Supplementary material for submission When Naïve Bayes Nearest Neighbors Meet Convolutional Neural Networks

Ilja Kuzborskij[1,2,3], Fabio Maria Carlucci[1], Barbara Caputo[1,2]
[1]Sapienza Rome University, Dept. of Computer, Control and Management Engineering, Italy
[2]Idiap Research Institute, Switzerland
[3]École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
{kuzborskij, fmcarlucci, caputo}@dis.uniroma1.it

1. Supplementary Experiments

Here, additional experimental results are provided both for Scene Recognition and Domain Adaptation.

1.1. Scene Recognition Experiments

Figure 1 contains, from top-left proceeding clockwise, results for NBNN on:

- Scene 15 [2] dataset, sparse sampling and Hybrid [5] features
- Scene 15 [2] dataset, sparse sampling and Imagenet [4] features
- UIUC Sports [3] dataset, sparse sampling and Places [5] features
- UIUC Sports [3] dataset, sparse sampling and Imagenet [4] features

1.2. Domain Adaptation Experiments

Tables 1 to 4 contain our full NBNN results on the Office + Caltech setting [1], both unsupervised and semi-supervised.

Table 1: Source only NBNN - non ReLU

		Amazon	Webcam	DSLR	Caltech
16px - 1 level	Amazon	88.23 ± 1.17	42.7 ± 7.2	38.4 ± 5.8	63.12 ± 1.16
	Webcam	48.67 ± 2.13	98.41 ± 1.21	88.59 ± 2.9	44.29 ± 1
	DSLR	26.25 ± 0.48	75.59 ± 0.84	93.75 ± 3.21	34.75 ± 0.68
	Caltech	78.5 ± 2.29	44.64 ± 4.8	41.65 ± 11.19	72.38 ± 1.37
16px - 2 levels	Amazon	87.67 ± 1.17	26.8 ± 7.85	35.09 ± 5.07	62.68 ± 1.49
	Webcam	48.14 ± 1.8	97.44 ± 1.78	88.9 ± 3.1	44.22 ± 1.22
	DSLR	34.9 ± 1.81	71.96 ± 0.99	92.18 ± 2.32	42.9 ± 0.749
	Caltech	74.05 ± 1.9	38.06 ± 5.6	41.84 ± 7.5	72.09 ± 1.54
16px - 3 levels	Amazon	89.43 ± 1.12	50.5 ± 1.81	58.91 ± 4.24	70.12 ± 1.61
	Webcam	52.48 ± 2.69	97.79 ± 1.89	96.17 ± 1.82	52.55 ± 1.55
	DSLR	60.76 ± 1.5	90.64 ± 0.62	94.37 ± 1.5	56.1 ± 0.91
	Caltech	75.9 ± 2.08	46.33 ± 4.64	57.89 ± 6.27	78.29 ± 1.45
32 px - 1 level	Amazon	87.44 ± 1.01	34.033 ± 5.68	35.15 ± 5.27	63.74 ± 1.26
	Webcam	50.63 ± 1.15	98.27 ± 1.78	89.93 ± 1.52	45.19 ± 1.38
	DSLR	29.85 ± 0.89	73.49 ± 1.27	92.34 ± 2.7	39.11 ± 0.7
	Caltech	75.78 ± 2.52	38.71 ± 6.77	41.97 ± 7.99	72.29 ± 1.87
32px - 2 levels	Amazon	88.98 ± 0.89	48.2 ± 3.21	55.66 ± 5.37	70.14 ± 1.67
	Webcam	54.06 ± 1.92	98.2 ± 1.08	95.79 ± 1.09	51.7 ± 1.8
	DSLR	43.69 ± 1.19	87.01 ± 0.76	95 ± 2.05	50.5 ± 0.78
	Caltech	74.79 ± 1.84	48.5 ± 6.38	59.04 ± 9.26	76.92 ± 1.9
32px - 3 levels	Amazon	89.85 ± 0.64	59.89 ± 2.55	68.28 ± 4.95	75.76 ± 1.33
	Webcam	66.53 ± 1.33	98.48 ± 1.29	98.21 ± 0.98	62.53 ± 1.25
	DSLR	67.27 ± 1.14	94 ± 0.57	97.18 ± 1.77	65.35 ± 0.66
	Caltech	78.81 ± 0.84	61.25 ± 4.4	66.75 ± 6.26	80.5 ± 1.42
64px - 1 level	Amazon	89.03 ± 0.92	49.83 ± 3.84	56.36 ± 5.3	69.1 ± 1.59
	Webcam	57.33 ± 1.98	98.55 ± 1.54	96.3 ± 1.26	51.42 ± 1.57
	DSLR	42.56 ± 0.87	84.88 ± 0.89	92.81 ± 2.91	47.7 ± 0.64
	Caltech	74.37 ± 1.69	49.32 ± 4.2	60.06 ± 7.2	76.37 ± 1.7
64px - 2 levels	Amazon	89.84 ± 0.81	60.54 ± 3.59	68.66 ± 4.95	75.33 ± 1.05
	Webcam	65.55 ± 1.41	98.55 ± 1.54	98.15 ± 1.14	62.57 ± 1.03
	DSLR	64.46 ± 0.92	94.74 ± 0.55	95.93 ± 1.67	63.57 ± 0.404
	Caltech	78.04 ± 1.04	60.16 ± 5.29	66.49 ± 6.17	80.23 ± 1.21
64px - 3 levels	Amazon	89.76 ± 0.95	60.23 ± 3.29	68.98 ± 4.54	75.2 ± 1.08
	Webcam	66.87 ± 1.2	98.48 ± 1.48	98.53 ± 0.73	63.33 ± 1.2
	DSLR	67.4 ± 0.94	93.93 ± 0.74	97.18 ± 2.05	65.22 ± 0.64
	Caltech	79.03 ± 0.93	61.28 ± 5.53	67.83 ± 2.0	80.46 ± 1.35

References

[1] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2012. 1

[2] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, IEEE Conference on*, 2006. 1

[3] L.-J. Li and L. Fei-Fei. What, where and who? classifying events by scene and object recognition. In *Computer Vision (ICCV), IEEE International Conference on*, 2007. 1

[4] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, pages 1–42, 2015. 1

[5] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems, NIPS*, 2014. 1

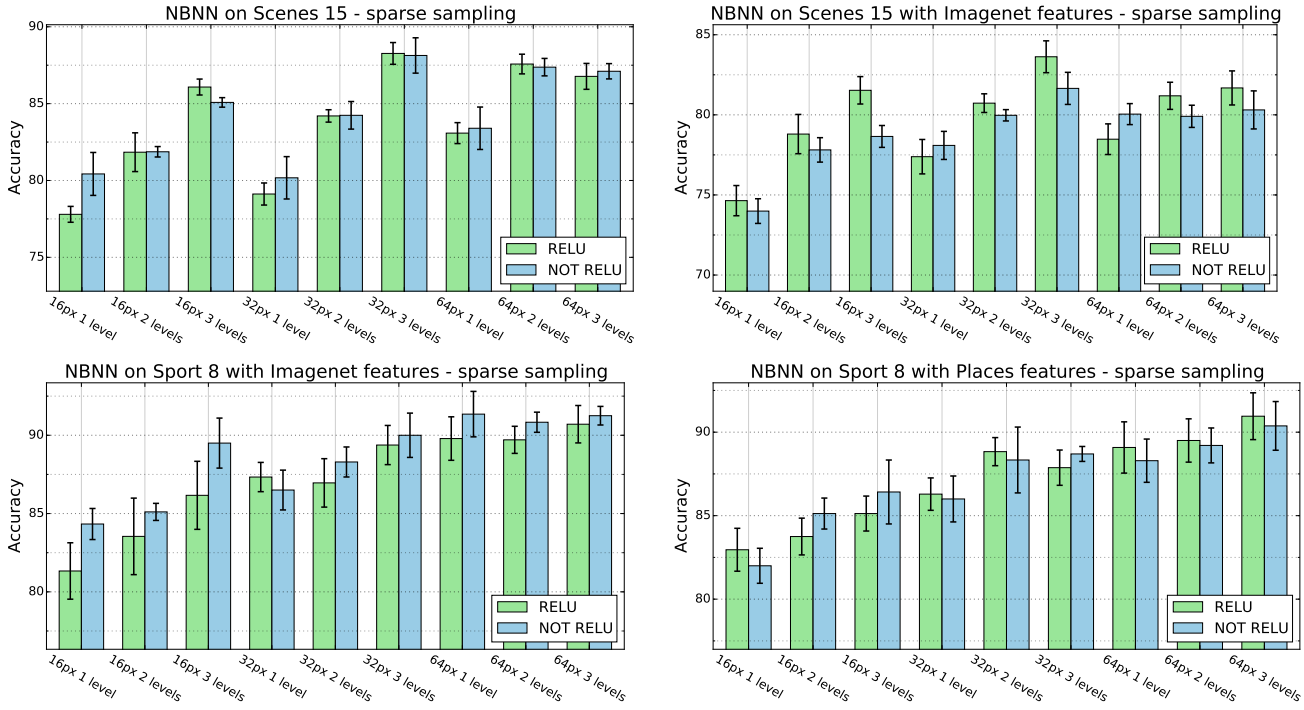

Figure 1: Results obtained by NBNN using CNN activations, based on different networks.