.15±3.43	91.26±3.11	91.01±3.26	87.53±3.02	90.23±3.21
ImageNet (RN-50)	88.90±3.44	89.05±3.3	88.04±3.23	86.04±3.24	88.00±3.30	89.01±4.10	89.65±3.55	89.02±3.44	86.00±3.12	88.42±3.55
MPCS-FP (RN-50)	89.23±3.21	90.11±2.32	89.30±3.12	88.90±3.15	89.39±2.95	89.05±3.33	90.05±3.55	89.56±3.65	88.21±3.56	89.21±3.52
MPCS-OP (RN-50)	91.02±2.54	91.08±3.18	90.05±2.78	88.21±3.22	90.09±3.22	91.05±3.66	91.01±2.70	90.10±2.32	88.00±2.60	90.04±2.82
MPCS-RP (RN-50)	90.01±3.22	90.10± 3.28	90.21±3.56	88.22±3.20	89.64±3.32	90.05±2.32	90.02±3.16	89.01±3.55	86.90±3.21	89.00±3.06

Table 5: Performance evaluation of the proposed methods in limited labelled data setting when fine-tuning on 40% labels of train set.

Method	Patient Level Accuracy (RR)				Mean	Image Level Accuracy				Mean
	40X	100X	200X	400X		40X	100X	200X	400X	
ImageNet (Eff-net b2)	87.45±3.04	88.00±3.22	87.52±3.56	85.95±3.11	87.03±3.23	88.05±3.11	89.14±3.22	88.00±3.19	85.02±3.00	87.55±3.13
MPCS-FP (Eff-net b2)	88.66±3.62	88.90±3.10	89.66±2.03	88.30±3.12	88.88±2.98	88.02±3.72	89.28±3.04	90.00±2.16	88.15±3.06	88.86±3.00
MPCS-OP (Eff-net b2)	89.75±3.56	90.63±2.66	90.32±3.11	88.96±3.55	89.92±3.22	88.82±3.22	90.04±2.89	90.60±3.43	88.40±3.89	89.47±3.36
MPCS-RP (Eff-net b2)	89.05±3.27	90.05± 3.42	89.80±2.82	88.70±2.65	89.40±3.04	90.25±3.21	91.00±3.36	90.05±3.11	87.40±3.45	89.70±2.53
ImageNet (RN-50)	88.01±3.65	88.50±3.43	87.09±3.62	85.88±3.01	87.37±3.43	88.10±3.23	88.60±3.53	88.02±3.00	85.60±3.22	87.78±3.25
MPCS-FP (RN-50)	88.20±3.61	88.17±2.62	89.00±3.81	88.90±3.56	88.58±3.4	87.90±3.25	88.10±3.21	88.50±3.61	88.10±3.56	88.15±3.41
MPCS-OP (RN-50)	89.42±2.13	90.28±3.64	89.92±2.18	88.02±3.22	89.41±2.79	89.10±3.05	90.00±2.85	89.90±2.52	87.80±2.65	89.20±2.78
MPCS-RP (RN-50)	89.01±3.32	89.10± 3.82	90.00±3.46	88.15±3.26	89.06±3.47	88.10±2.35	89.30±3.20	88.90±3.05	86.85±3.22	88.29±2.96

this section adds the results for the same setting considering 40%, 60%, and 80% label utilization in fine-tuning, results described in Tables 3, 4, and 5, respectively. The most important observation is that MPCS methods consistently outperform the ImageNet model over the range of labels provided and specifically, the ordered pair method remains best performing in largely. Figure 1 and 2 shows the comparisons for Efficient-net b2 encoder for ILA and PLA accuracy. Similarly, Figure 3 and 4 shows the comparisons for ResNet-50 encoder for ILA and PLA accuracy. A common trend is evident that MPCS methods based models consistently performs better than ImageNet based model for entire range of labels.

It clearly shows that self-supervised learned representations improve fine-tuning task performance overall range of available labels, similar to the trend observed on the BACH dataset. Besides being able to obtain relatively higher accuracy on limited label settings, more label additions are largely beneficial to self-supervised pre-trained models than ImageNet pre-trained models.

C. Experimentation statistics

An extensive experimentation strategy was designed, and experiments were performed To evaluate all the variants of the proposed self-supervised pre-training method MPCS. Specifically, 15 pre-training experiments for Efficient-net b2 and 18 pre-training experiments for ResNet-50 were performed on the BreakHis dataset. Further learned representations from pre-training models are evaluated by 800 downstream task training experiments on the BreakHis dataset covering all four magnifications (40x, 100x, 200x, and 400x), 5-cross folds, and on a wide range of labels (5% to 100% train set labels). One hundred forty downstream task training experiments were performed on the BACH dataset using BreakHis MPCS pre-trained ResNet-50 models. Finally, 30 downstream tasks experiments were performed for the Breast Cell Cancer Dataset using ResNet-50 pre-trained models covering fine-tuning and linear evaluation. In this

Table 6: Experimentation statistics for proposed method MPCS

Dataset	Experiment Training Type	No. of Experiments
BreakHis	SSL pretrain - Efficient-net b2	15
	SSL pretrain - ResNet-50	18
	Downstream task- Efficient-net b2	400
	Downstream task- ResNet-50	400
BACH	Downstream task- ResNet-50	140
Breast Cell Cancer Dataset	Downstream task- ResNet-50	30
Total		**1003**

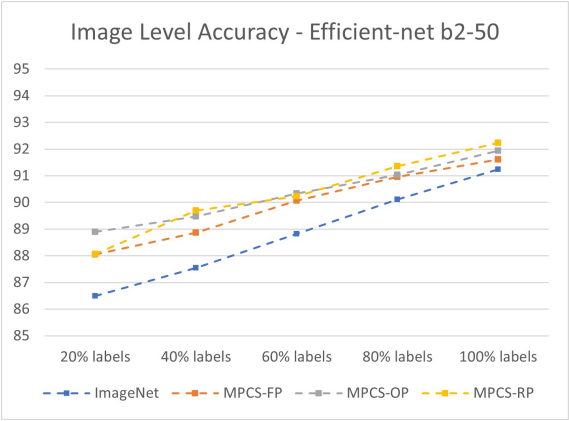

Figure 1: Performance comparison (ILA accuracy, Efficient-net b2 model) for MPCS pre-trained models with ImageNet pre-trained model over range of labels used.

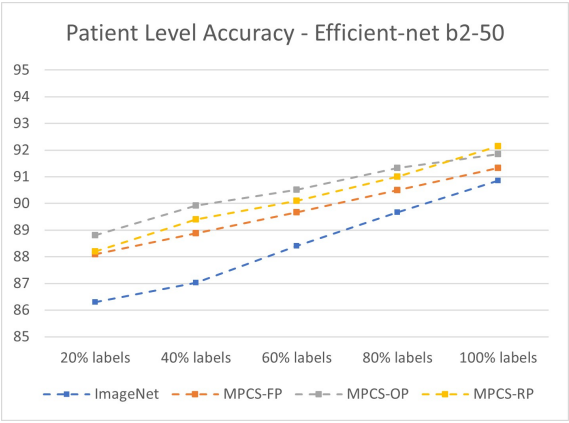

Figure 2: Performance comparison (PLA accuracy, Efficient-net b2 model) for MPCS pre-trained models with ImageNet pre-trained model over range of labels used.

way, 1003 experiments are performed in the current work. Details are mentioned in Table 6.

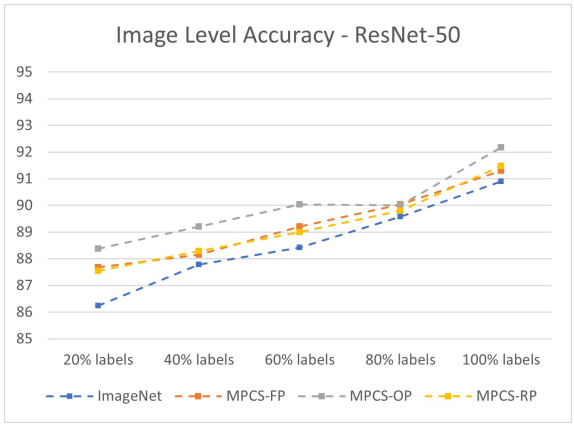

Figure 3: Performance comparison (ILA accuracy, ResNet-50 model) for MPCS pre-trained models with ImageNet pre-trained model over range of labels used.

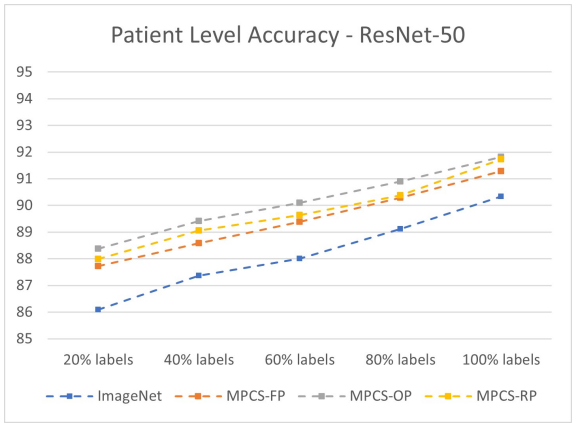

Figure 4: Performance comparison (PLA accuracy, ResNet-50 model) for MPCS pre-trained models with ImageNet pre-trained model over range of labels used.