Supplementary information for: Machine learning of optical properties of materials - predicting spectra from images and images from spectra

Helge S. Stein,[a,*] Dan Guevara [a], Paul F. Newhouse [a], Edwin Soedarmadji [a], John M. Gregoire [a,*]

[a.] *Joint Center for Artificial Photosynthesis, California Institute of Technology, Pasadena, California 91125 (USA)*
*stein@caltech.edu, gregoire@caltech.edu

A visualization of the models described in the main text and methods section is shown in Figure S1 as graph representations. The lambda layer in the VAE model samples the latent space via:

$$\square = \square + \square * \square * \square^{\square/2}$$

as described in the main text. The code for all models and the trained weights are included as separate SI files.

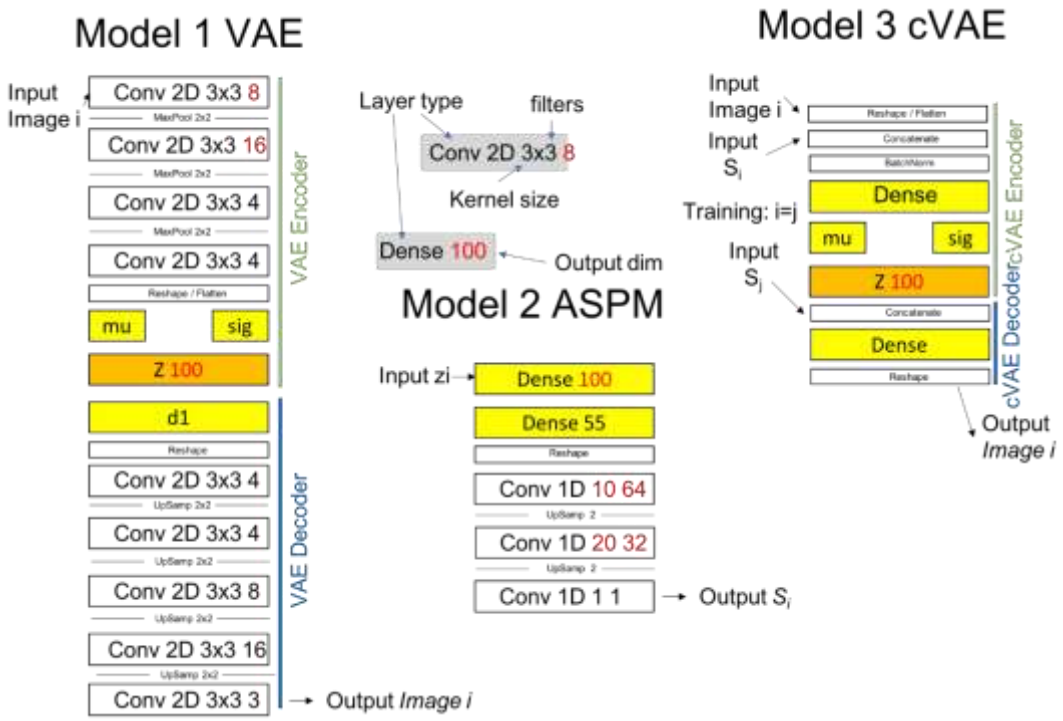

Figure S1: Graph representation of the *neural nets used throughout the manuscript. In split models like the VAE and cVAE the output of the first model is fed as input into the second model. Yellow layers constitute dense layers. The output dimensionality of d1 in model 1 depends on the output shape of the last conv 2d layer of the encoder and is computed during model creation. Red numbers were optimized during hyperparameter optimization. The output dimensions of mu and sig are the same as z.*

The training losses for model 1-3 are plotted in Figures S2-S4. A histogram showing the binary crossentropy per image as a histogram is shown in Figure S5

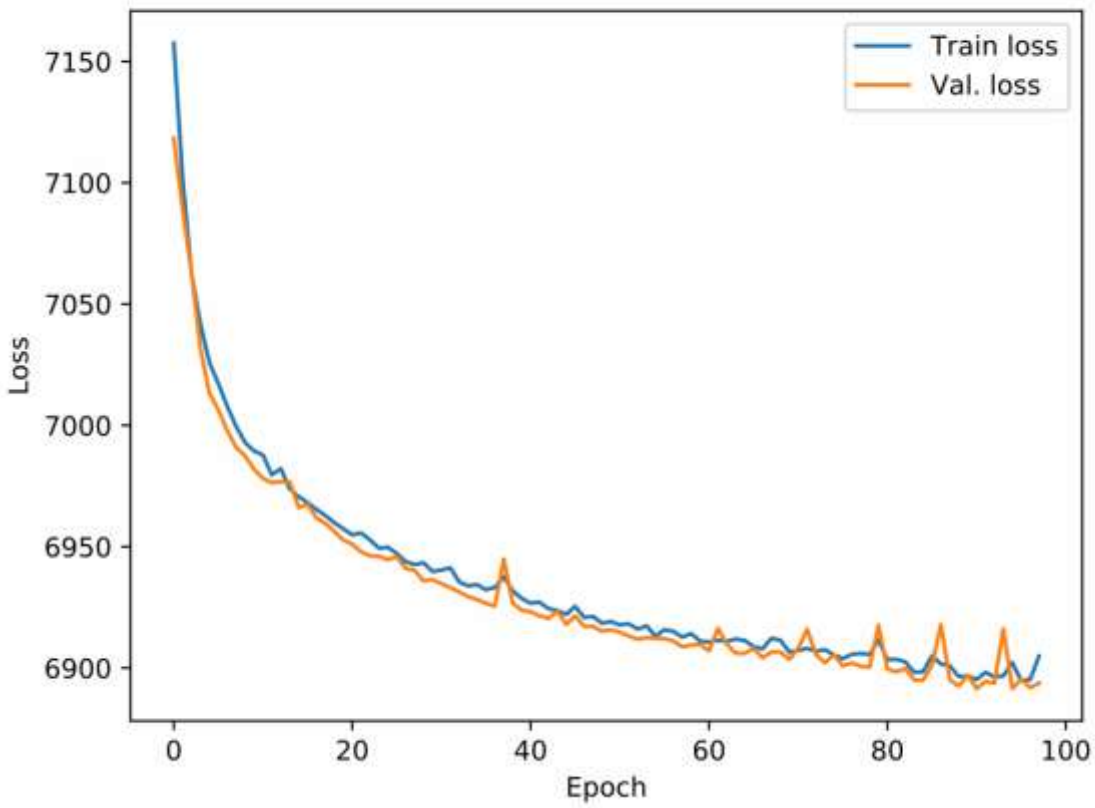

Figure S2: Training loss of model 1 (VAE). Training was stopped after 100 epochs.

Hyperparameter for model 1 one optimized by considering the model with the lowest per pixel loss. The hyperparameter optimization proceed via grid search, as detailed in Table S1. The best and second best model only had a minute difference in per pixel loss but the second best model had fewer trainable parameters (23887 instead of 42059 parameters for a 0.001% increase in loss), leading to our choice of these parameters for the model.

Table S1: Results from hyperparameter optimization. Shown are the hyperparameters and the train loss (no model selection is based on the test loss within this manuscript but only test set results are presented) per pixel.

Filter Layer 1	Filter Layer 2	Latent Space Dimension	Loss per pixel
32	16	50	1.68899
32	16	100	1.68733
32	16	150	1.68677
32	8	50	1.6883
32	8	100	1.69224
32	8	150	1.68991
32	4	50	1.68917
32	4	100	1.6895

32	4	150	1.69105
16	16	50	1.6872
16	16	100	1.68727
16	16	150	1.69131
16	8	50	1.69165
16	8	100	1.68879
16	8	150	1.68826
16	4	50	1.68731
16	4	100	1.69273
16	4	150	1.69012
8	16	50	1.69101
8	16	100	1.68678
8	16	150	1.68683
8	8	50	1.69249
8	8	100	1.6921
8	8	150	1.68761
8	4	50	1.69298
8	4	100	1.69101
8	4	150	1.69375

Table S2: Results from hyperparameter training where a1 denotes the output dimension of the first dense layer and f1 and f2 the filters in the second layer after training for 100 epoch. The two best models were trained for an additional 1000 epochs in which the second-best model (marked yellow) performed slightly better (see main text) and was chosen for the subsequent analysis

a1	f1	f2	loss	r2_loss
100	64	64	0.205	0.651
100	**64**	**32**	**0.197**	**0.628**
100	32	64	0.198	0.639
100	32	32	0.267	0.726
50	64	64	**0.194**	**0.627**
50	64	32	0.200	0.641
50	32	64	0.209	0.650
50	32	32	0.203	0.646

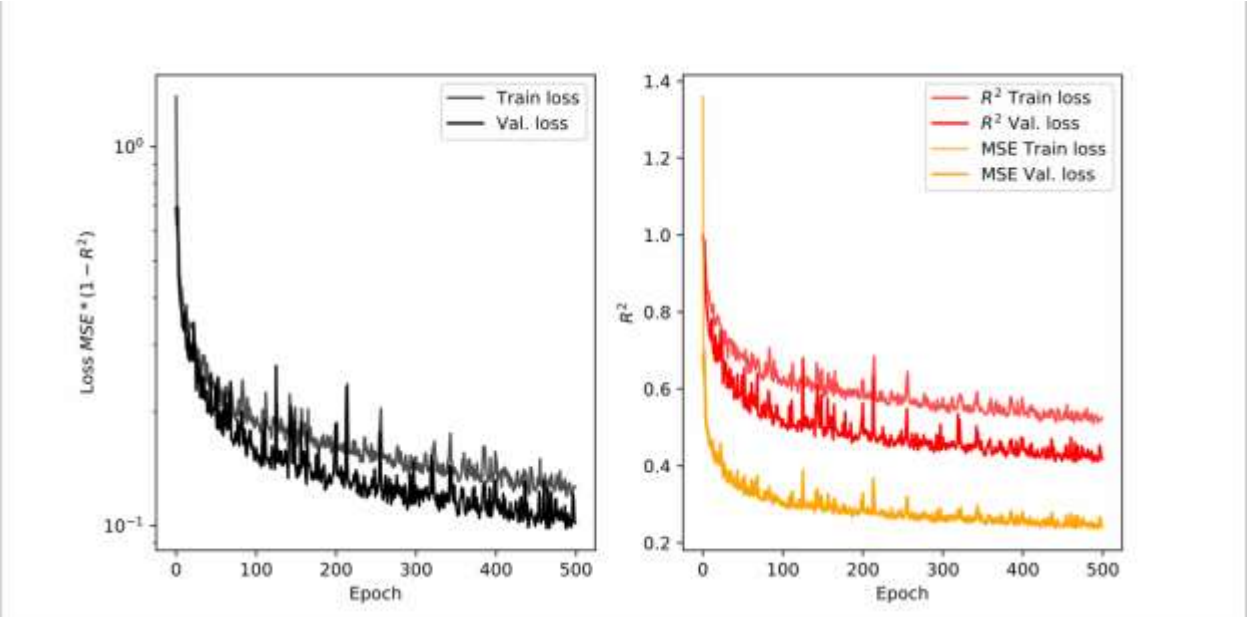

Figure S3: Training loss of model 2 (Spectrum Prediction). Stopped after 500 epochs.

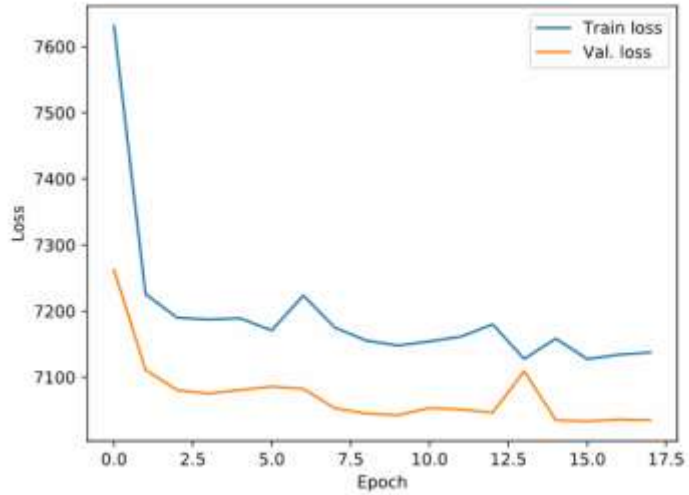

Figure S4: Training loss model 3 (cVAE). Stopped via early stopping.

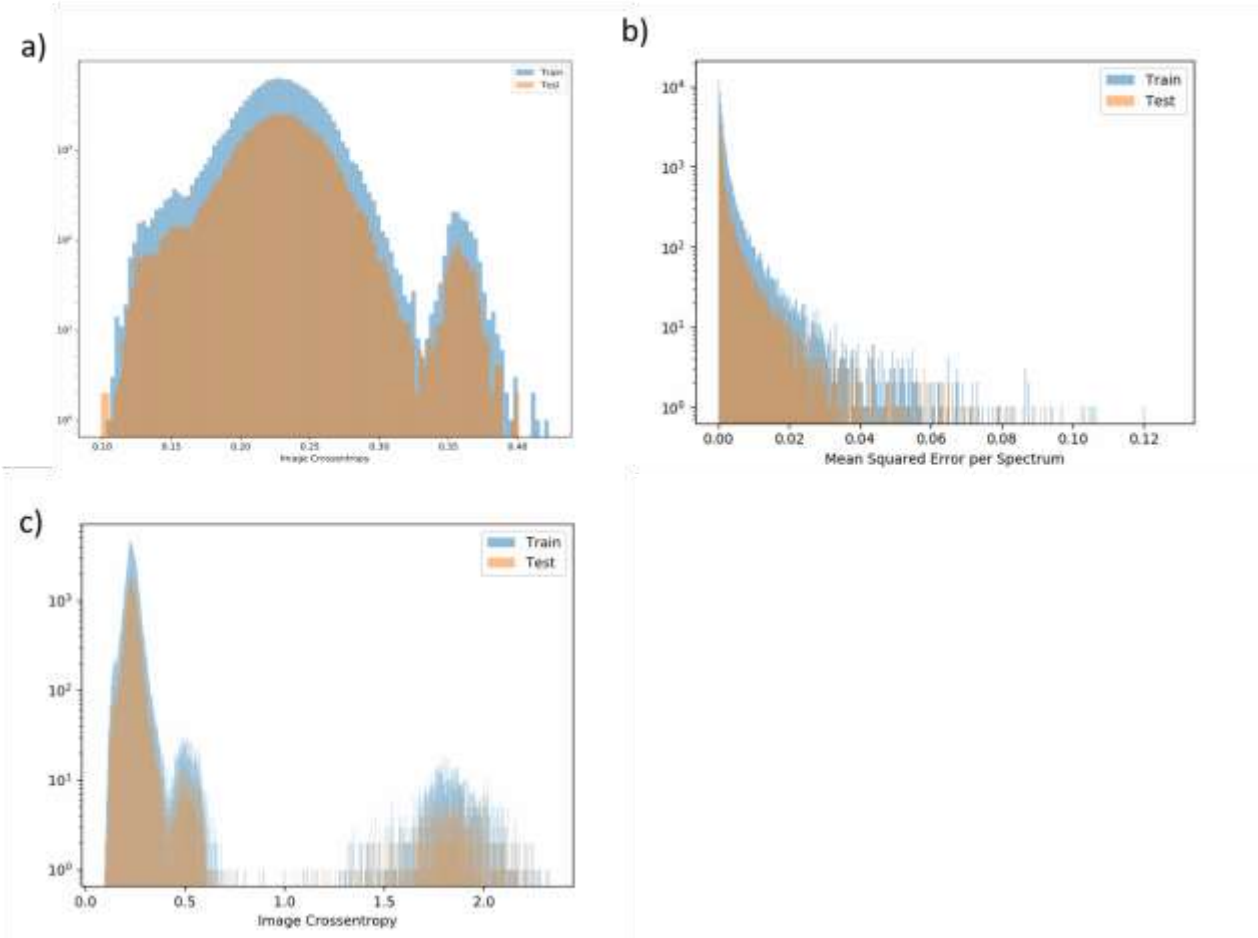

Figure S5: Loss histograms for model 1,2 and 3. a. For the VAE, most images are well reconstructed, and the distribution of errors is quite similar for the test and train sets. b. The absorption spectra are well reconstructed with a more than exponential decay in MSE loss per sample. c. The loss histogram for the cVAE is quite similar to the VAE with the addition of a small fraction of outliers with considerably higher image crossentropy.

Concerning the use of model 3 as a generative model for a conditional absorption spectrum, the resulting material image is, at least visually, relatively insensitive to the choice of latent space coordinate compared to the choice of conditional absorption spectrum, as shown in Figure S6 where 50 randomly chosen samples were used to gather 50 arbitrary latent space coordinates. For each starting latent space coordinate, images were generated according to 11 different conditional spectra (the same conditional spectra as the middle row in Figure 7a). There is relatively minor differences between different columns.

Figure S6: cVAE predictions using randomly chosen starting images and their latent space encoding variable. Each row uses the same conditional spectrum, and each column uses a different image/latent space vector from which the image is generated.

To highlight that the cVAE model also needs a very large dataset to learn from, cVAE predictions from the conditional spectra in Figure 7a are shown after the cVAE model was trained on truncated datasets of 30, 45, 65, and 95 thousand images. The results indicate that at least 65 thousand images are necessary to predict images from spectra, especially for the less frequently occurring lower bandgap materials.

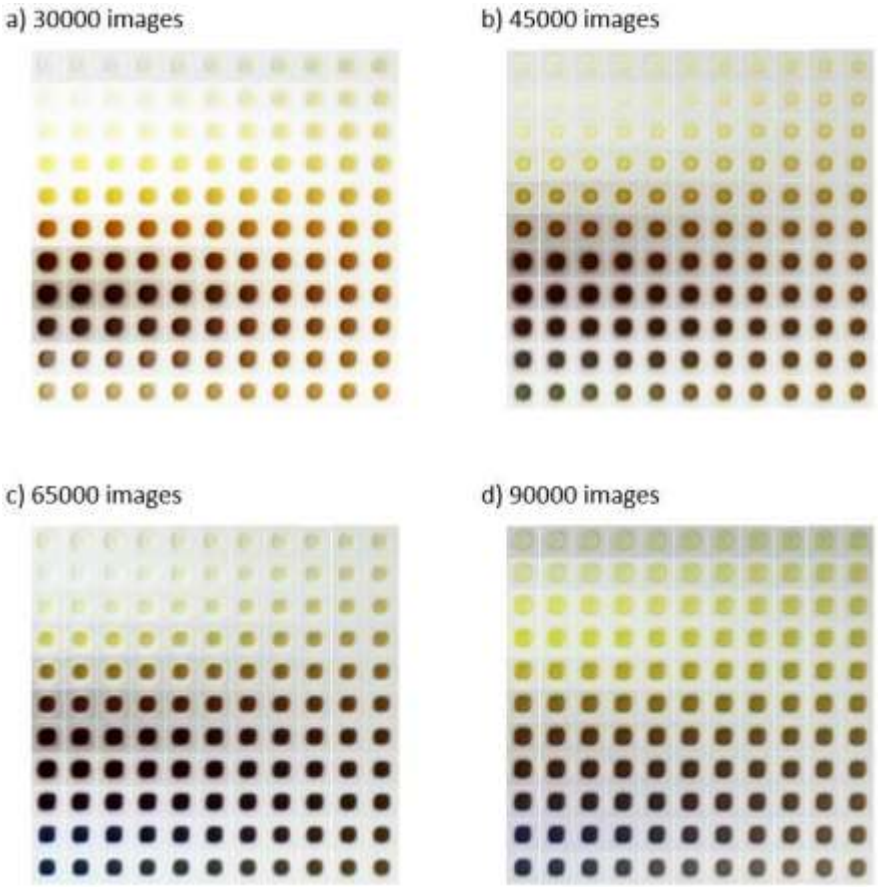

Figure S7: Comparison for the image from spectrum prediction in model 3 run with 30000, 45000, 65000 and 90000 randomly chosen images with the same array of absorption spectra shown in Figure 7a. Visual inspection of the results indicates that minimally about 65000 images and spectra are necessary to train the cVAE to predict the low bandgap materials, and there are visual improvements for 90000 images and more. All images were reconstructed from the same latent space variable that corresponded to an image/spectrum pair whose measured absorption spectrum has lowest L_2 loss from the sigmoidal absorption curve from the spectrum corresponding to the center image.

Additional details on dataset

The materials in the dataset are all metal oxides made by combining 1, 2, 3, or 4 different metal precursors and converting to metal oxides with annealing in air or other O_2-containing atmosphere with annealing temperatures between 300 and 800 °C. Within a given composition space, i.e. combination of metal elements, a grid of composition was synthesized on a library plate. A common design of this grid is the complete quaternary space of 4 metals with 5 at.% intervals. The resulting 1771 compositions (all deposited on a single library plate and annealed together) contain all the individual metal oxides and the mixed metal oxides containing 2, 3 or 4 of the metals. The choices of elements, grid of compositions, and annealing temperatures were varied neither systematically nor randomly to generate the full dataset. These choices were made for pursuing various research projects related to identification of metal oxide photoanodes with a visible band gap. The datasets contains analogous variation of other

synthesis details such as the ink formulation, metals loading, drying procedure, print head, batch of glass substrates, etc., none of which were used in the models.