Supplementary material for "Robust and Explainable Fine-Grained Visual Classification with Transfer Learning: A Dual-Carriageway Framework"

Zheming Zuo[1], Joseph Smith[1], Jonathan Stonehouse[2], Boguslaw Obara[1]
[1]School of Computing, Newcastle University, UK [2]Procter and Gamble, UK
{zheming.zuo,j.smith57,boguslaw.obara}@newcastle.ac.uk stonehouse.jr@pg.com

A. Optimal Padding Scheme Determination

Table 1. Detailed prediction performance (in %) comparisons on the testing sets of datasets $\mathcal{A}$ and $\mathcal{B}$, *i.e.* $\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{B}}$ regarding the labels 'F1' and 'F2', using the ResNet34 [1] (*i.e.* $\mathcal{M}$) trained on the training set of dataset $\mathcal{A}$ with `zero` padding (*i.e.* $\widehat{\mathcal{A}}$) that was processed by each of the six padding schemes. Results yielded by the optimal padding scheme are marked in bold and highlighted with .

Padding Scheme	run	Acc$\big(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{A}}_{\text{F1}}, \widetilde{\mathcal{A}}_{\text{F2}})\big)$	Acc$\big(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{A}})\big)$	mean ± std.	Acc$\big(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{B}}_{\text{F1}}, \widetilde{\mathcal{B}}_{\text{F2}})\big)$	Acc$\big(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{B}})\big)$	mean ± std.
	1	(93.16, 96.02)	94.59		(65.91, 80.73)	73.32	
	2	(92.58, 97.55)	95.06		(71.36, 76.74)	74.05	
`zero`	3	(95.25, 85.63)	90.44	94.05 ± 2.03	(82.73, 50.17)	66.45	71.79 ± 3.15
	4	(94.17, 95.72)	94.94		(77.27, 65.61)	71.44	
	5	(**95.03**, **95.41**)	**95.22**		(**74.09**, **73.26**)	**73.68**	
	1	(94.17, 96.02)	95.09		(73.18, 73.59)	73.39	
	2	(94.46, 93.88)	94.17		(75.00, 73.75)	74.38	
`RGB-mean`	**3**	(**94.10**, **96.64**)	**95.37**	95.07 ± 0.53	(**77.73**, **73.59**)	**75.66**	74.59 ± 0.86
	4	(94.31, 96.02)	95.16		(76.82, 71.93)	74.38	
	5	(94.46, 96.64)	95.55		(74.55, 75.75)	75.15	
	1	(95.10, 96.02)	95.56		(75.00, 68.44)	71.72	
	2	(94.74, 96.02)	95.38		(70.91, 73.92)	72.41	
`LAB-mean`	3	(94.67, 97.86)	96.27	95.59 ± 0.48	(75.00, 70.10)	72.55	72.61 ± 1.21
	4	(93.38, 98.17)	95.78		(75.00, 68.44)	71.72	
	5	(**94.24**, **95.72**)	**94.98**		(**72.73**, **76.58**)	**74.66**	
	1	(93.09, 96.94)	95.02		(75.00, 77.41)	76.20	
	2	(**94.74**, **95.41**)	**95.07**		(**75.00**, **79.73**)	**77.37**	
`white`	3	(94.82, 95.11)	94.97	94.89 ± 0.20	(73.18, 76.58)	74.88	75.78 ± 1.41
	4	(93.81, 95.72)	94.77		(75.00, 78.24)	76.62	
	5	(94.10, 95.11)	94.60		(70.45, 77.24)	73.84	
	1	(94.67, 96.64)	95.66		(79.55, 74.42)	76.98	
	2	(93.74, 95.41)	94.57		(75.91, 78.90)	77.41	
`grey`	**3**	(**93.81**, **96.02**)	**94.91**	95.10 ± 0.41	(**74.55**, **81.89**)	**78.22**	77.10 ± 1.06
	4	(94.74, 95.41)	95.07		(75.00, 75.75)	75.38	
	5	(94.60, 96.02)	95.31		(74.09, 80.90)	77.50	
	1	(95.10, 99.08)	97.09		(75.91, 83.72)	79.81	
	2	(95.18, 98.17)	96.68		(77.73, 80.23)	78.98	
`reflection`	**3**	(**95.03**, **98.17**)	**96.60**	**96.58 ± 0.35**	(**78.18**, **85.55**)	**81.87**	**79.85 ± 1.18**
	4	(94.82, 97.55)	96.19		(78.18, 79.90)	79.04	
	5	(95.10, 97.55)	96.32		(77.73, 81.40)	79.56	

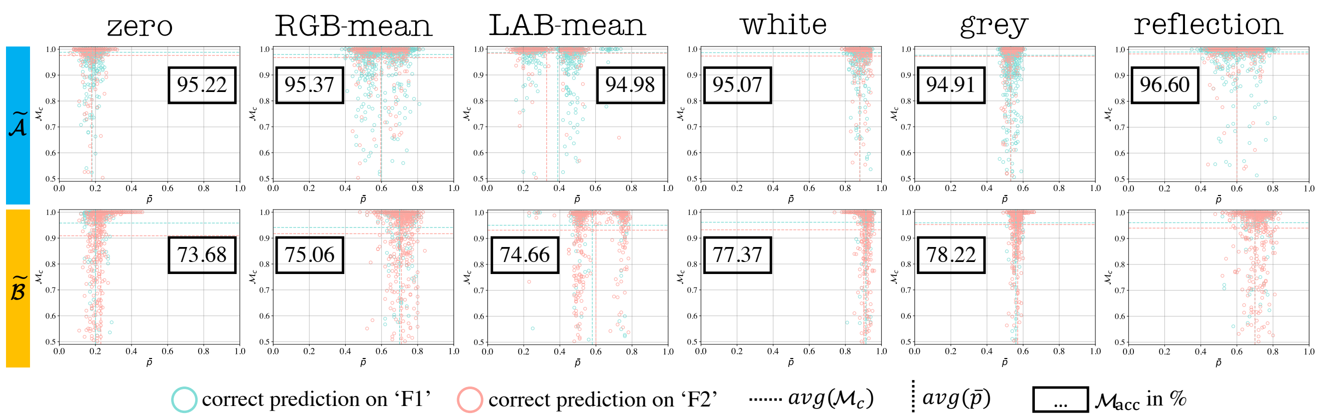

Figure 1. Quantitative explanation of the relationship between input pixel value and confidence of correct predictions given by the ResNet34 under varying padding schemes presented in Table 1. $avg(\cdot)$ calculates the average for the element regarding the total number of correctly predicted samples of the subset (grouped by label) within the testing set, $\bar{p}$ (normalised to the range of $[0, 1]$) represents mean pixel value, $\mathcal{M}_c$ denotes model confidence and $\mathcal{M}_{\text{acc}}$ corresponds to model prediction accuracy (in %). Best viewed in colour and zoomed mode.

B. Optimal Training Pathway Selection

Table 2. Detailed prediction performance (in %) comparisons on the testing sets of datasets $\mathcal{A}$ and $\mathcal{B}$, *i.e.* $\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{B}}$, using the ResNet18 [1] trained on the training set of dataset $\mathcal{A}$ (*i.e.* $\widehat{\mathcal{A}}$) with `reflection` padding (as summarised in Table 1) under five different training pathways in the proposed DCF. Results yielded by the optimal padding scheme are marked in bold and highlighted with ▒.

Model	setting	run	$\text{Acc}(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{A}}_{\text{F1}}, \widetilde{\mathcal{A}}_{\text{F2}}))$	$\text{Acc}(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{A}}))$	mean ± std.	$\text{Acc}(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{B}}_{\text{F1}}, \widetilde{\mathcal{B}}_{\text{F2}}))$	$\text{Acc}(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{B}}))$	mean ± std.
ResNet18	1	**1**	**(94.60, 98.78)**	**96.69**	96.80 ± 0.07	**(78.64, 86.21)**	**82.42**	81.81 ± 0.53
		2	(94.53, 99.08)	96.81		(81.36, 83.06)	82.21	
		3	(94.67, 99.08)	96.88		(76.36, 86.88)	81.62	
		4	(94.53, 99.08)	96.81		(79.55, 82.56)	81.06	
		5	(93.95, 99.69)	96.82		(77.27, 86.21)	81.74	
	2	1	(95.61, 96.33)	95.97	**96.38 ± 0.32**	(84.09, 99.34)	91.72	**92.32 ± 0.85**
		2	(95.03, 97.25)	96.14		(87.27, 98.17)	92.72	
		3	(94.74, 98.47)	96.60		(84.55, 98.84)	91.69	
		4	(95.25, 98.17)	96.71		(85.00, 98.67)	91.84	
		5	**(94.82, 98.17)**	**96.50**		**(90.91, 96.35)**	**93.63**	
	3	1	(91.79, 96.64)	94.22	94.25 ± 0.83	(82.73, 99.00)	90.87	91.49 ± 0.44
		2	(91.79, 96.02)	93.91		(84.09, 99.17)	91.63	
		3	(92.01, 96.33)	94.17		(84.55, 98.84)	91.69	
		4	**(91.79, 99.39)**	**95.59**		**(83.64, 98.84)**	**91.24**	
		5	(91.29, 95.41)	93.35		(85.00, 99.00)	92.00	
	4	1	(90.78, 86.54)	88.66	88.02 ± 1.82	(80.91, 98.50)	89.70	90.15 ± 0.82
		2	**(91.14, 88.69)**	**89.91**		**(84.55, 98.34)**	**91.44**	
		3	(92.51, 85.02)	88.77		(80.91, 98.34)	89.62	
		4	(90.06, 80.12)	85.09		(82.27, 98.67)	90.47	
		5	(90.06, 85.32)	87.69		(81.82, 97.18)	89.50	
	5	1	(95.10, 98.47)	96.78	96.11 ± 0.66	(79.09, 91.03)	85.06	86.04 ± 0.79
		2	(94.17, 97.55)	95.86		(81.36, 91.36)	86.36	
		3	(95.10, 98.17)	96.63		(77.27, 93.69)	85.48	
		4	(92.73, 97.55)	95.14		(78.64, 95.51)	87.08	
		5	**(94.74, 97.55)**	**96.14**		**(79.55, 92.86)**	**86.20**	

Table 3. Detailed prediction performance (in %) comparisons on the testing sets of datasets $\mathcal{A}$ and $\mathcal{B}$ using the ResNet34 [1] and Inception-v3 [2] trained on the training set of dataset $\mathcal{A}$ with `reflection` padding under five different training pathways. Results yielded by the optimal padding scheme are marked in bold and highlighted with ▭.

Model	setting	run	$\mathrm{Acc}\big(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{A}}_{\mathrm{F1}}, \widetilde{\mathcal{A}}_{\mathrm{F2}})\big)$	$\mathrm{Acc}\big(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{A}})\big)$	mean ± std.	$\mathrm{Acc}\big(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{B}}_{\mathrm{F1}}, \widetilde{\mathcal{B}}_{\mathrm{F2}})\big)$	$\mathrm{Acc}\big(\mathcal{M}^{\widehat{\mathcal{A}}}(\widetilde{\mathcal{B}})\big)$	mean ± std.
ResNet34	1	1	(95.10, 99.08)	97.09		(75.91, 83.72)	79.81	
		2	(95.18, 98.17)	96.68		(77.73, 80.23)	78.98	
		3	(**95.03**, **98.17**)	**96.60**	96.58 ± 0.35	(**78.18**, **85.55**)	**81.87**	79.85 ± 1.18
		4	(94.82, 97.55)	96.19		(78.18, 79.90)	79.04	
		5	(95.10, 97.55)	96.32		(77.73, 81.40)	79.56	
	2	1	(95.25, 98.47)	96.86		(85.91, 99.17)	92.54	
		2	(95.46, 97.25)	96.35		(90.00, 95.85)	92.92	
		3	(95.46, 97.86)	96.66	**96.51 ± 0.40**	(87.27, 97.84)	92.56	**92.80 ± 0.23**
		4	(95.46, 96.33)	95.89		(88.64, 97.18)	92.91	
		5	(**95.46**, **98.17**)	**96.81**		(**87.27**, **98.84**)	**93.06**	
	3	1	(94.02, 96.64)	95.33		(85.00, 98.67)	91.84	
		2	(**94.53**, **96.64**)	**95.59**		(**84.09**, **99.17**)	**91.63**	
		3	(93.38, 93.88)	93.63	94.65 ± 1.07	(83.64, 98.17)	90.91	91.67 ± 0.62
		4	(92.80, 97.86)	95.33		(84.09, 98.67)	91.38	
		5	(94.96, 91.74)	93.35		(86.36, 98.84)	92.60	
	4	1	(87.90, 94.19)	91.05		(84.09, 98.17)	91.13	
		2	(89.27, 93.27)	91.27		(82.27, 98.67)	90.47	
		3	(**92.30**, **92.97**)	**92.63**	90.94 ± 1.15	(**83.64**, **98.67**)	**91.16**	91.16 ± 0.43
		4	(89.63, 90.52)	90.07		(85.45, 97.67)	91.56	
		5	(90.42, 88.99)	89.70		(85.45, 97.51)	91.48	
	5	1	(94.67, 99.08)	96.88		(80.00, 96.18)	88.09	
		2	(94.96, 98.78)	96.87		(80.00, 97.18)	88.59	
		3	(95.03, 98.47)	96.75	96.81 ± 0.06	(81.36, 97.18)	89.27	88.73 ± 0.58
		4	(95.10, 98.47)	96.78		(79.09, 97.51)	88.30	
		5	(**94.74**, **98.78**)	**96.76**		(**80.45**, **98.34**)	**89.40**	
Inception-v3	1	1	(94.82, 97.86)	96.34		(77.27, 87.38)	82.32	
		2	(**94.89**, **99.69**)	**97.29**		(**80.00**, **89.37**)	**84.69**	
		3	(94.38, 97.55)	95.97	96.14 ± 0.93	(77.73, 89.87)	83.80	83.33 ± 1.27
		4	(93.45, 96.02)	94.73		(79.09, 89.20)	84.15	
		5	(94.89, 97.86)	96.38		(76.36, 87.04)	81.70	
	2	1	(95.82, 97.55)	96.69		(86.82, 98.67)	92.75	
		2	(95.03, 97.86)	96.44		(88.64, 97.67)	93.16	
		3	(**95.54**, **98.17**)	**96.86**	**96.77 ± 0.40**	(**90.91**, **95.85**)	**93.38**	**93.11 ± 0.48**
		4	(95.32, 97.55)	96.44		(88.64, 98.84)	93.74	
		5	(95.75, 99.08)	97.41		(87.73, 97.34)	92.53	
	3	1	(95.39, 94.80)	95.09		(84.55, 99.00)	91.78	
		2	(**95.25**, **96.94**)	**96.09**		(**86.36**, **98.50**)	**92.43**	
		3	(94.10, 98.78)	96.44	95.96 ± 0.68	(85.45, 98.50)	91.97	91.88 ± 0.38
		4	(94.89, 96.02)	95.45		(84.55, 99.17)	91.86	
		5	(94.38, 99.08)	96.73		(84.55, 98.17)	91.36	
	4	1	(93.81, 83.49)	88.65		(84.09, 98.17)	91.13	
		2	(85.17, 89.60)	87.38		(85.00, 97.18)	91.09	
		3	(92.01, 85.02)	88.52	88.69 ± 0.90	(83.64, 98.67)	91.16	90.90 ± 0.69
		4	(90.64, 87.46)	89.05		(81.36, 98.01)	89.69	
		5	(**87.69**, **92.05**)	**89.87**		(**84.55**, **98.34**)	**91.44**	
	5	**1**	(**94.96**, **99.08**)	**97.02**		(**84.09**, **96.51**)	**90.30**	
		2	(94.67, 99.69)	97.18		(78.64, 96.51)	87.58	
		3	(94.74, 98.47)	96.60	96.97 ± 0.27	(82.27, 94.02)	88.14	89.02 ± 1.13
		4	(94.82, 99.69)	97.25		(81.82, 96.84)	89.33	
		5	(95.39, 98.17)	96.78		(83.64, 95.85)	89.75	

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. CVPR*, pages 770–778, 2016. 1, 2, 3

[2] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proc. CVPR*, pages 2818–2826, 2016. 3