Supplementary Material

# Dataset

The Ultra High Carbon Steel micrographs data set consists of 961 micrographs in all. As discussed in the main paper, depending upon which phases appear as primary, there are seven classes. The number of micrographs in each of these classes is show in Table SI.

Table SI: UHCS Dataset description

| Primary phase(s) | Number of images |
| --- | --- |
| Spheroidite | 374 |
| carbide network | 212 |
| Pearlite | 124 |
| pearlite+spheroidite | 107 |
| spheroidite+Widmansttten | 81 |
| pearlite+Widmansttten | 27 |
| martensite/bainite | 36 |

As discussed in the main paper, the original images are of size 645x522, of which a band of size 38 at the bottom contains the metadata generated by the microscope. Removing this, we get images of size 645x484. Following the work by DeCost et al. who introduced this data set, we crop four patches of 224x224 from the center of each image. This gives us a data set of size 3844. We split the data set into approximately 60% train, 20% validation and 20% test. The details of cross validation are discussed in a separate section below.

# Implementation Details

The VGG16 network was originally developed for the task of object detection and localisation as part of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014 and was ranked first and second on these tasks respectively [1]. Since then, the trained network has been transferred for various other data sets and tasks. While transferring the pre-trained network, generally the top output layer is replaced with a new layer. The VGG16 network consists of five convolutional blocks followed by two fully connected layers. Each convolution block contains a few (2 or 3) convolutional layers followed by a MaxPooling layer. The convolutional part of the network acts as feature extractor and the last two fully connected layers perform the end task. We remove the task-specific fully connected layers from the top and add a new fully connected layer of size 4096. Along with training this new layer, we also fine-tune the three convolutional layers immediately below it – that is block5\_conv1 to block5\_conv3. The weights of lower convolutional layers are not updated because the features extracted at lower layers appear to be commonly useful across tasks.

 The Siamese network consists of two copies of the VGG16 obtained as above, joined at the top with a layer that computes L1 distance (sum of dimension-wise absolute differences) followed by an output sigmoid layer. The weights of the two identical networks are tied, in the sense that the gradient is additive across the two.

## Initialization and training

Following advice from previous research [2], we initialize the weights of the new fully connected layer and the output sigmoid layer from a Gaussian distribution with mean 0 and standard deviation 0.01. We train the network using the Adam optimization algorithm with learning rate 0.0001 on minibatches of size 128. We also add regularization to the newly added layers in order to prevent overfitting. The weights of the regularization terms are empirically tuned. We found that a regularization weight of 0.0001 works best. To decide when to stop training, we periodically evaluate the network for the target classification task on a validation set. We stop when the validation set accuracy drops. The network is trained on 1800 minibatches.

## Cross validation

We perform a stratified 4-fold cross validation in order to get statistically significant numbers and report the average test accuracy over four folds. We split the total data set into four folds, ensuring the percentage of samples from each class is preserved (stratified). In each iteration, we use one of the folds as test set and the remaining folds as train set. We further split the train set into 80% train and 20% validation. The validation set is only used to determine when to stop training, as discussed in previous section. The average accuracy over four different test sets is reported in the paper.

# References

1. K. Simonyan and A. Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR.* **abs/1409.1556** (2014).
2. G. Koch, R. Zemel and R. Salakhutdinov: Siamese neural networks for one-shot image recognition. *Proceedings of the 32nd International Conference on Machine Learning* (Lille, 2015).