Motivation
Multi-focus Image Fusion (MIF) is an important technique to reconstruct a fully focused image (FFI) from two or more partially focused images of the same scene. In semiconductor industry, as chip design gets more complicated, capturing FFI gets harder due to topological differences on a wafer (varying heights of the structures). Traditional Computer Vision techniques take multiple source images from the same location at different focal offsets to generate a useful FFI for inspection tools and is time consuming. We propose a deep supervised model for the generation of FFI to solve semiconductor inspection image defocus issue in less time in order to increase productivity throughout these tools.

Data
The data set was collected using multiple customer wafers at multiple focal offsets and at various high topology feature locations, using KLA-Tencor’s proprietary high-resolution semiconductor wafer inspection system.
• Collected 30 images at different focal offsets at 300 different sites on the wafer, for a total of 2000 images.
• From the 300 different sites, 90 sites (1800 images) were used for training, 5 sites (500 images) as dev set, and 5 sites (300 images) for test/validation.
• Each image is a colored image of size 640 x 480 pixels.
• The maximum topological difference at a given site was about 20 microns. Fully Focused images (Ground truth) were generated using KLA-Tencor’s proprietary Software for each high topology site.

Features
For our project, the input images fed to the network are raw color images taken from different sites on a semiconductor wafer. No pre-processing was required on the input images as it can compromise the final image quality. Image quality is of utmost importance for semiconductor inspection systems.

Network Architecture
Our proposed network is Model 1, however we also evaluated Model 2 shown below.

Model 1
The goal was to minimize the number of input images to reconstruct the fully focused image during inference. Thus, our loss function L is as follows:

\[ L(i_f, i_j) = \left| i_f - i_j \right| \]

Where \( i_f \) - Final fused image (from the network) and \( i_j \) = Ground truth Image.

Model 2

Discussion
Our model (Model 1) is designed to input 1’ number of multimodal defocused images and outputs the fully focused image by fusing focused features extracted from individual images. This helps in fusing images with wide range of focus offsets at a given site on a semiconductor wafer. In order to speed up transfer learning, the fusion layer is modularized such that at any given point of time in the future, a new feature fusion methodology can easily be integrated with our network to get a better result. Model 2 is designed for Bi-modal image fusion (background and foreground) and can’t be extended for multi-modal image fusion. This model works on small image patches which has an adverse effect on the memory consumption and computational efficiency. These shortcomings have been addressed in our network (Model 1).

Future
Performance of our network on a site with high density features, degrades slightly. Currently our feature fusion strategy is a simple addition and L1 Norm. Future work includes a better feature fusion methodology to improve the performance on different semiconductor wafer images.

References