Quasi-Unsupervised Color Constancy

Supplementary material

Simone Bianco
University of Milano-Bicocca
simone.bianco@unimib.it

Claudio Cusano
University of Pavia
claudio.cusano@unipv.it

We report here some additional data and examples that, for the sake of brevity, we did not include in the paper. Most of the material is related to learning scenarios (such as in-db, and supervised learning) that are not the main focus of the paper, but that can be useful anyway to have a comprehensive view of the experimental results.

Comparison with the state of the art

Figure S1 summarizes the comparison of the proposed method with other methods in the state of the art. It extends of Figure 5 of the paper, by including all the six categories of methods considered in the paper and reported in its Table 3. We remind that the three "in-db" groups had access to training data coming from the evaluation dataset in two different forms: data only for the unsupervised (in-db) group, data and illuminant ground truths for the parametric (in-db) and supervised (in-db) groups.

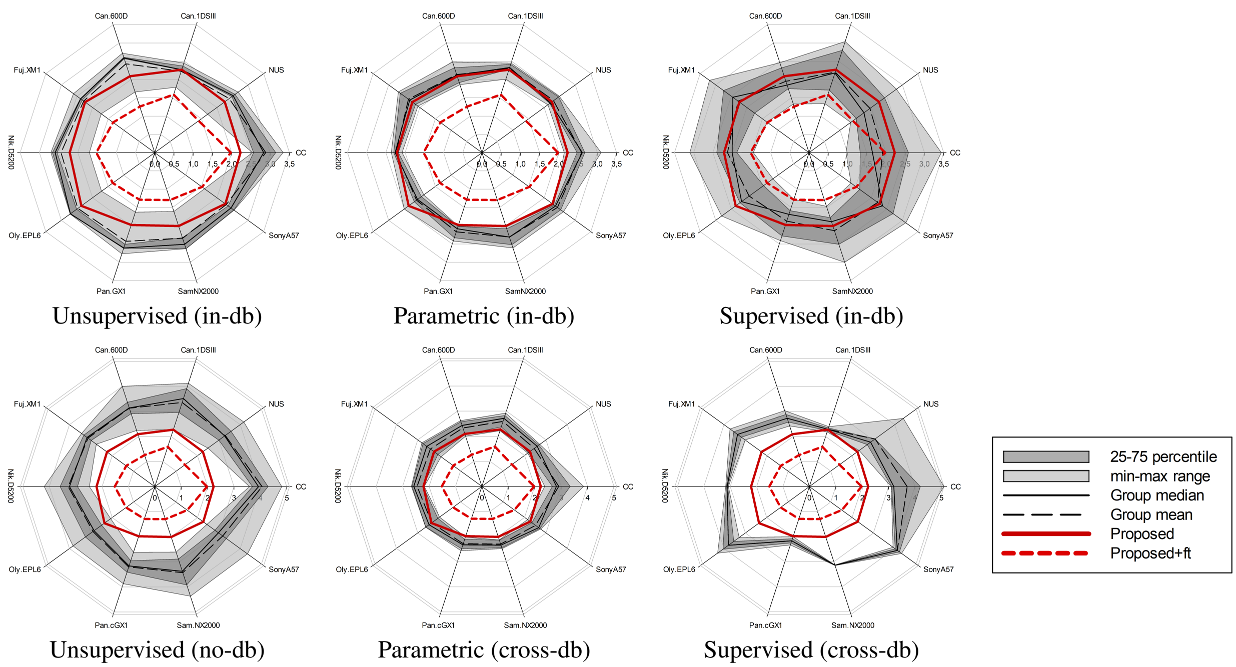

Figure S1. Visual summary of the median errors of the proposed method, with and without fine tuning, on CC, NUS and individual NUS cameras. The method is compared with six groups of algorithms. For each group are drawn the best and worst median error, the interquartile range, the median ad the mean.

Supervised fine tuning

Table S1 extends Table 2 in the paper by reporting the results obtained by the neural networks trained on the three training data sets and fine tuned on Color Checker and NUS. Results have been computed with a three-fold cross validation. For each combination of training and test data sets we considered two variants: the first uses the equalized grayscale image as input; the second uses the combination of the grayscale image with gradient directions.

Results show that in all cases the fine tuning procedure reduced the mean angular error. Moreover, only in two cases out of twelve (training on Flickr100k, fine tuning on Color Checker) the median error increased and in both cases very slightly (by 0.01 and 0.03). The benefit of fine tuning on NUS is generally larger than that of fine tuning on CC. This may depend on the limited size of CC (568 images) compared to NUS (1853 images).

Training set	Test set	Input	Mean	Median	Max
Ilsvrc12	CC	Grayscale	3.82 (-0.22)	2.50 (-0.17)	19.56 (-8.3)
	CC	Grayscale + Directions	**2.91** (-0.55)	**1.98** (-0.25)	19.96 (-1.2)
Places365	CC	Grayscale	3.76 (-0.25)	2.36 (-0.24)	25.73 (-2.0)
	CC	Grayscale + Directions	3.10 (-0.50)	2.33 (-0.12)	**14.84** (-6.6)
Flickr100k	CC	Grayscale	3.94 (-0.15)	2.68 (+0.01)	24.96 (-2.1)
	CC	Grayscale + Directions	3.08 (-0.51)	2.28 (+0.03)	20.43 (+0.4)
Ilsvrc12	NUS	Grayscale	2.55 (-0.59)	1.77 (-0.47)	19.79 (-2.6)
	NUS	Grayscale + Directions	1.97 (-1.03)	**1.41** (-0.86)	20.52 (+1.6)
Places365	NUS	Grayscale	2.54 (-0.70)	1.81 (-0.51)	16.66 (-6.0)
	NUS	Grayscale + Directions	**1.95** (-1.12)	1.46 (-0.74)	**12.95** (-4.2)
Flickr100k	NUS	Grayscale	2.45 (-0.82)	1.68 (-0.70)	20.37 (-0.9)
	NUS	Grayscale + Directions	2.03 (-0.95)	1.52 (-0.64)	14.50 (-1.4)

Table S1. Statistics of angular errors (in degrees) obtained by the network trained on Ilsvrc12, places365 and Flickr100k and fine tuned on CC and NUS. The values in brackets report the differences with respect to those obtained in the quasi-unsupervised setting. The best statistics for each test data set are reported in bold.

Visual comparison of the variants of the method

Figures S2 and S2 show a few images in the Color Checker and NUS data sets processed by twelve variants of the method. The variants differ in the data set used for training (Ilsvrc12, Places365 or Flickr100k), in the type of input (equalized grayscale with and without gradient directions) and in the learning setting (quasi-unsupervised, or supervised fine tuning). The figures suggest which are the main differences in the behavior of the variants:

- the networks processing only grayscale images select pixels in regions that are easier to recognize with respect to those obtained using also gradient directions.
- Fine tuned networks tend to select less pixels than the others. This may depend on the fact that, during the fine tuning, we removed the noise term that pushed the network to output large weights.
- The output of the networks only marginally depend on the training set used. In particular the weights found by networks trained on Ilsvrc12 and Places365 are, most of the times, very similar.
- With a small number of exceptions, the balanced images look very similar to those corrected using the ground truth illuminant.

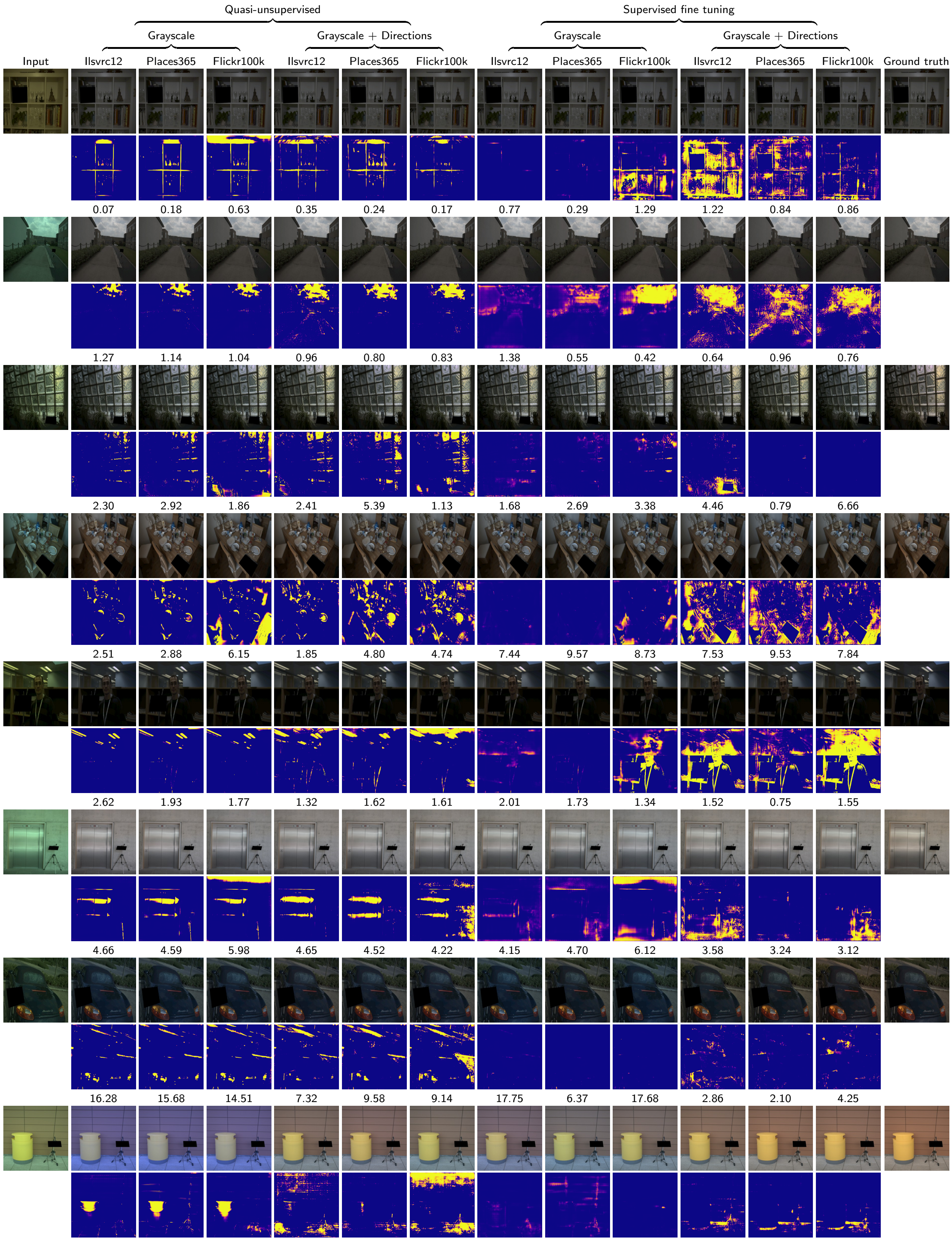

Figure S2. Color Checker images processed by variants of the method. For each variant are shown the image balanced according to the estimated illuminant, the weights produced by the neural network, and the angular error.

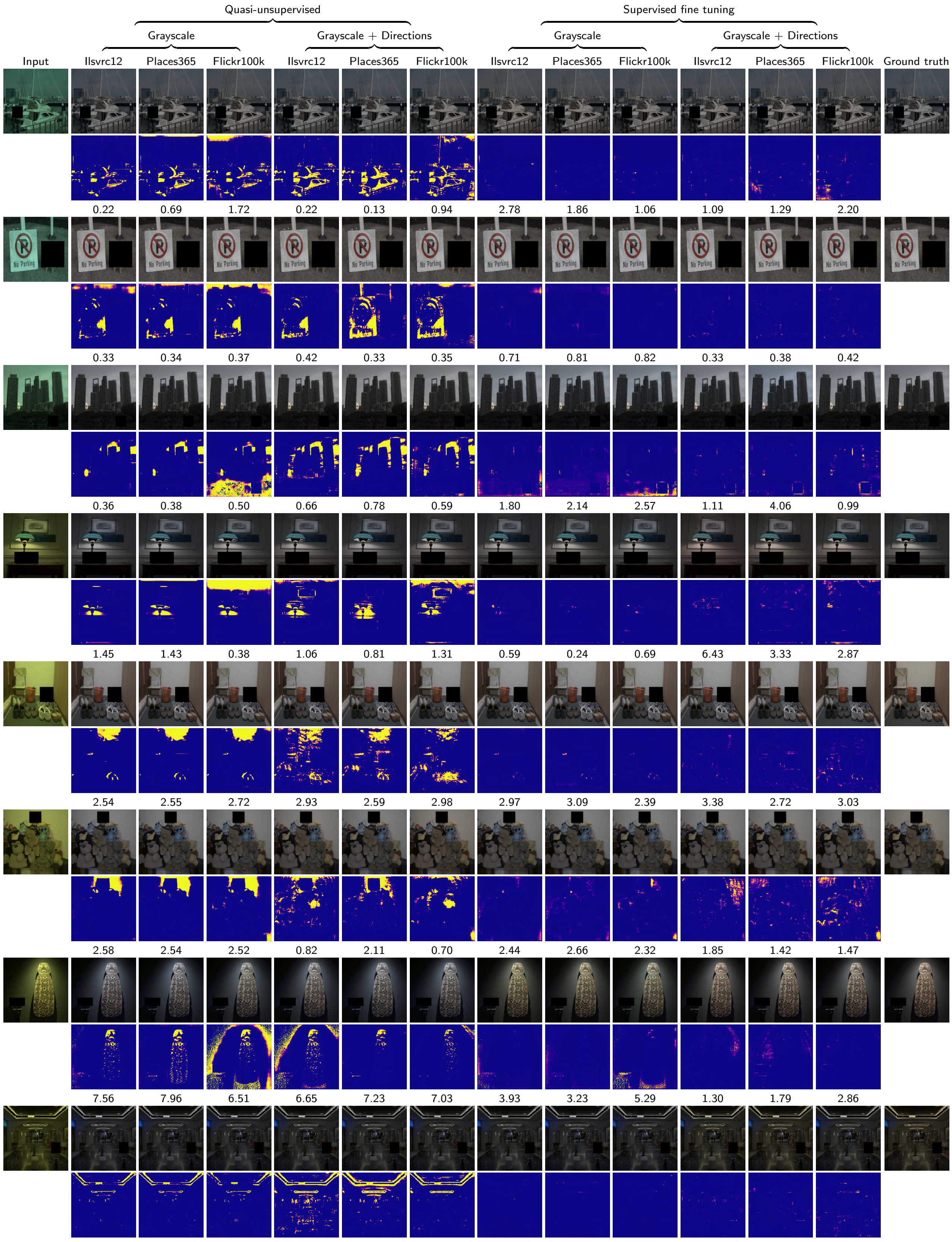

Figure S3. NUS images processed by variants of the method. For each variant are shown the image balanced according to the estimated illuminant, the weights produced by the neural network, and the angular error.