Interpreting mechanisms of prediction for skin cancer diagnosis using multi-task learning
Supplementary Material

Davide Coppola
Bioinformatics Institute
A*STAR, Singapore
davidec@bii.a-star.edu.sg

Hwee Kuan Lee
Bioinformatics Institute
A*STAR, Singapore
leehk@bii.a-star.edu.sg

Cuntai Guan
SCSE
NTU, Singapore
CTGuan@ntu.edu.sg

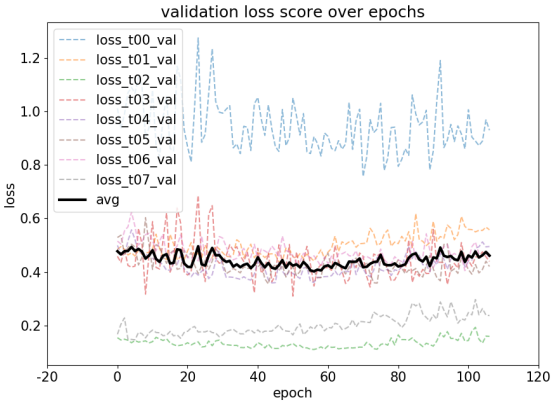
Network Architecture

block	**layer (kernel/stride)**	**# output channels**		
		`standard`	`double`	`half`
convolutional1	`conv1` (5x5/2)	2	4	1
	`maxpool1` (2x2/2)	2	4	1
gated1	-	16	32	8
convolutional2	`conv2` (1x1/1)	8	16	4
	`conv3` (3x3/1)	4	8	2
	`maxpool3` (2x2/2)	4	8	2
gated2	-	32	64	16
convolutional3	`conv4` (1x1/1)	16	32	8
	`conv5` (3x3/1)	8	16	4
	`maxpool5` (2x2/2)	8	16	4
gated3	-	64	128	32
convolutional4	`conv6` (1x1/1)	32	64	16
	`conv7` (3x3/1)	16	32	8
	`maxpool7` (2x2/2)	16	32	8
gated4	-	128	256	64
final	`conv8` (1x1/1)	64	128	32
	`avgpool8`	64	128	32
	`fcl9`	64	128	32
	`fcl10`	J^t	J^t	J^t

Table S1: Network architecture. The convolutional blocks have the same architecture for each task, but do not share weights. The last columns indicate the number of channels in the two architectures used in the experiments.

Training Curves

(a) Training loss.



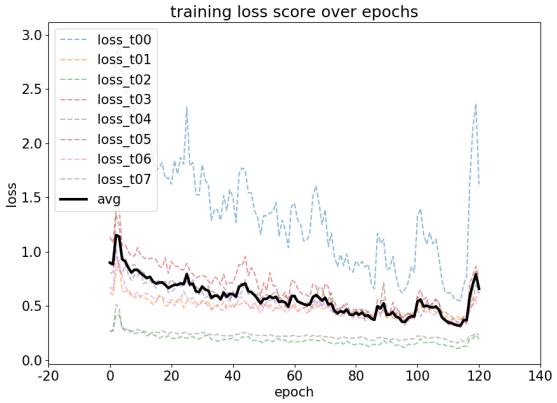
(b) Validation loss.

Figure S1: Training curves for experiment `standard`. Note that the scale is different because during testing phase the mini-batch weights are note evaluated.

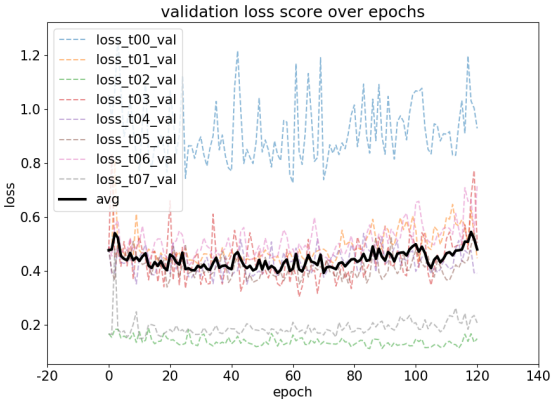
(a) Training loss.

(b) Validation loss.

Figure S2: Training curves for experiment `half`. Note that the scale is different because during testing phase the mini-batch weights are note evaluated.



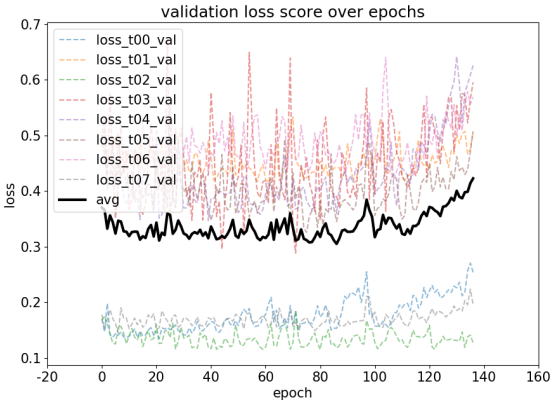
(a) Training loss.



(b) Validation loss.

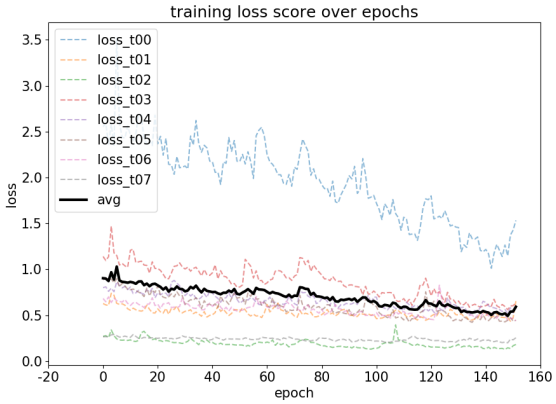
Figure S3: Training curves for experiment `double`. Note that the scale is different because during testing phase the mini-batch weights are note evaluated.

(a) Training loss.

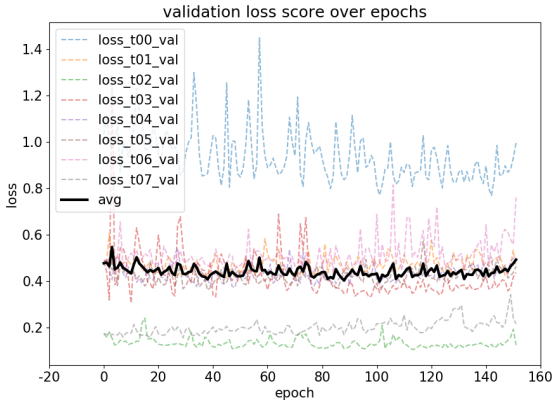


(b) Validation loss.

Figure S4: Training curves for experiment `binary`. Note that the scale is different because during testing phase the mini-batch weights are note evaluated.



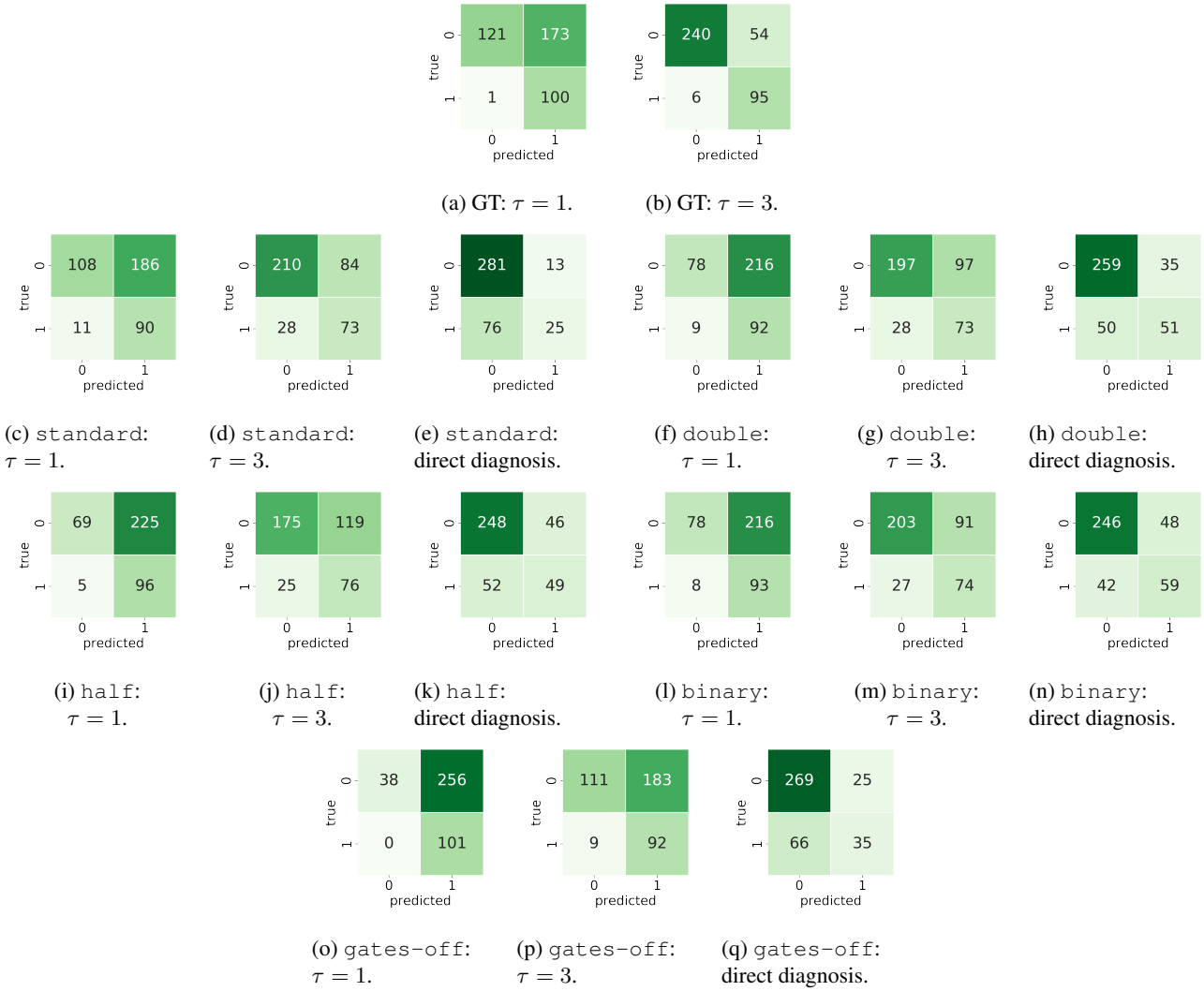
(a) Training loss.



(b) Validation loss.

Figure S5: Training curves for experiment `gates-off`. Note that the scale is different because during testing phase the mini-batch weights are note evaluated.

Melanoma diagnosis confusion matrices



(a) GT: $\tau = 1$. (b) GT: $\tau = 3$.

(c) `standard`: $\tau = 1$. (d) `standard`: $\tau = 3$. (e) `standard`: direct diagnosis. (f) `double`: $\tau = 1$. (g) `double`: $\tau = 3$. (h) `double`: direct diagnosis.

(i) `half`: $\tau = 1$. (j) `half`: $\tau = 3$. (k) `half`: direct diagnosis. (l) `binary`: $\tau = 1$. (m) `binary`: $\tau = 3$. (n) `binary`: direct diagnosis.

(o) `gates-off`: $\tau = 1$. (p) `gates-off`: $\tau = 3$. (q) `gates-off`: direct diagnosis.

Figure S6: Confusion matrices of the binary melanoma diagnosis for various experiments. τ is the threshold chosen for the 7-point checklist method [1, 2]. 0: not melanoma; 1: melanoma.

References

[1] Giuseppe Argenziano, Gabriella Fabbrocini, Paolo Carli, et al. Epiluminescence Microscopy for the Diagnosis of Doubtful Melanocytic Skin Lesions: Comparison of the ABCD Rule of Dermatoscopy and a New 7-Point Checklist Based on Pattern Analysis. *Arch Dermatol*, 134(12):1563–1570, Dec. 1998. 6

[2] Jeremy Kawahara, Sara Daneshvar, Giuseppe Argenziano, et al. Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. *IEEE Journal of Biomedical and Health Informatics*, 23(2):538–546, Mar. 2019. 6