Supplemental Material for the paper
"Learning Separable Filters"

IEEE Conference on Computer Vision and Pattern Recognition, 2013

Roberto Rigamonti Amos Sironi Vincent Lepetit Pascal Fua
CVLab, École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
name.surname@epfl.ch

Additional Results

In this appendix we present some additional details on the results we obtained that, owing to space limitations, did not fit in our submission. We will first detail the results we obtained on the pixel classification task, and then we will analyze the approximation capabilities of our approach in the considered denoising framework.

1. Pixel Classification

Figure 1 illustrates some filter banks learned on the DRIVE dataset. In particular, it shows how an example learned filter bank can be replaced by its rank-1 approximation obtained using the SVD decomposition (*SEP-SVD*).

Figure 2 shows examples of 3D filter banks learned on the OPF dataset. In Figure 3 the central slice of the filter banks are given for a better comparison.

We report the detailed pixel classification results in Table 1. To provide a fair comparison of the different approaches, we have evaluated the quality of the classification we obtained on both the 2D and the 3D datasets by using several analytic measures. In particular, we considered the following:

- Area Under Curve (AUC): represents the area subtended by the ROC curve. It assumes values in $[0, 1]$, the higher the better;
- F-measure [6];
- Variation of Information (VI) [2]: assumes values in $[\ 0, \infty)$, the lower the better;
- Rand Index (RI) [5]: assumes values in $[0, 1]$, the higher the better.

We also report the time it takes to compute the convolutions with the filters or, in the OOF case, the tubularity score. Both the VI and the RI values are computed on the classification results thresholded at the value found using the F-measure. The corresponding PR and ROC curves for the 2D case are depicted by Figure 4 and Figure 5 respectively. In the 3D case, since the number of negative samples is much larger than the number of positive samples (about 99% of total volume), the ROC curve is always close to the ideal curve and therefore not a representative measure of accuracy (as shown by the large AUC values in Table 3). Thus, in the 3D case we preferred to show a plot of the F-measure and the computational time instead of ROC curves (Figure 13).

From the table and the graphs it can be inferred that

- Filtering methods are more effective, on the considered datasets, than state-of-the-art hand-crafted approaches. This has been shown to be true also in [3].
- Expressing the full-rank learned filter bank in terms of few separable filters leads to a significant speed-up at no cost in terms of accuracy.

Table 2 shows the time required to filter a $512 \times 512 \times 64$ 3D volume with different strategies. Separability emerges as the key factor for improving speed, leading to a $8\times$ reduction, while a subsequent reduction in the number of filters involved, as done in *SEP-COMB**, leads to an additional $2\times$ speed-up. The reduction in the number of filters, by expressing a filter bank on a shared, smaller spanning basis, is clearly not sufficient if the separability property is not considered. Also, this does not remove the dependency on the dimensionality of the input.

Figures 6,7,8,9,10,11,12 show the pixel classification results for some images from the considered 2D datasets. The results of our method are indistinguishable from the classifications obtained by the non-separable approach, and vastly outperform the results achieved by the hand-crafted method.

2. Image Denoising

We have chosen image denoising as a task to evaluate the quality of the approximation provided by our separable filters. Indeed, a bad approximation of the filter bank would be reflected by poor denoising performances. The denoising results for the standard set of test images adopted by [1] is given in Table 4.

Figure 14 shows a test image before and after denoising, as well as the original and the approximated denoising filter banks. As noted in the text, the two filter banks differ only in the aspect of some very noisy filter, that result a bit smoothed. As these filters are not very relevant for the denoising process itself, this smoothing turned out to be irrelevant in terms of the Peak Signal-to-Noise Ratio score obtained. They could have been, however, retrieved with more precision by a larger separable filter bank.

References

[1] M. Elad and M. Aharon. Image Denoising via Sparse and Redundant Representations Over Learned Dictionaries. *TIP*, 2006. 2, 13, 15

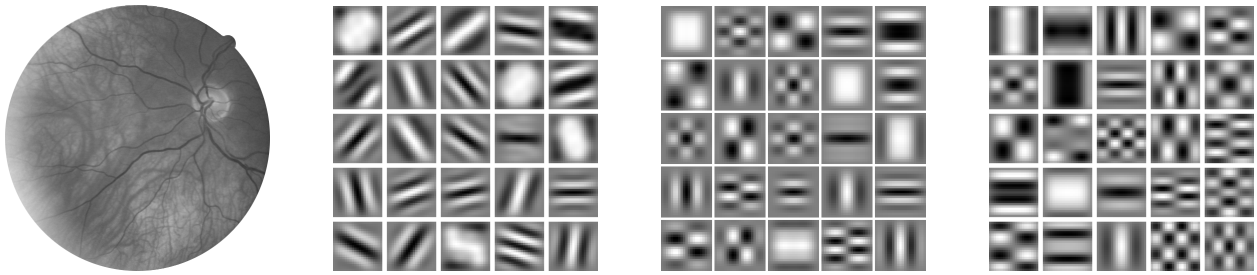
[2] M. Meilă. Comparing Clusterings - an Information Based Distance. *JMVA*, 2007. 1

[3] R. Rigamonti, E. Türetken, G. González, P. Fua, and V. Lepetit. Filter Learning for Linear Structure Segmentation. Technical report, EPFL, 2011. 1, 5

[4] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken. Ridge Based Vessel Segmentation in Color Images of the Retina. *TMI*, 2004. 3

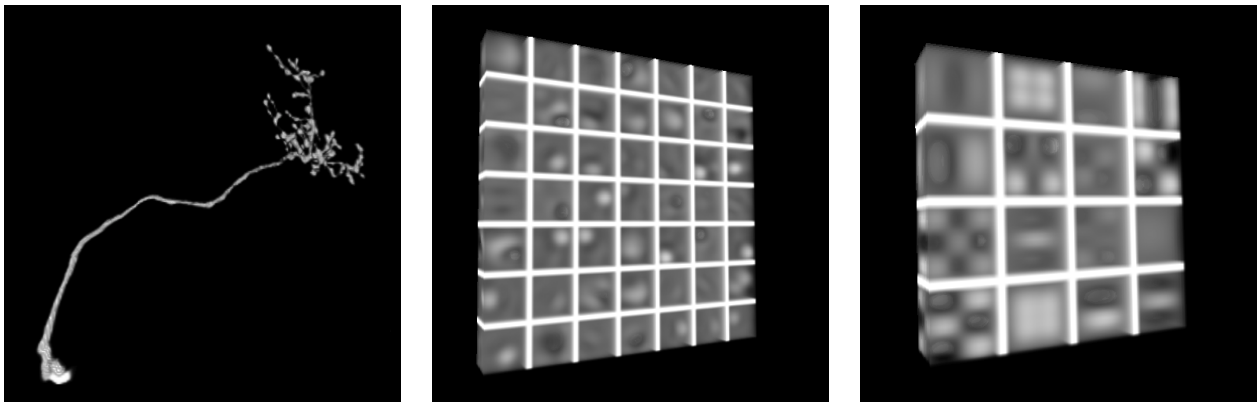
[5] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward Objective Evaluation of Image Segmentation Algorithms. *PAMI*, 2007. 1

[6] C. van Rijsbergen. Foundation of Evaluation. *Journal of Documentation*, 1974. 1



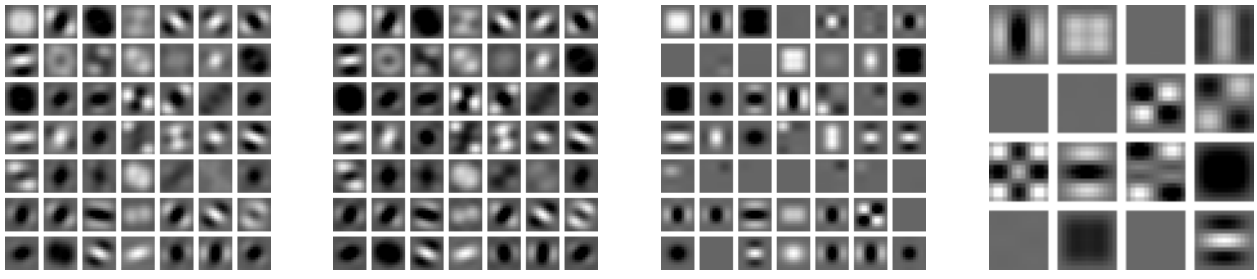
(a) **(b)** **(c)** **(d)**

Figure 1. Examples of non-separable and separable filter banks, learned on the DRIVE dataset [4]. **(a)** One of the training images. **(b)** Example non-separable filter bank learned by optimizing Eq. (1). **(c)** Its rank-1 approximation via SVD. **(d)** The separable filter bank learned by optimizing Eq. (2) of the paper.



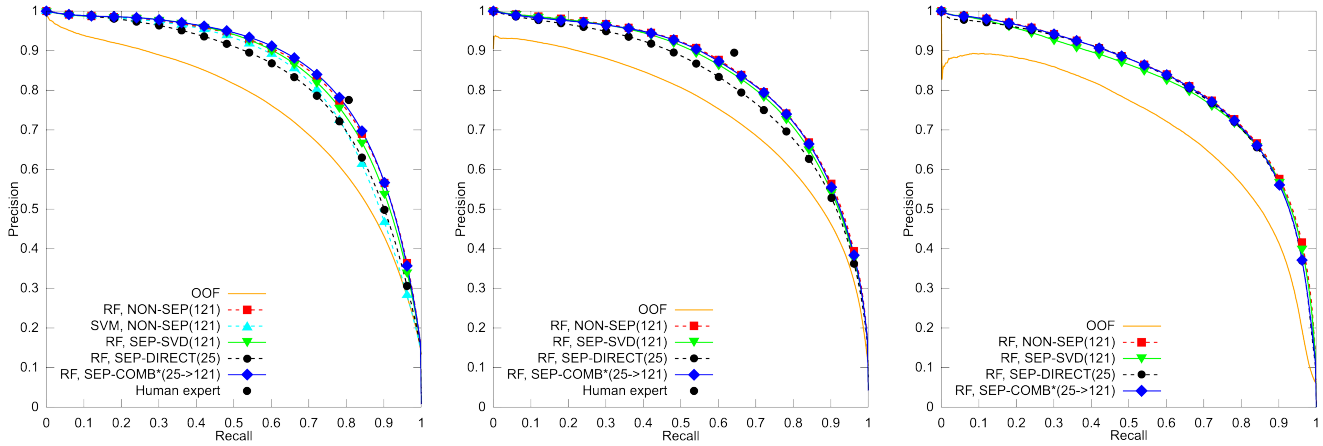
(a) **(b)** **(c)**

Figure 2. Three dimensional filter banks learned on the OPF dataset: **(a)** one of the images of the dataset. **(b)** Non separable filter bank learned optimizing Eq. (1). **(c)** Separable filter bank learned optimizing Eq. (5) to approximate (b).



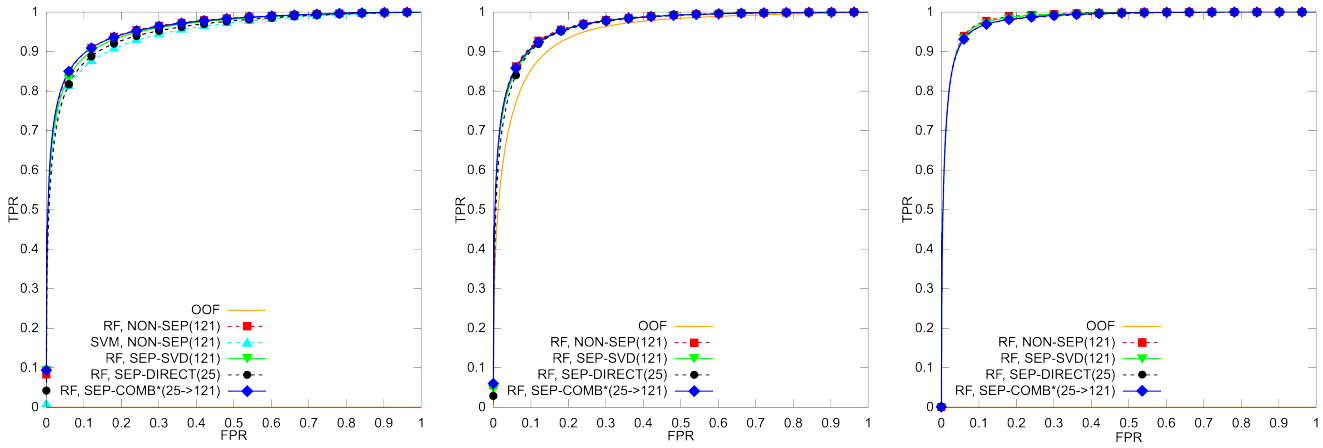
(a) **(b)** **(c)** **(d)**

Figure 3. Central slices of different 3D filter banks: **(a)** The learned filter bank of Figure 2(b). **(b)** Reconstruction obtained using the separable filter bank of Figure 2(c). **(c)** Rank one approximation using CP-Decomposition. **(d)** The separable filter bank of Figure 2(c).



(a) **(b)** **(c)**

Figure 4. PR curves evaluating the pixel classification accuracy on 2D biomedical images. The performance of our *SEP-COMB**(25) approach are virtually indistinguishable from those of the non-separable filter bank. **(a)** Retinal scans from the DRIVE dataset. **(b)** Retinal scans from the STARE dataset. **(c)** Neural structures from the BF2D dataset.



(a) **(b)** **(c)**

Figure 5. ROC curves evaluating the pixel classification accuracy on 2D biomedical images. **(a)** Retinal scans from the DRIVE dataset. **(b)** Retinal scans from the STARE dataset. **(c)** Neural structures from the BF2D dataset.

Table 1. Analytic measures of the quality of the pixel classification task over the different datasets. The VI and the RI values are computed on the classification thresholded at the value found using the F-measure. VI assumes values in $[\,0,\,\infty)$, the lower the better, and RI assumes values in $[0,1]$, the higher the better. The values are averages over 5 random trials and over the whole dataset images. For the learning-based approaches, a training set of 50,000 positive and 50,000 negative samples and a Random Forests classifier have been used, except for the SVM case where the number of samples was limited to 2,500 as in [3]. In all the considered datasets, the performance of *SEP-COMB**(25) is virtually identical to that of the non-separable filter bank of *NON-SEP*(121), but the computational costs are reduced by at least a factor of 10.

Method	**AUC**	**F-measure**	**VI**	**RI**	**Time[s]**
DRIVE					
Ground truth	-	0.788	0.380	0.930	-
OOF	0.933	0.695	0.569	0.770	5.70
NON-SEP(121)	0.959	0.782	0.554	0.890	2.22
NON-SEP(121)-SVM	0.944	0.764	0.590	0.800	2.22
NON-SEP-FFT(121)	0.959	0.782	0.554	0.890	4.79
SEP-SVD(121)	0.955	0.773	0.563	0.887	1.02
SEP-DIRECT(25)	0.948	0.756	0.602	0.879	0.23
SEP-COMB(25)	0.959	0.785	0.541	0.894	0.33
SEP-COMB*(25)	0.959	0.785	0.541	0.894	0.23
BF2D					
OOF	0.958	0.677	0.325	0.891	15.88
NON-SEP(121)	0.983	0.754	0.300	0.945	11.42
NON-SEP-FFT(121)	0.983	0.754	0.300	0.945	23.04
SEP-SVD(121)	0.982	0.749	0.306	0.943	6.67
SEP-DIRECT(25)	0.980	0.750	0.306	0.944	1.44
SEP-COMB(25)	0.981	0.752	0.301	0.944	2.11
SEP-COMB*(25)	0.981	0.752	0.301	0.944	1.44
STARE					
Ground truth	-	0.740	0.424	0.909	-
OOF	0.946	0.691	0.488	0.815	6.12
NON-SEP(121)	0.968	0.769	0.537	0.885	2.74
NON-SEP-FFT(121)	0.968	0.769	0.537	0.885	4.89
SEP-SVD(121)	0.965	0.760	0.548	0.882	1.28
SEP-DIRECT(25)	0.963	0.743	0.580	0.873	0.28
SEP-COMB(25)	0.966	0.767	0.539	0.885	0.41
SEP-COMB*(25)	0.966	0.767	0.539	0.885	0.28

Table 2. Time required to filter a $512 \times 512 \times 64$ 3D volume with different strategies. The volume size has been chosen to favor the competing FFT implementation. A filter bank composed by 49 $21 \times 21 \times 21$ filters is considered. Timings are averaged over 10 random experiments. From the table, separability emerges as the key factor for improving speed, while the reduction in the number of filters alone is clearly not sufficient. Moreover, expressing a filter bank on a smaller spanning basis not remove the dependency on the dimensionality of the input.

Method	**Time required**
FFT-based convolution(49)	268.00 ± 1.82
FFT-based convolution with spanning basis($16 \rightarrow 49$*)*	89.13 ± 0.78
Separable convolution(49)	32.88 ± 0.63
Separable convolution with spanning basis($16 \rightarrow 49$*)*	13.85 ± 0.26

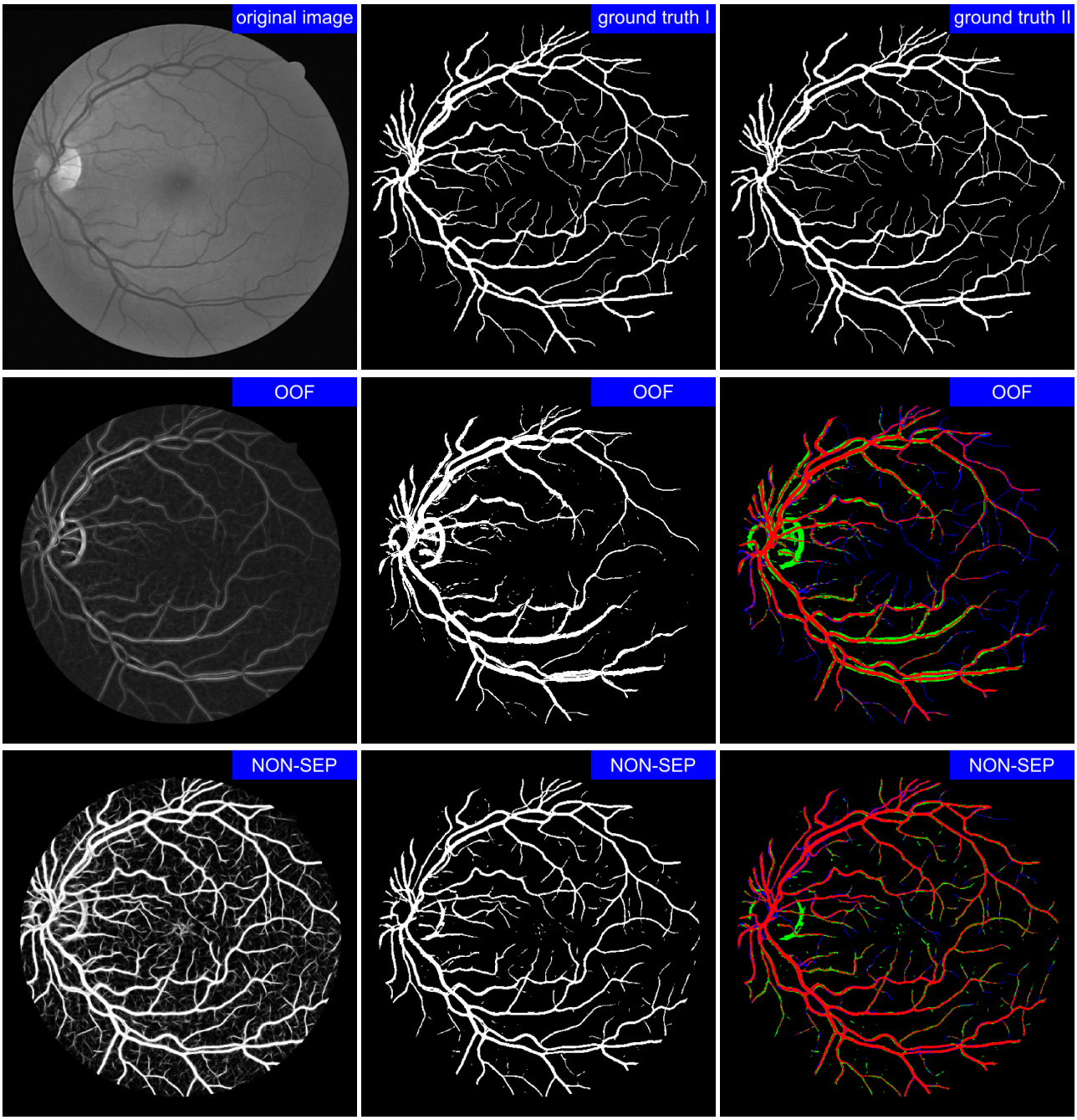


Figure 6. Classification examples for a randomly selected image from the DRIVE dataset (Part I). The first column presents the score image, the central column the binarized version at the threshold where the highest F-measure is attained. In the colorized segmentations (right-most column), true positives are outlined in red, false positives in green, and false negatives in blue. **(Top row):** Original test image along with the two ground truth provided by two ophthalmologists. **(Central row):** Optimally Oriented Flux filter (OOF) response. **(Bottom row):** Response obtained by the original, non-separable filter bank.

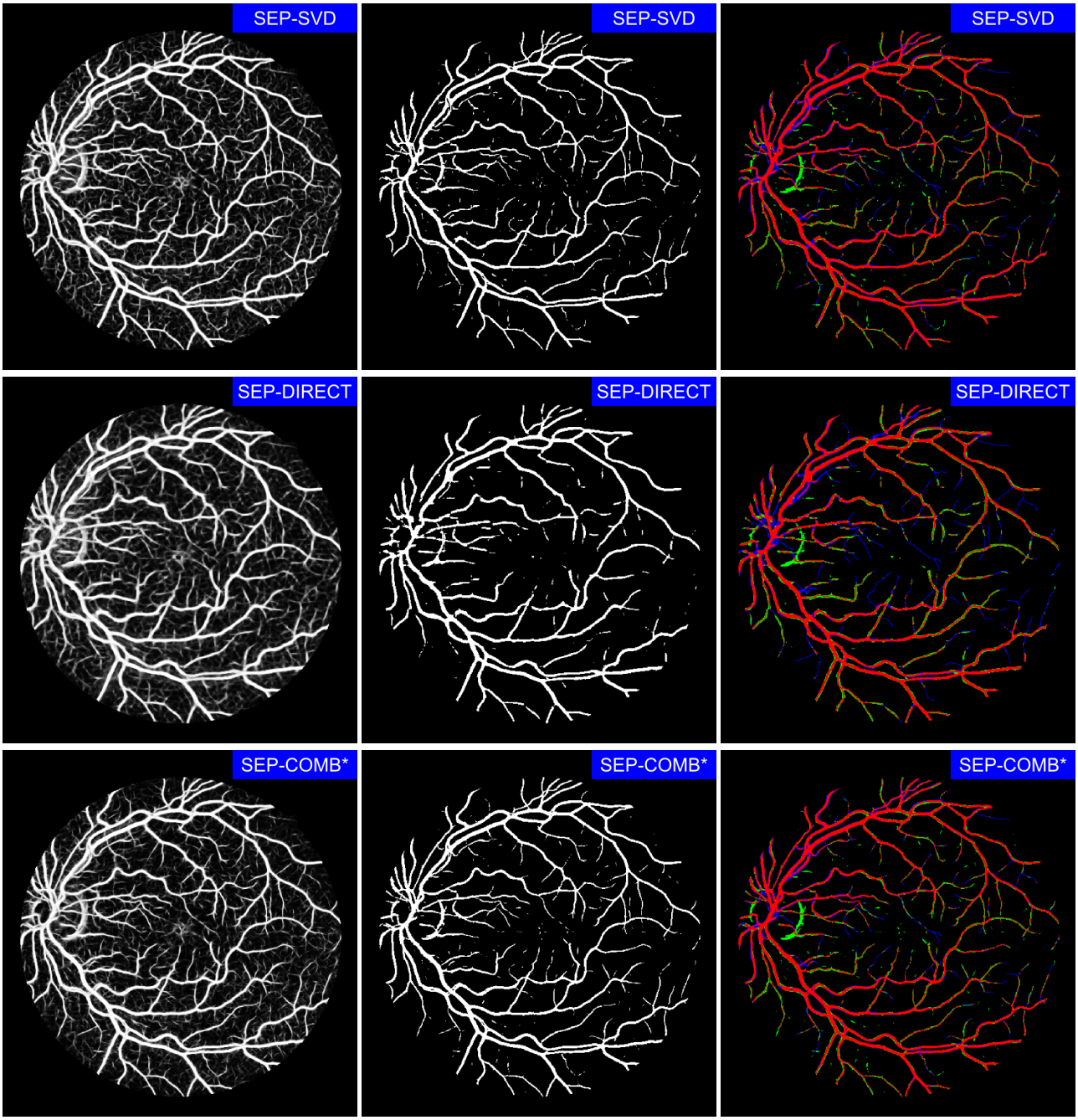


Figure 7. Classification examples for a randomly selected image from the DRIVE dataset (Part II). The first column presents the score image, the central column the binarized version at the threshold where the highest F-measure is attained. In the colorized segmentations (right-most column), true positives are outlined in red, false positives in green, and false negatives in blue. **(Top row):** Response obtained by the *SEP-SVD*(121) approach. **(Central row):** Response obtained by the *SEP-DIRECT*(25) approach. **(Bottom row):** Response obtained by the *SEP-COMB**(25) approach. As shown in the figures, the score image obtained by *SEP-COMB** is indistinguishable from the response of the non-separable filter bank, and gives excellent classification results.

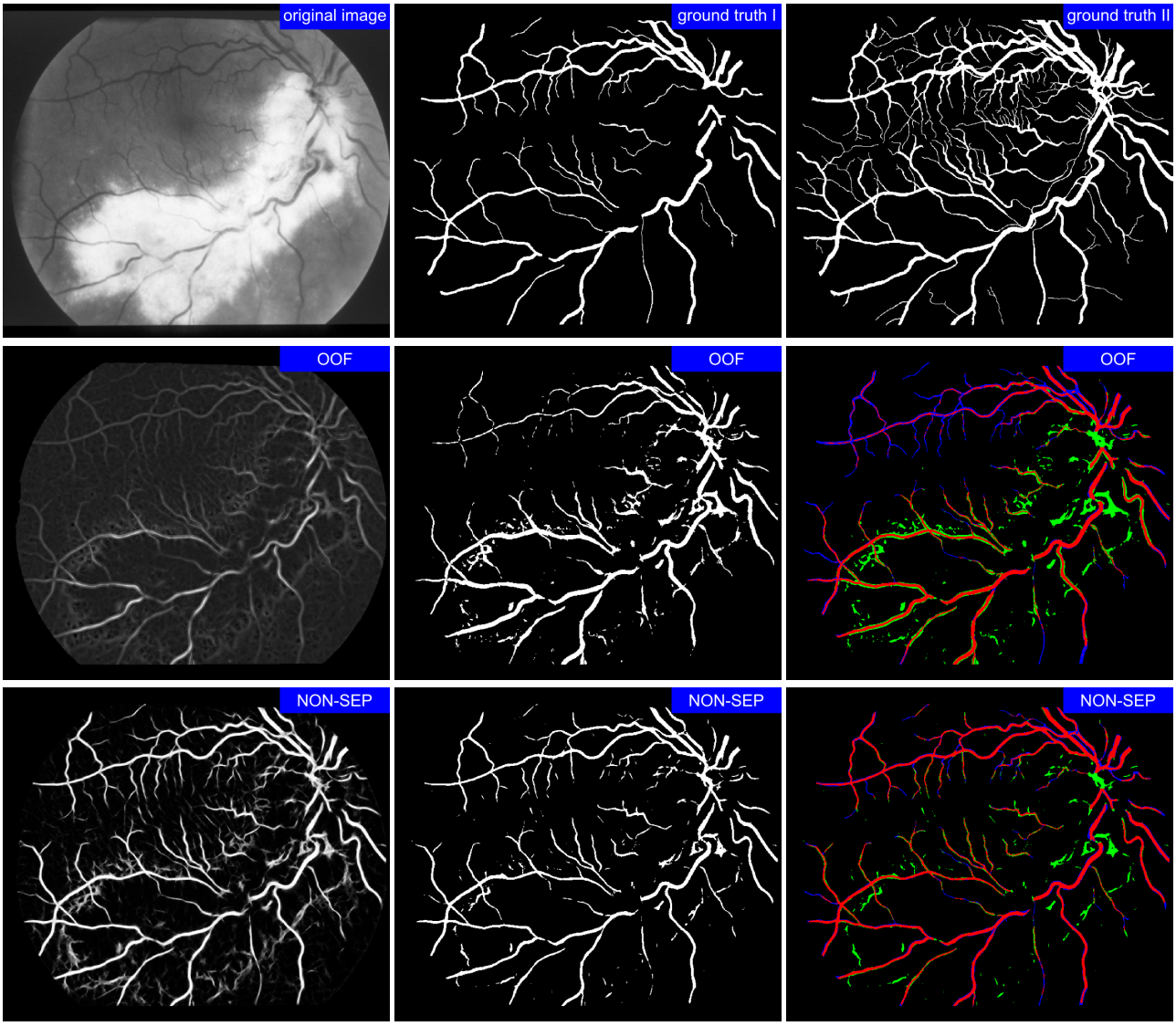


Figure 8. Classification examples for a randomly selected image from the STARE dataset (Part I). The first column presents the score image, the central column the binarized version at the threshold where the highest F-measure is attained. In the colorized segmentations (right-most column), true positives are outlined in red, false positives in green, and false negatives in blue. **(Top row):** Original test image along with the two ground truth provided by two ophthalmologists. **(Central row):** Optimally Oriented Flux filter (OOF) response. **(Bottom row):** Response obtained by the original, non-separable filter bank.

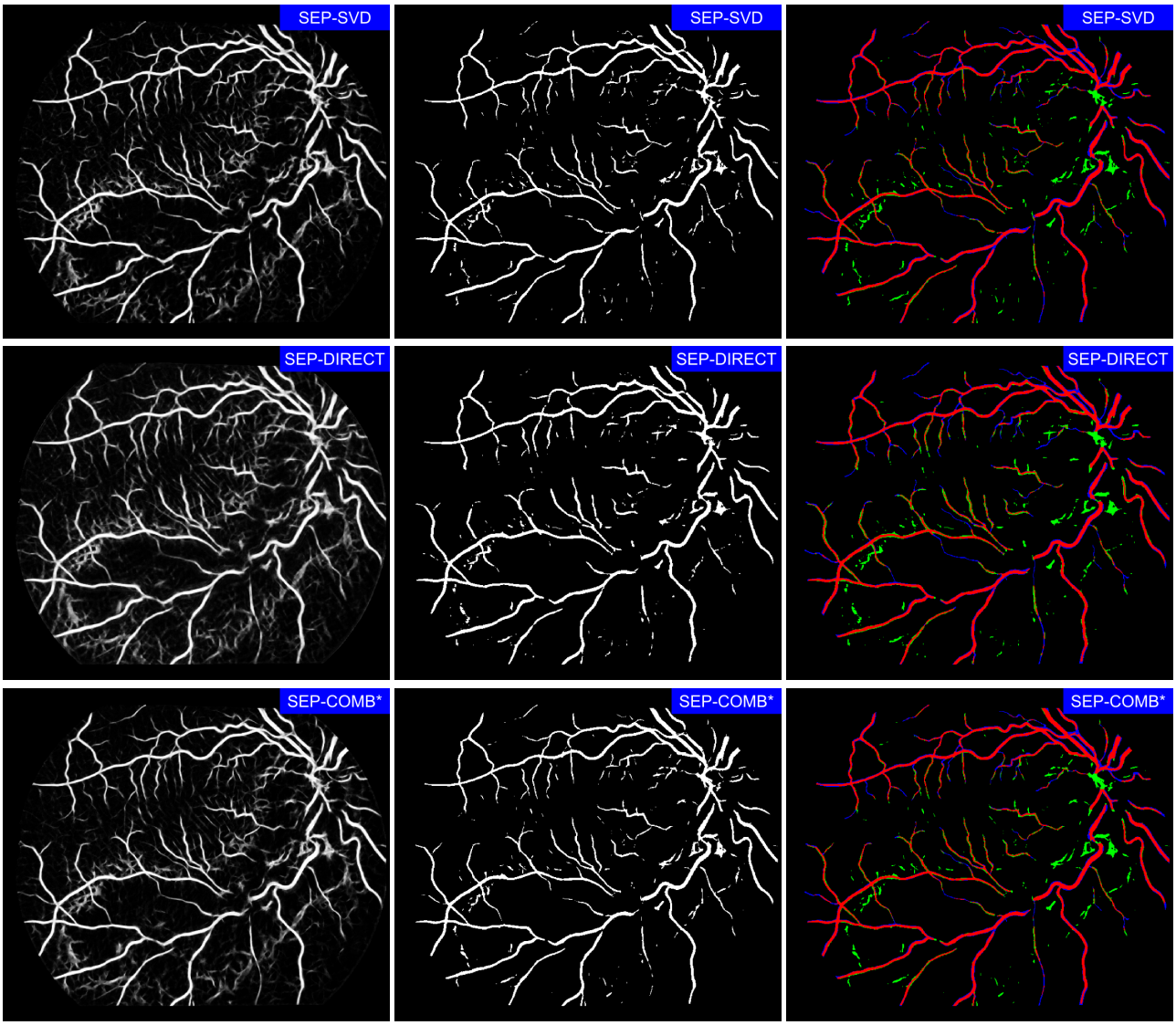


Figure 9. Classification examples for a randomly selected image from the STARE dataset (Part II). The first column presents the score image, the central column the binarized version at the threshold where the highest F-measure is attained. In the colorized segmentations (right-most column), true positives are outlined in red, false positives in green, and false negatives in blue. **(Top row):** Response obtained by the *SEP-SVD*(121) approach. **(Central row):** Response obtained by the *SEP-DIRECT*(25) approach. **(Bottom row):** Response obtained by the *SEP-COMB**(25) approach. As shown in the figures, the score image obtained by *SEP-COMB** is indistinguishable from the response of the non-separable filter bank, and gives excellent classification results.

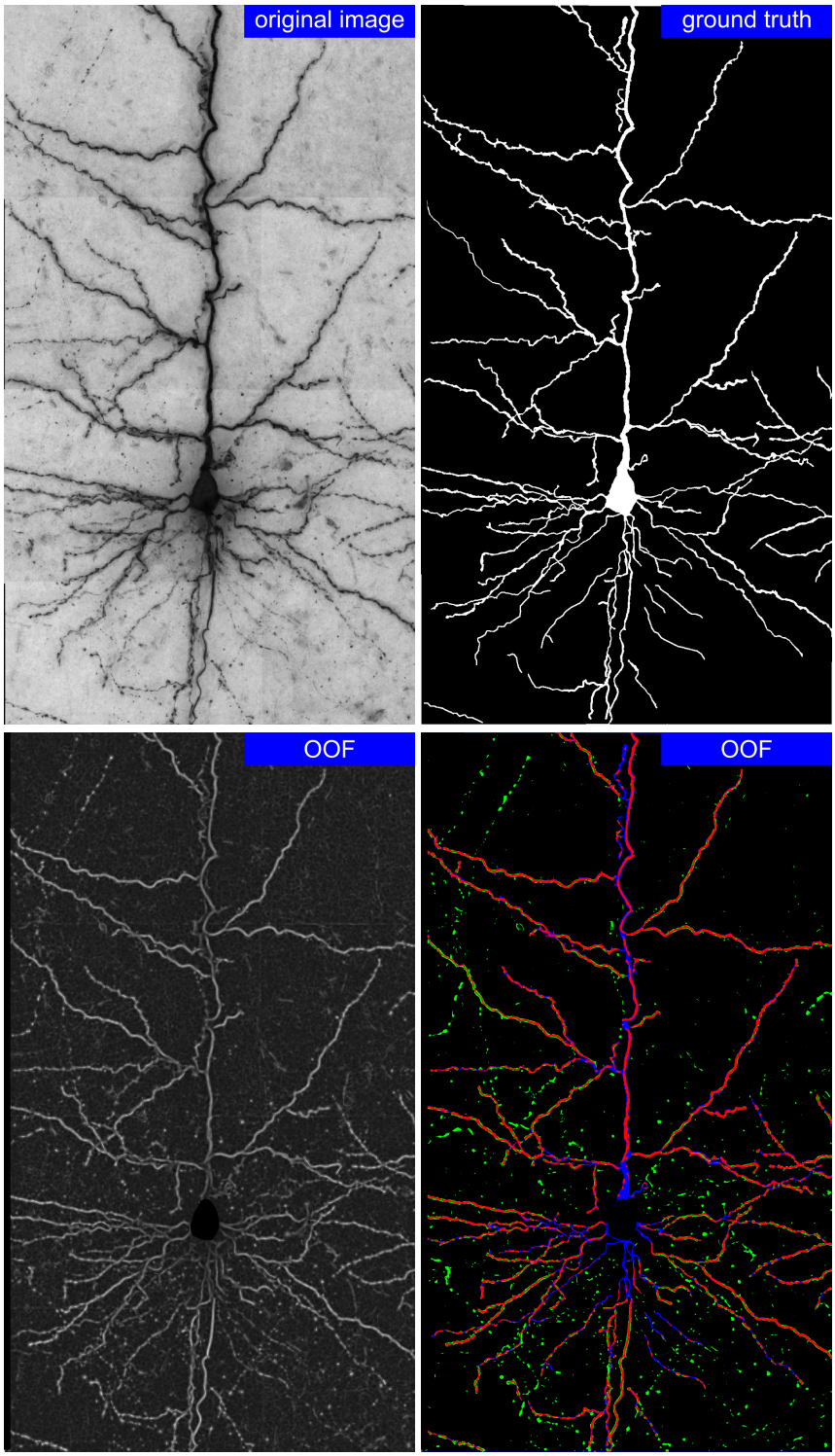


Figure 10. Classification examples for the test image of the BF2D dataset (Part I). The first column presents the score image, the second column the colorized segmentations, where true positives are outlined in red, false positives in green, and false negatives in blue. **(Top row):** Original test image along with the ground truth. **(Bottom row):** Optimally Oriented Flux filter (OOF) response.

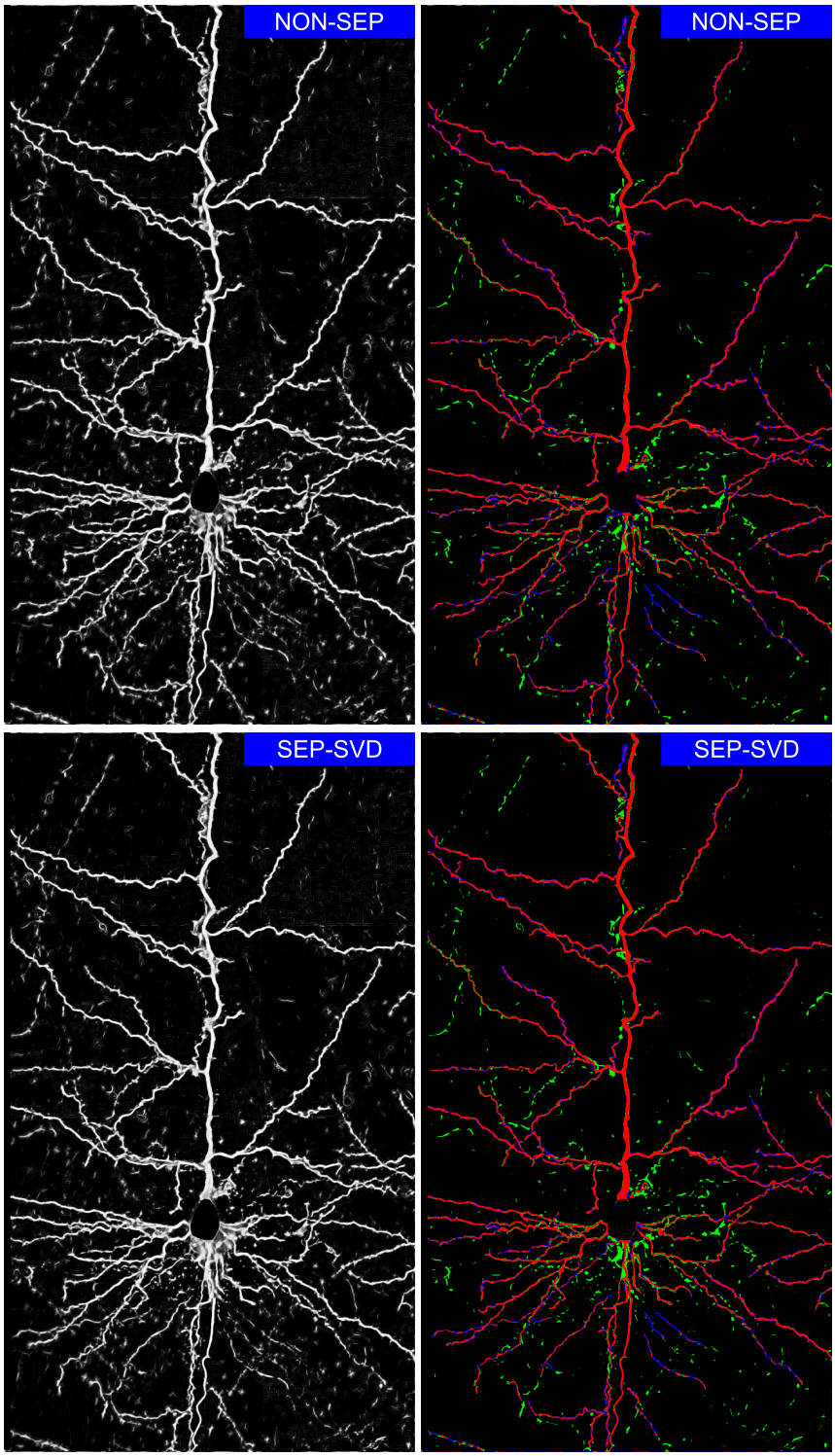


Figure 11. Classification examples for the test image of the BF2D dataset (Part II). The first column presents the score image, the second column the colorized segmentations, where true positives are outlined in red, false positives in green, and false negatives in blue. **(Top row):** Response of the original, non-separable filter bank. **(Bottom row):** Response obtained by the *SEP-SVD*(121) approach.

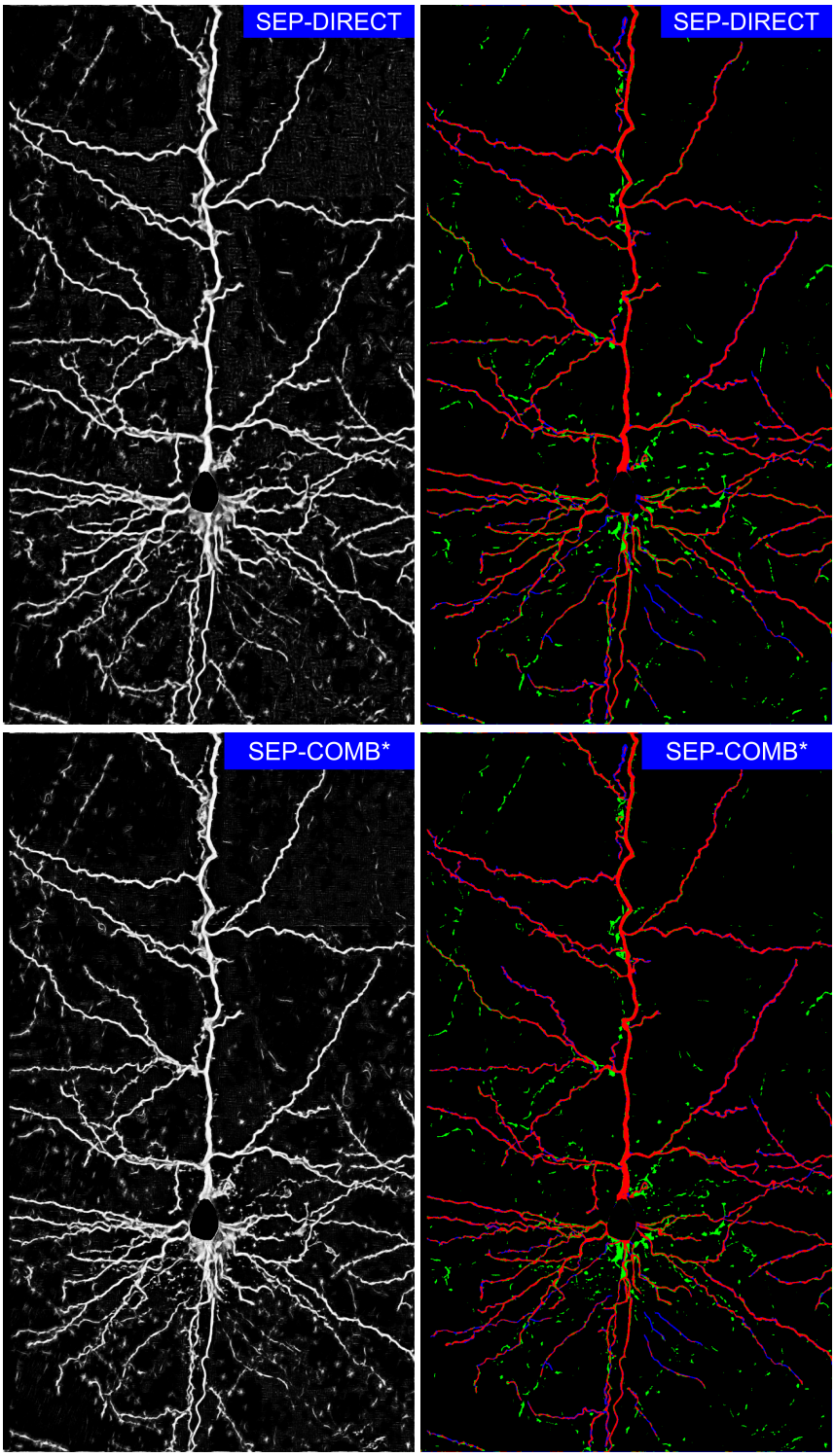


Figure 12. Classification examples for the test image of the BF2D dataset (Part III). The first column presents the score image, the second column the colorized segmentations, where true positives are outlined in red, false positives in green, and false negatives in blue. **(Top row):** Response obtained by the *SEP-DIRECT*(25) approach. **(Bottom row):** Response obtained by the *SEP-COMB**(25) approach.

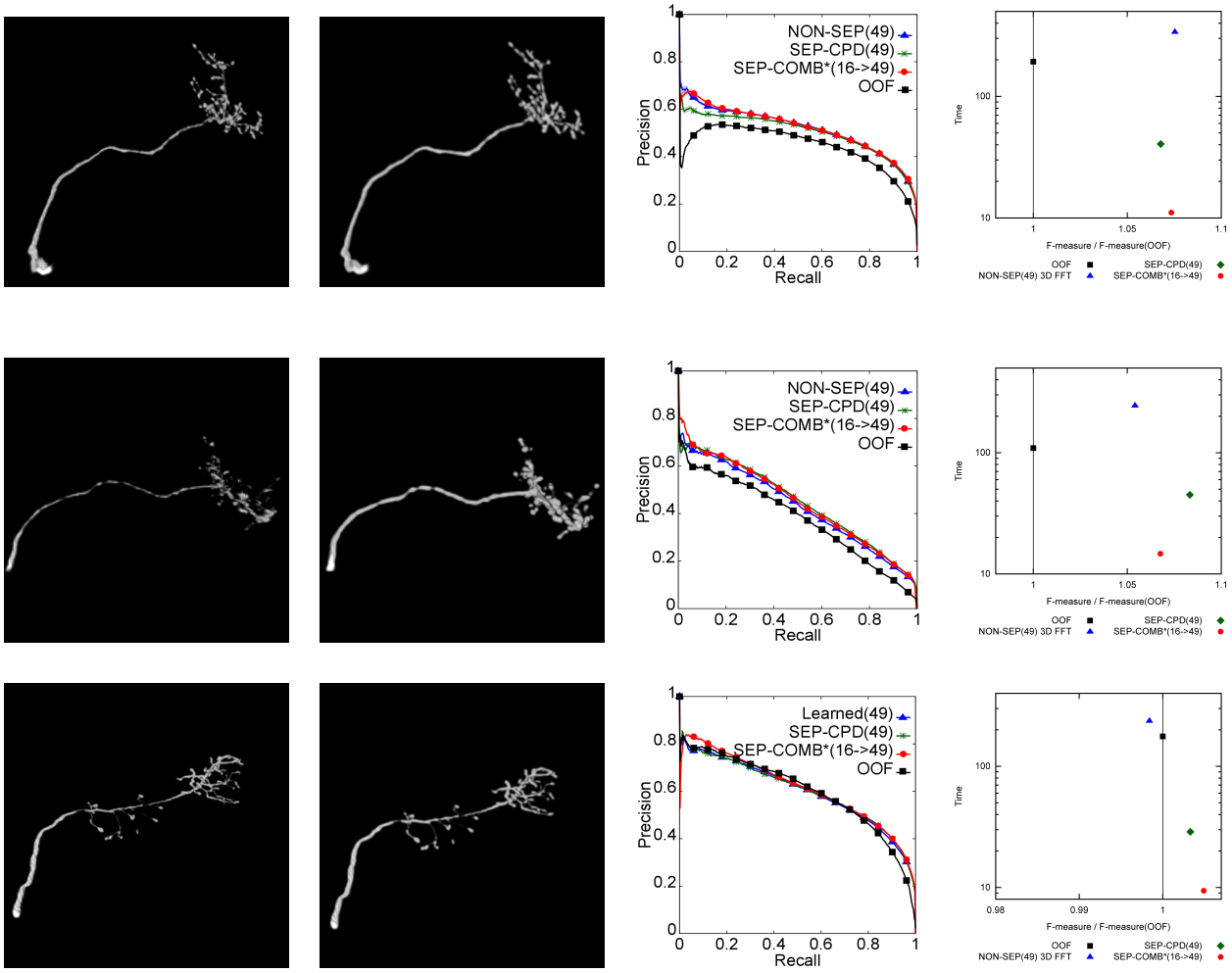
Table 3. Analytic measures of the quality of the voxel classification task on the OPF dataset. As in the 2D case, the VI and the RI values are computed on the classification thresholded at the value found using the F-measure. The values are averages over 5 random trials and over the whole dataset images. For the learning-based approaches, a training set of 200,000 randomly selected samples and ℓ_1-regularized logistic regressors classifier have been used.

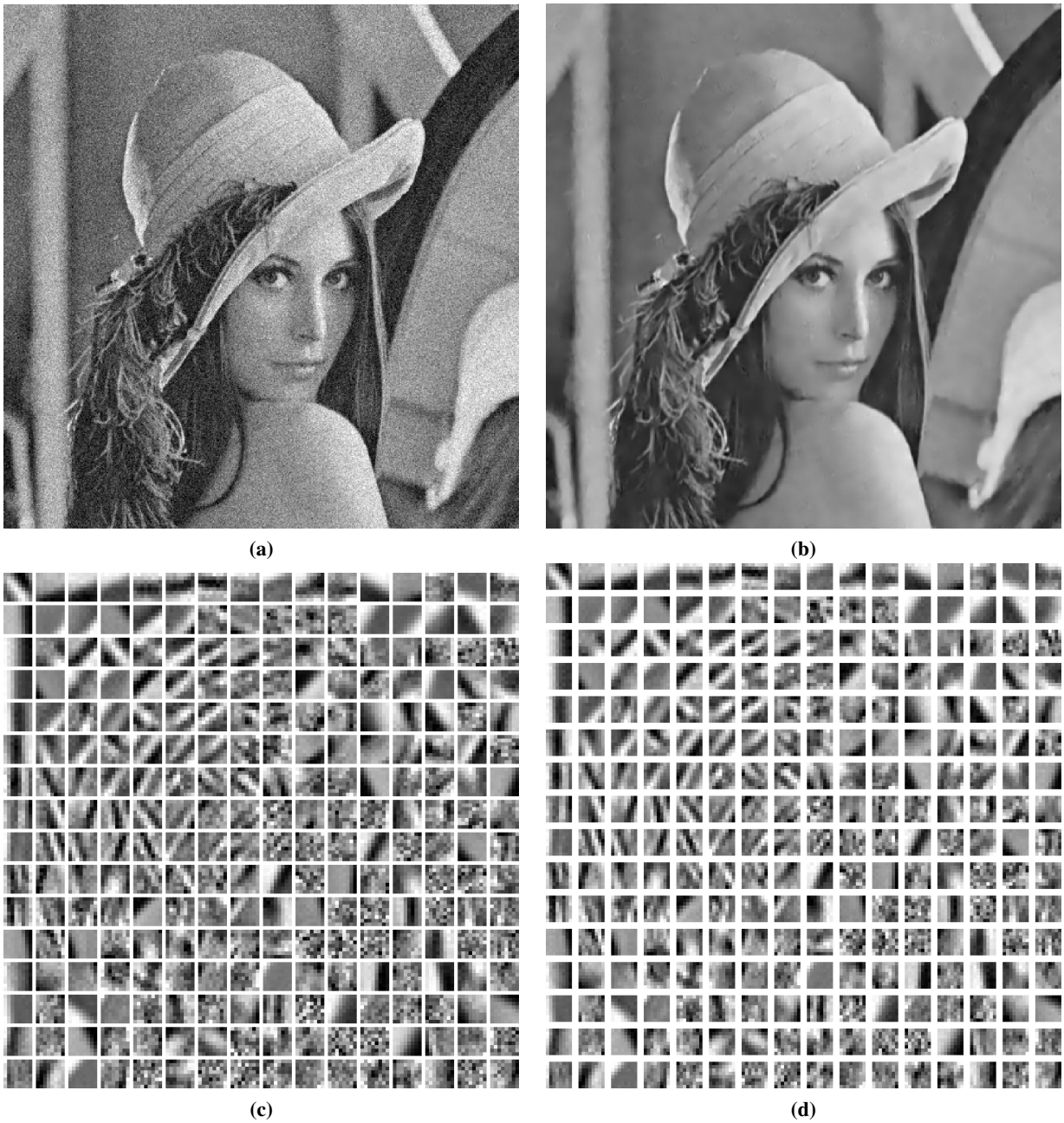
Method	**AUC**	**F-measure**	**VI**	**RI**	**Time [s]**
OPF: Image 1					
OOF	0.997	0.607	0.012	0.998	176.53
NON-SEP(49)	0.997	0.606	0.013	0.998	237.06
CPD(49)	0.998	0.609	0.012	0.998	28.79
SEP-COMB*(16→49)	0.998	0.610	0.014	0.998	9.40
OPF: Image 4					
OOF	0.997	0.531	0.012	0.998	193.05
NON-SEP(49)	0.997	0.571	0.013	0.998	339.01
CPD(49)	0.997	0.567	0.013	0.998	40.06
SEP-COMB*(16→49)	0.997	0.570	0.013	0.998	11.08
OPF: Image 9					
OOF	0.997	0.444	0.005	0.999	109.28
NON-SEP(49)	0.998	0.468	0.015	0.998	245.16
CPD(49)	0.997	0.481	0.014	0.998	44.97
SEP-COMB*(16→49)	0.998	0.474	0.014	0.998	14.68

Table 4. Results for the image denoising task. The values in the upper part of the table are expressed in decibels, and represent the image Peak Signal-to-Noise Ratio (PSNR). The images were artificially corrupted by additive white Gaussian noise with standard deviation 20. The results have been obtained by replacing the filter bank learned by the K-SVD algorithm with the approximated one in the code provided by [1]. *SEP-COMB-Barbara* denotes the strategy where, instead of grounding the reconstruction on the approximating filter bank corresponding to the image to denoise, the approximating filter bank from the Barbara image is used. In the lower part of the table, the average reconstruction error for the different filter banks, measured as $\mathbb{E}[\|\mathbf{f}^j - \sum_k w_k^j \mathbf{s}_k\| / (\|\mathbf{f}^j\| \|\sum_k w_k^j \mathbf{s}_k\|)]$, is reported. In all of the experiments no tuning of the parameters of neither the approximation nor of the denoising algorithms was performed.

	Barbara	**Boat**	**House**	**Lena**	**Peppers**
Noisy image	22.12	22.09	22.06	22.09	22.13
K-SVD	30.88	30.36	33.34	32.42	32.25
SEP-SVD(256)	30.23	30.20	32.73	32.08	32.06
SEP-COMB(25)	30.21	30.27	33.13	32.40	31.99
SEP-COMB(36)	30.77	30.36	33.24	32.42	32.08
SEP-COMB(49)	30.87	30.36	33.32	32.42	32.17
SEP-COMB(64)	30.88	30.36	33.34	32.42	32.25
SEP-SVD-Barbara(256)	-	30.01	31.92	32.06	31.82
SEP-COMB-Barbara(36)	-	30.26	32.41	32.43	31.97
SEP-COMB-Barbara(64)	-	30.36	33.28	32.43	32.23
Average filter bank reconstruction error					
SEP-COMB(36)	0.188	0.157	0.262	0.234	0.291
SEP-COMB(64)	0.060	0.054	0.055	0.051	0.050

Figure 13. Results obtained on the OPF dataset using images number 1,4 and 9 for testing. First column: original image; Second column: response of the classifier; Third column: PR curves for different algorithms. The number between brackets indicates the number of filters used (in the case of *SEP-COMB** the first number is the number of separable filters and the second is the number of filters the original filter bank); Fourth column: F-measure normalized w.r.t. the F-measure of *OOF* plotted against time. While all the methods are almost equivalent in terms of F-measure, our approach is significantly faster.





(a) **(b)** **(c)** **(d)**

Figure 14. **(a)** Lena image corrupted with additive white Gaussian noise of standard deviation 20. The PSNR is 22.09dB. **(b)** Lena image denoised using our approximation within the algorithm of [1]. The PSNR is 32.42dB. **(c)** Filter bank learned by the K-SVD algorithm of [1] on the Lena image. **(d)** Filter bank approximated by *SEP-COMB*(36).