Machine Learning Pipeline for Segmentation and Defect Identification from High Resolution Transmission Electron Microscopy Data

Supplementary Information

C.K. Groschner[1], Christina Choi[1], M.C. Scott[1,2,*]

[1]Department of Materials Science and Engineering, University of California Berkeley, Berkeley, CA 94720

[2]Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

[*]mary.scott@berkeley.edu

1 Supplementary Information

1.1 Results From Different Training Sets

Segmentation was found to improve when a more varied dataset was used. Results for training solely on CdSe compared to CdSe and Au are presented in Table 1 and in Figure 1.

	DICE	Precision	Recall
U-Net with only CdSe Data	0.43	0.44	0.42
U-Net with Combined Data	0.59	0.56	0.62

Table 1: Performance metrics for U-Net on CdSe data when the U-Net was trained with different datasets. The same number of images were used for each training instance.

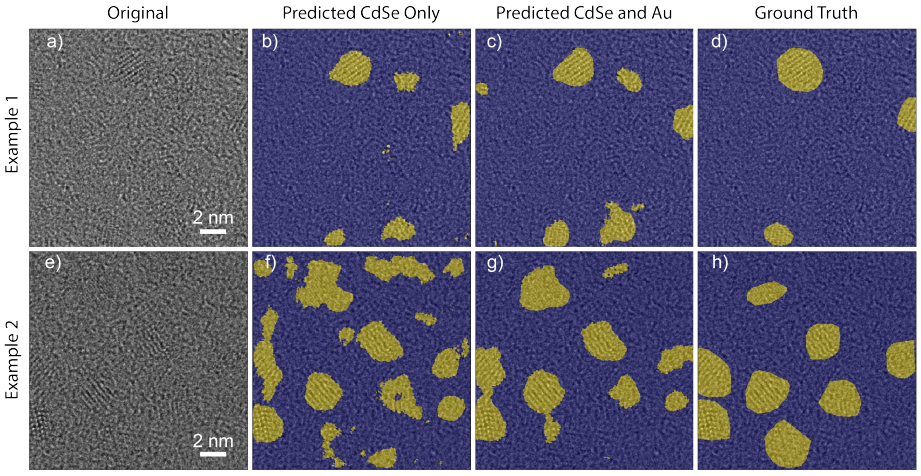
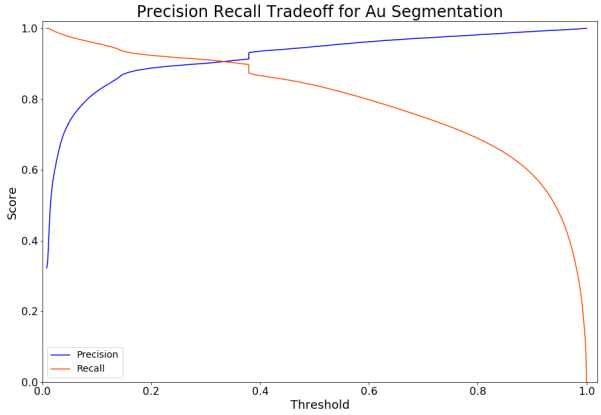


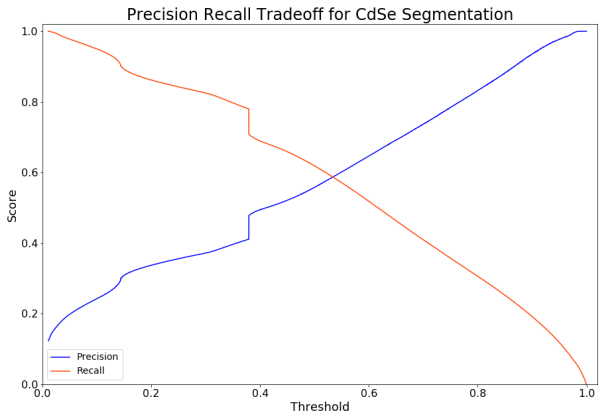
Figure 1: Two sample micrographs of CdSe particles and the resulting segmentation maps when trained with only CdSe data and when trained with Au and CdSe data.

1.2 Threshold Determination

Precision versus threshold and recall versus threshold was plotted for both the Au nanoparticle and CdSe nanoparticle segmentation. Based on the plots the optimal threshold was chosen. The plots are shown in Figure 2



(a)



(b)

Figure 2: Precision and recall vs threshold for (a) segmentation of Au nanoparticles and (b) segmentation of CdSe nanoparticles.