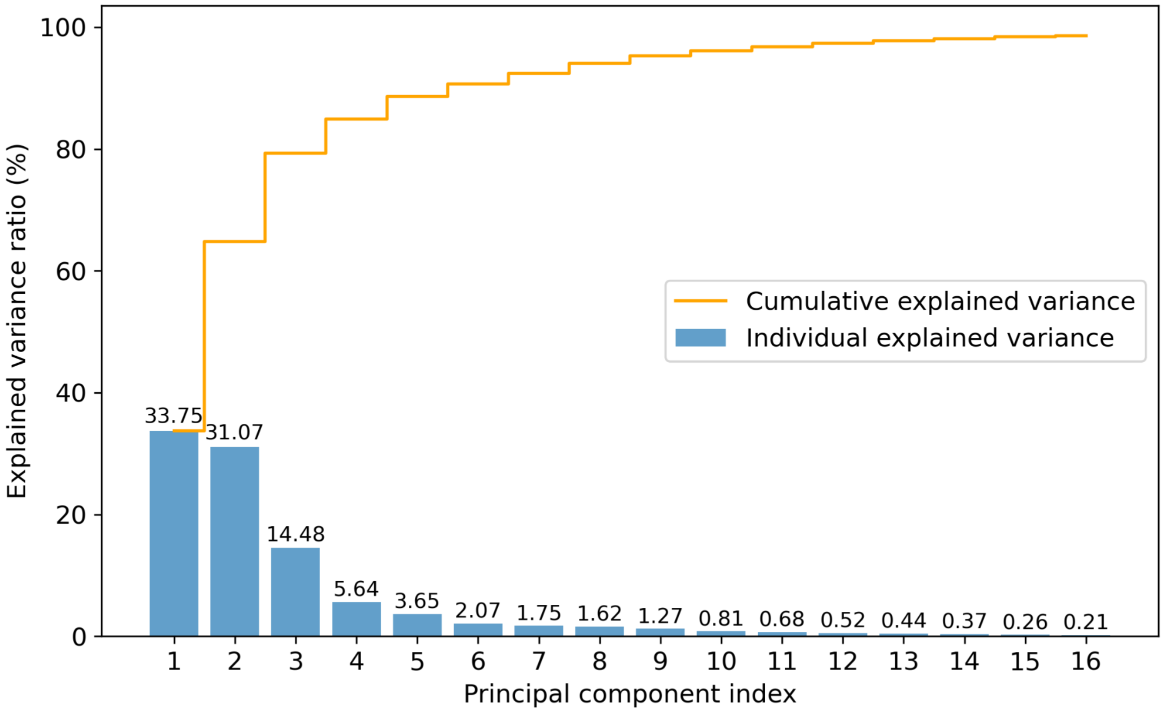
Dimensionality reduction for deep learning in infrared microscopy: A comparative computational survey

Supplementary Material

Dajana Müller, David Schuhmacher, Stephanie Schörner, Frederik Großerueschkamp, Iris Tischoff, Andrea Tannapfel, Anke Reinacher-Schick, Klaus Gerwert, Axel Mosig

List of Supplementary Figures

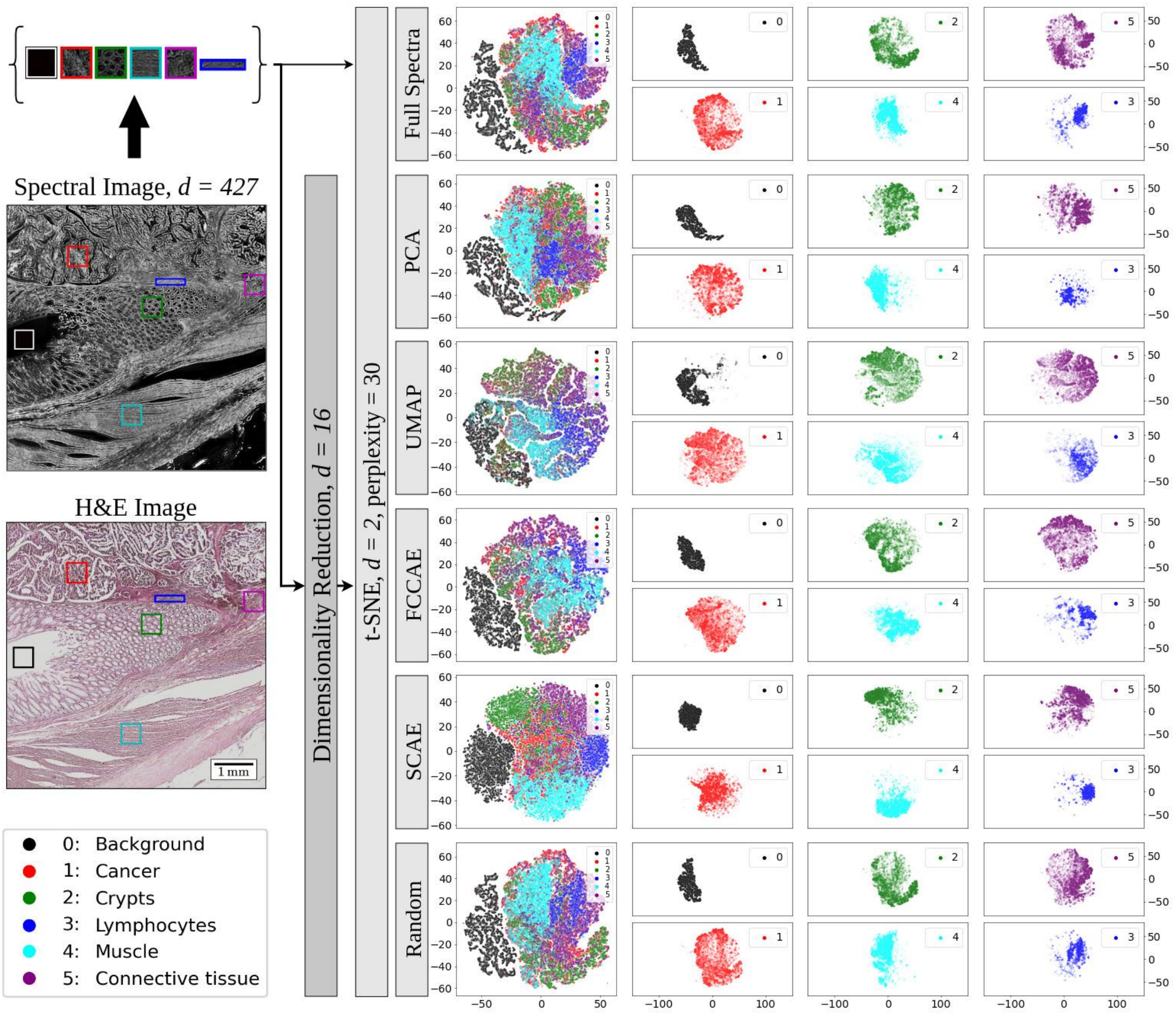
List of Supplementary Tables



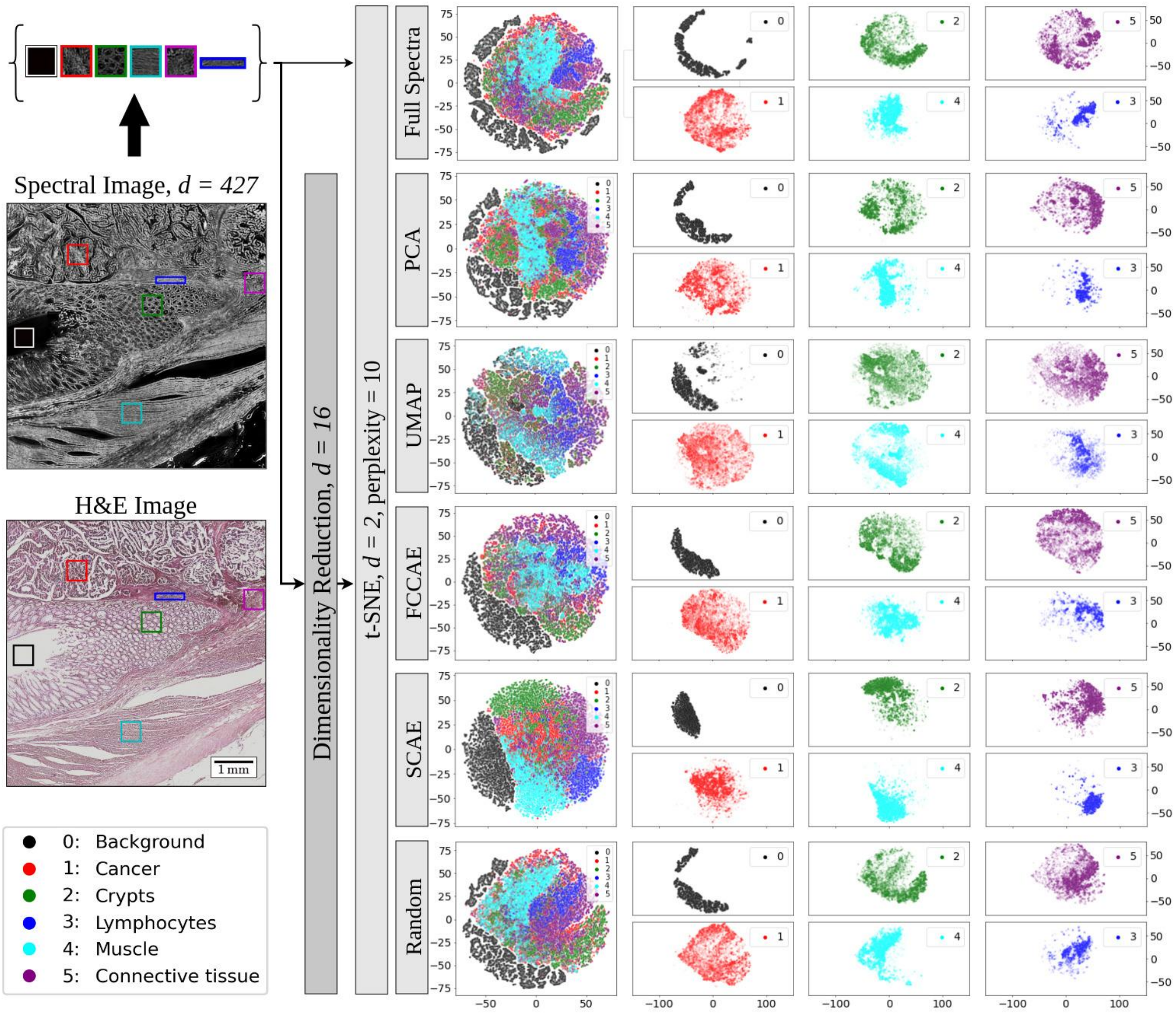
Supplementary Figure 1. Explained variance ratio of the principal components. Individual explained variance ratio (%) is plotted in blue bars and the cumulative explained variance in yellow for the first 16 principle components (PCs). The sum of the first 9 PCs exceeds 95% cumulative explained variance.

Supplementary Table 1. Training time and classification results of CompSegNet trained with dimensionality reduced datasets.

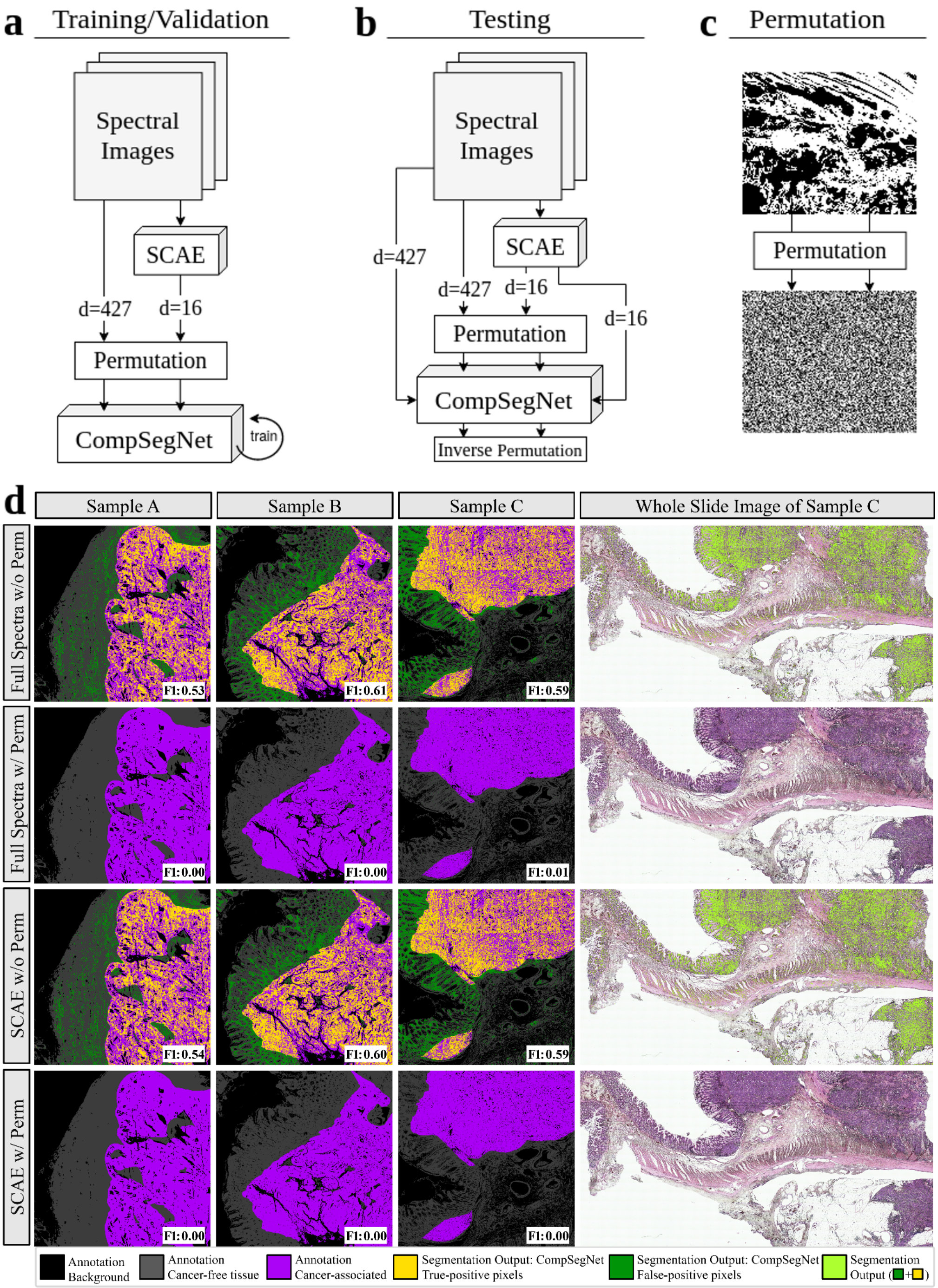
	num. of dim.	*t* train model (h)	Validation data				Testing data			
			Sens (%)	Spec (%)	Acc (%)	F1 (%)	Sens (%)	Spec (%)	Acc (%)	F1 (%)
PCA	4	~21	89.97	87.62	88.78	88.82	85.70	86.07	86.63	83.71
PCA	9	~23	98.38	98.41	98.40	98.38	94.83	96.10	95.60	94.42
PCA	32	~30	**99.03**	**100.0**	**99.52**	**99.51**	96.12	**99.16**	**97.97**	97.38
Random	4	~21	97.10	95.87	96.47	96.46	90.95	90.81	90.86	88.66
Random	9	~23	98.71	98.41	98.56	98.55	95.26	96.39	95.94	94.85
Random	32	~30	96.44	**100.0**	98.24	98.19	**96.98**	98.61	**97.97**	**97.40**



Supplementary Figure 2. Comparison of encodings with t-SNE in euclidean space. Spectral data of 6 different (tissue) classes (1, background, black; 2, cancer, red; 3, crypts, green; 4, muscle, cyan; 5, connective tissue, purple; 6, lymphocytes, blue) from a test patient have been encoded with principal component analysis (PCA), uniform manifold approximation and projection (UMAP), fully-connected contractive autoencoder (FCCAE), stacked contractive autoencoder (SCAE), and random projection into 16 dimensions and were embedded into euclidean space with t-SNE (perplexity = 30). t-SNE was also conducted on full spectra for comparison.



Supplementary Figure 3. Comparison of encodings with t-SNE in euclidean space. Spectral data of 6 different (tissue) classes (1, background, black; 2, cancer, red; 3, crypts, green; 4, muscle, cyan; 5, connective tissue, purple; 6, lymphocytes, blue) from a test patient have been encoded with principal component analysis (PCA), uniform manifold approximation and projection (UMAP), fully-connected contractive autoencoder (FCCAE), stacked contractive autoencoder (SCAE), and random projection into 16 dimensions and were embedded into euclidean space with t-SNE (perplexity = 10). t-SNE was also conducted on full spectra for comparison.



Suppementary Figure 4. Workflow and whole-slide cancer segmentation on permutated infrared samples and embeddings. a: Training and validation of pixel permuted full and encoded spectra with the stacked contractive autoencoder (SCAE). b: Testing of normal and pixel permuted whole slide images on full spectra and encodings. c: Example of permutation used. d: Segmentations of 3 different samples of the independent test cohort of the ColoPredict Plus 2.0 registry study with workflow b. Black: background of sample with no tissue. Grey: Annotation of cancer-free tissue.. Purple: Cancer-associated tissue that is not detected by the neural network (false-negative). Yellow: Segmentation output of cancerous pixels (true-positive). Green: Segmentation output of hypothetical cancerous pixels (false-positive). Lime: Segmentation output (green + yellow). Annotations were performed on H&E slides by a pathologist.

Supplementary Table 2. Classification results of CompSegNet trained on permuted infrared samples and embeddings.

	num. of dim.	Validation data				Testing data			
		Sens (%)	Spec (%)	Acc (%)	F1 (%)	Sens (%)	Spec (%)	Acc (%)	F1 (%)
Full Spectra	427	98.38	97.77	98.08	98.06	96.98	93.87	95.09	93.95
SCAE	16	98.38	96.83	97.60	97.59	98.28	93.59	95.43	94.41