

Supplementary Data

For article:

DULoc: quantitatively unmixing protein subcellular location patterns in immunofluorescence images based on deep learning features

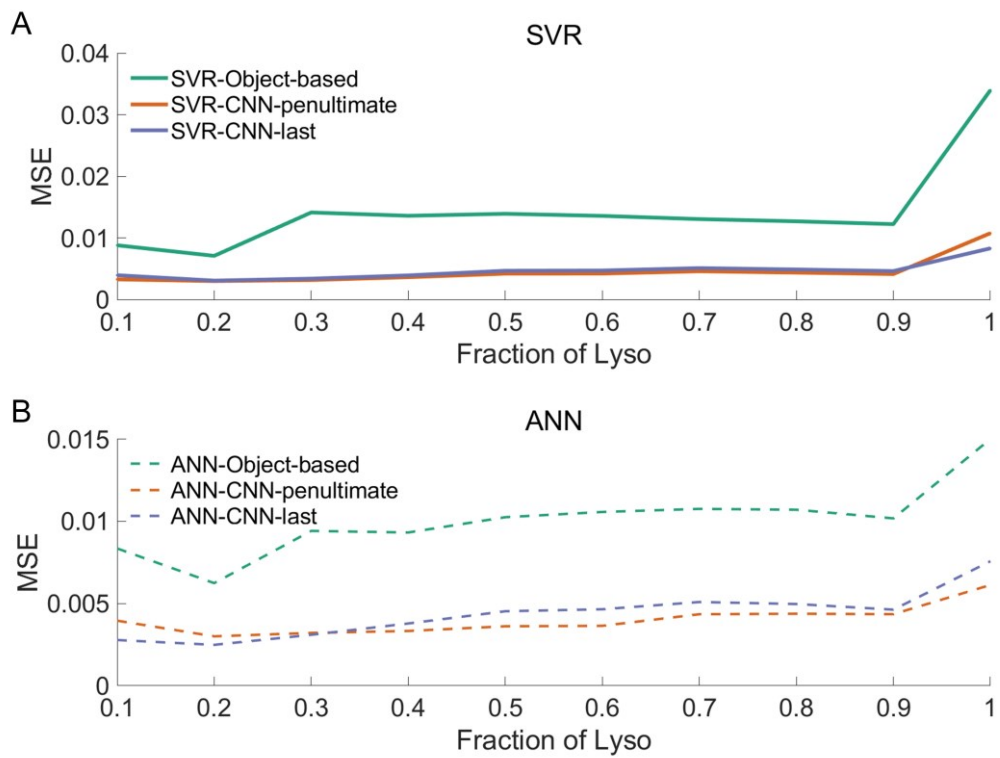
Min-Qi Xue^{1,2}, Xi-Liang Zhu^{1,2}, Ge Wang^{1,2}, and Ying-Ying Xu^{1,2,*}

¹School of Biomedical Engineering and Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou 510515, China, ²Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology, Southern Medical University, Guangzhou 510515, China

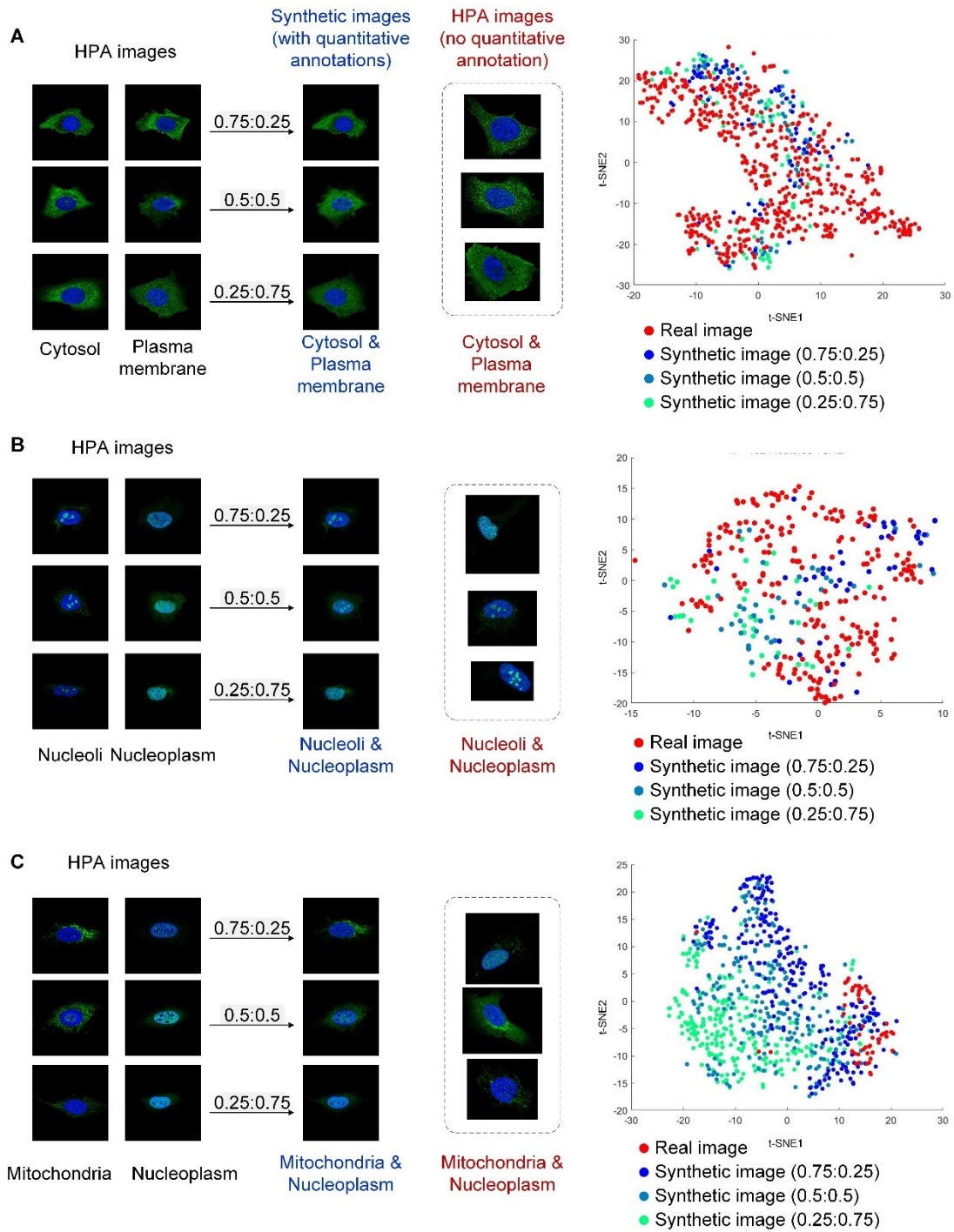
*To whom correspondence should be addressed. yyxu@smu.edu.cn

Contents

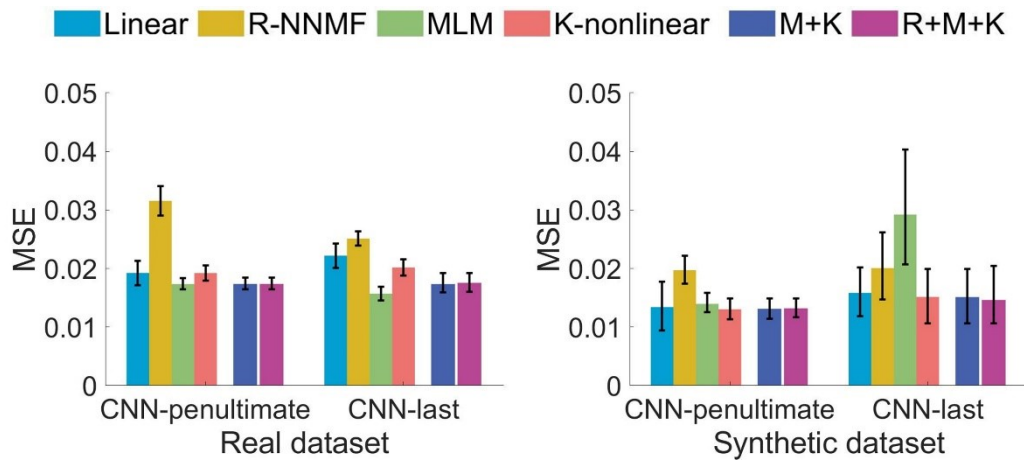
Content	Description	Page
Supplementary Figure 1	Regression results of the images with different fraction patterns in the real dataset	2
Supplementary Figure 2	Comparison of synthetic and real cell images of different combinations	3
Supplementary Figure 3	MSE results of different pattern unmixing methods	4
Supplementary Figure 4	Performance on different fraction patterns when using the pattern unmixing methods	5
Supplementary Table 1	Statistics of misclassifying multi-label proteins as single-locational	6
Supplementary Table 2	Averaged computing time required to decompose an image	7
Supplementary Table 3	Comparison of features from the penultimate and last layers used in DULoc model for unmixing HPA images	8



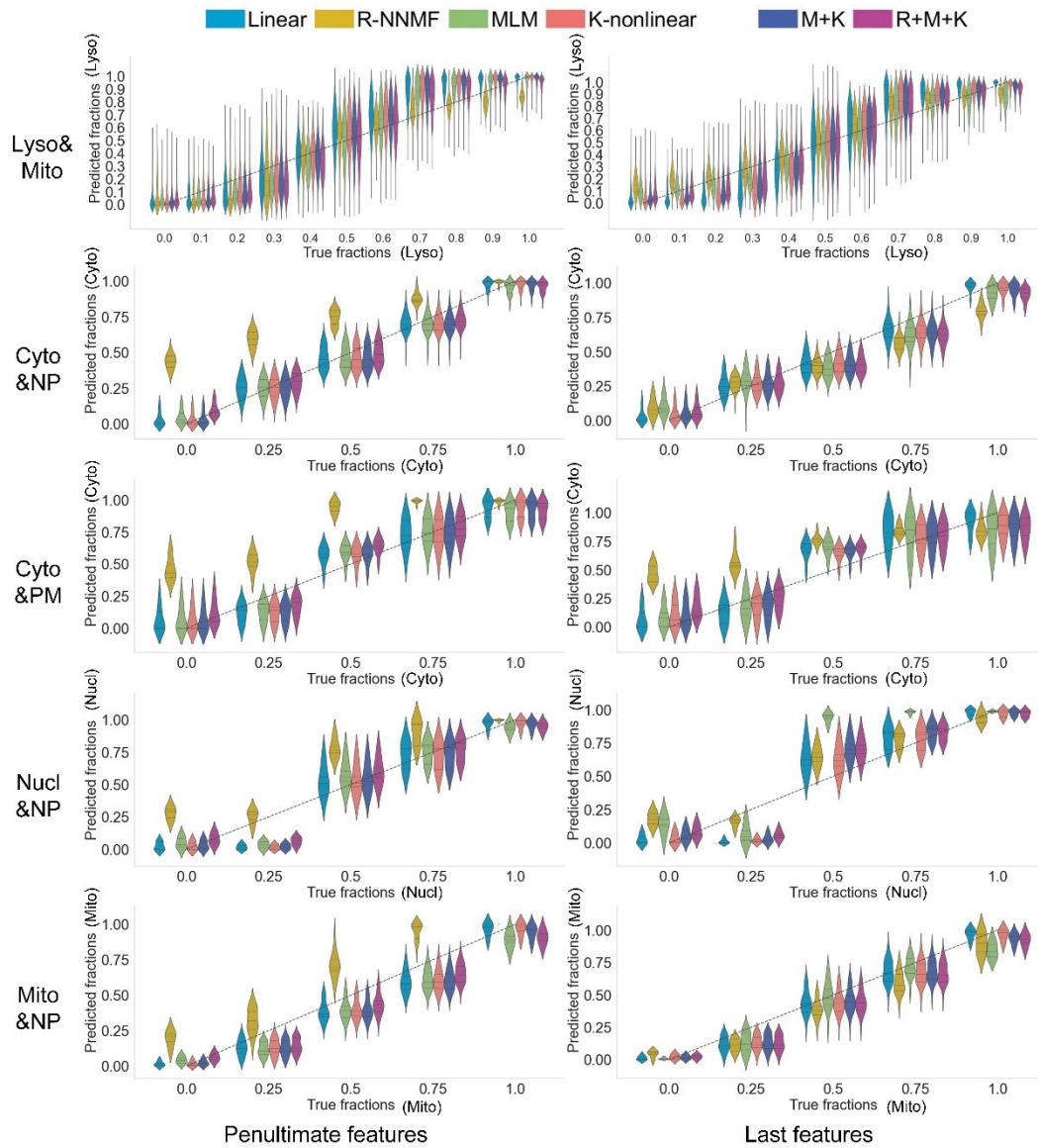
Supplementary Figure 1. Regression results of the images with different fraction patterns in the real dataset. (A) Results of SVR. (B) Results of ANN.



Supplementary Figure 2. Comparison of synthetic and real cell images of different combinations. Example images and t-SNE visualization of combination (A) cytosol and plasma membrane, (B) nucleoli and nucleoplasm, and (C) mitochondria and nucleoplasm are shown.



Supplementary Figure 3. MSE results of different pattern unmixing methods



Supplementary Figure 4. Performance on different fraction patterns when using the pattern unmixing methods. For the synthetic dataset, the unmixing methods showed poor performance on the patterns “Cytosol & Plasma membrane” and “Nucleoli & Nucleoplasm”, because the location pairs in the two combinations are too similar, and difficult to decompose.

Supplementary Table 1. Statistics of misclassifying multi-label proteins as single-locational. The statistics was based on the 1816 multi-label images in the real dataset.

Methods	Penultimate layer features		Last layer features	
	number of misclassified samples	Ratio of misclassification	number of misclassified samples	Ratio of misclassification
Linear	200	11.01%	287	15.8%
R-NNMF	0	0%	3	0.17%
MLM	36	1.98%	2	0.11%
K-nonlinear	1	0.06%	2	0.11%
M+K	0	0%	0	0%
R+M+K	0	0%	0	0%

Supplementary Table 2. Averaged computing time required to decompose an image

Output layer of the bestfitting	Number of features	Linear	R-NNMF	MLM	K-nonlinear
Penultimate	1024	0.0010s	3.2305s	0.0589s	0.0771s
Last	28	0.0011s	0.0307s	0.0195s	0.000064s

Supplementary Table 3. Comparison of the features from the penultimate and last layers used in DULoc model for unmixing HPA images

Cell line	Number of labels	Ratio of correct predictions (DULoc using the penultimate layer features)	Ratio of correct predictions (DULoc using the last layer features)
U-2 OS	Double	79.43%	81.96%
	Triple	56.83%	52.40%
	Quadruple	31.25%	37.50%
A-431	Double	77.49%	81.32%
	Triple	55.07%	58.15%
	Quadruple	28.00%	28.00%
U-251 MG	Double	72.00%	72.73%
	Triple	54.25%	52.94%
	Quadruple	42.86%	42.86%
In total		69.05%	70.52%