1. Supplementary Material

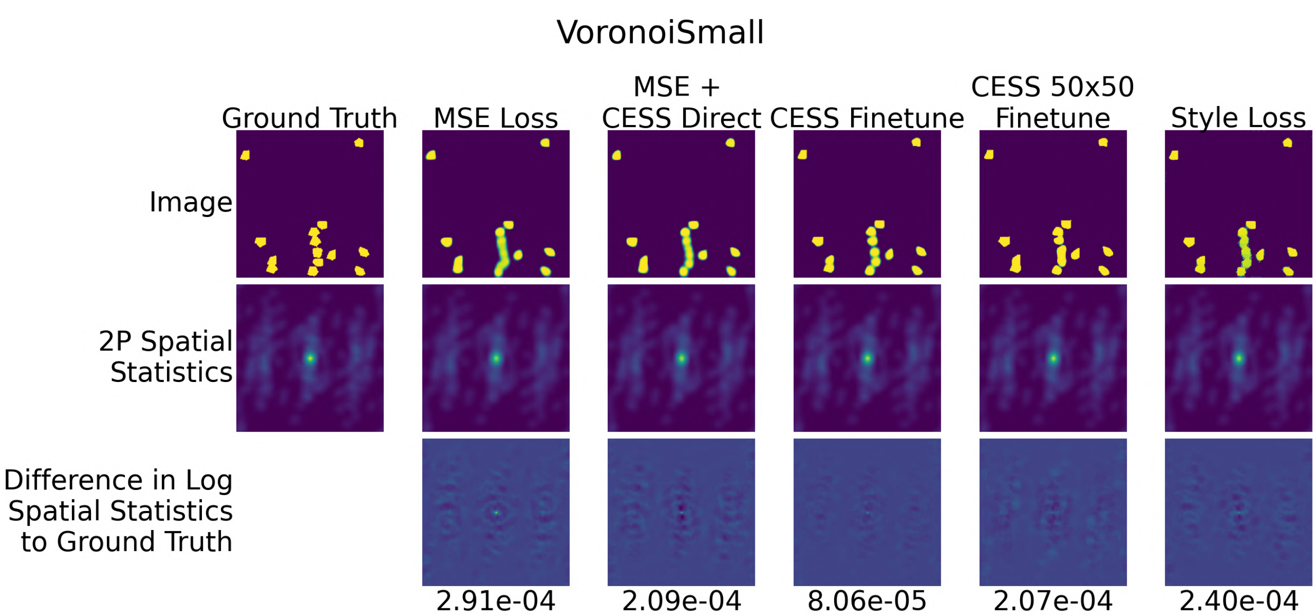
1.1. Optimal hyperparameters from Bayesian sweep

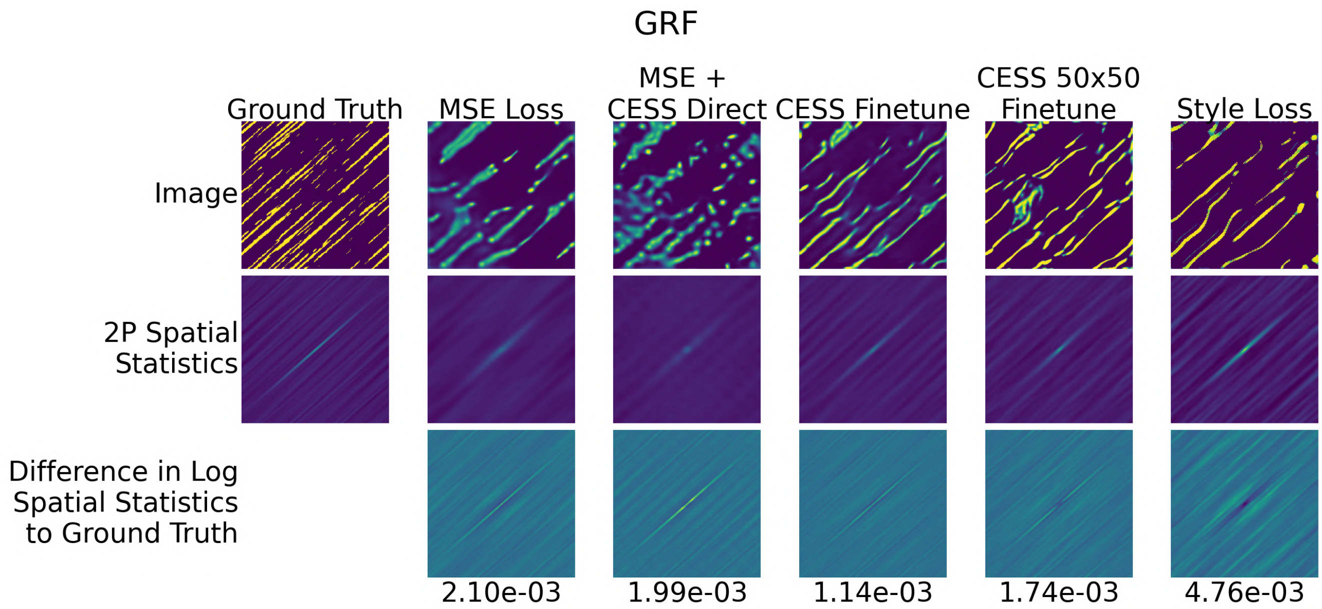
	MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Learning Rate	7.3×10^{-5}	2.5×10^{-4}	5.1×10^{-5}	4.8×10^{-5}	1.2×10^{-4}
Depth	4	3	4	4	4
CESS Loss Weight	0	2.3×10^{3}	6.3×10^{3}	5.2×10^{3}	4.5×10^{-3}
Gradient Clipping	1	0.1	1	1	0.1

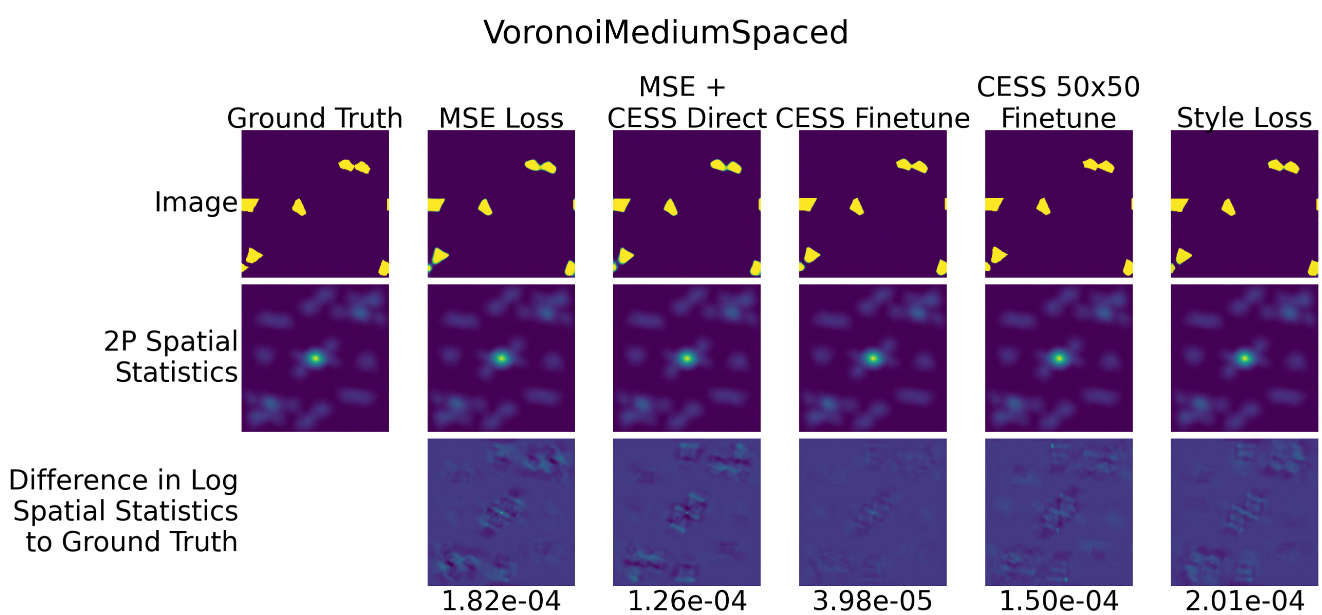
Table 1. Best hyperparameters for each model as determined by a Bayesian sweep. For the MSE + CESS model, the weight of MSE loss was fixed at 1×10^{-3}.

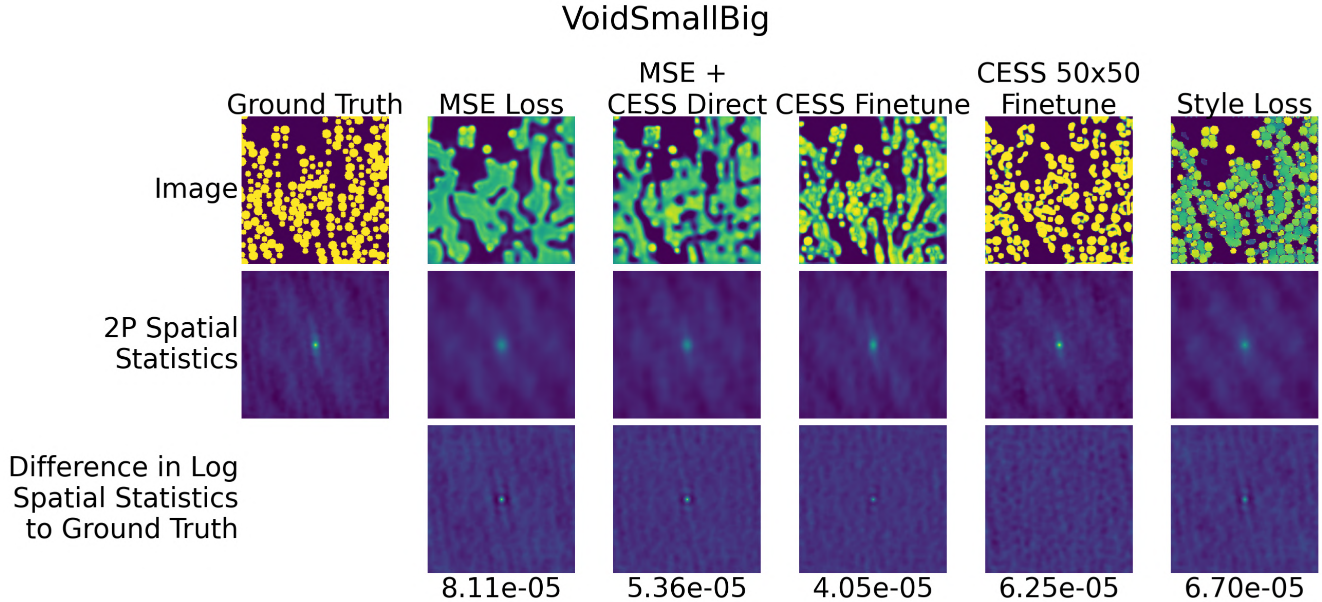
1.2. Reconstructions of a sample from each class with spatial statistics comparison

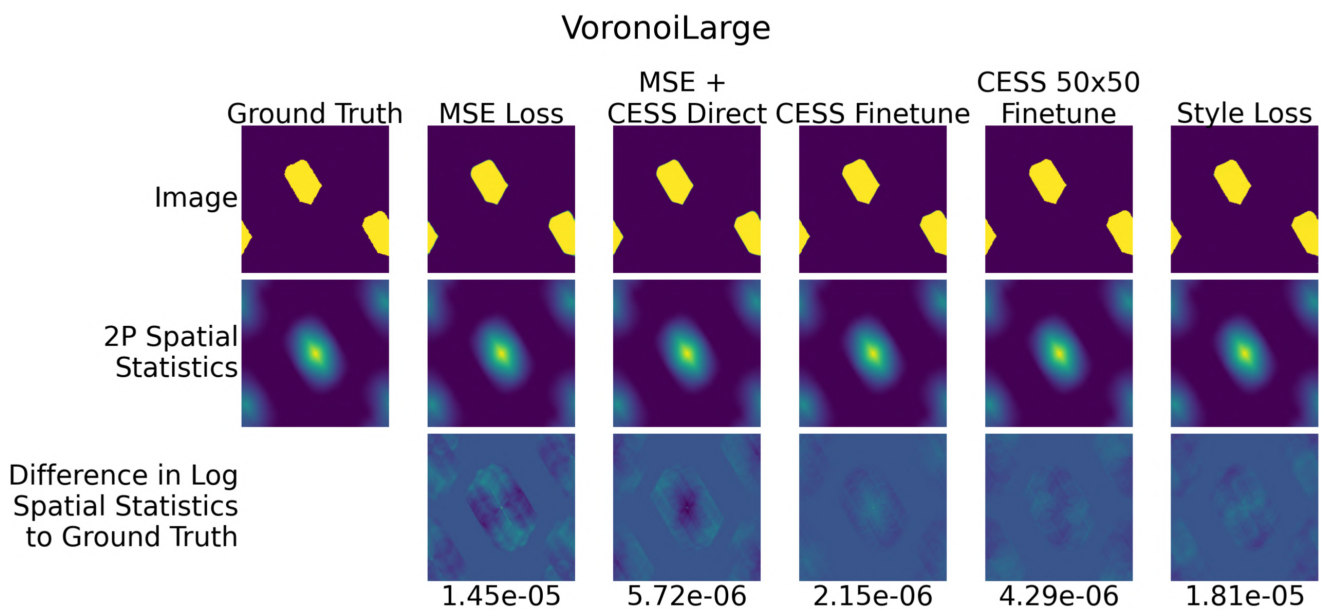
Qualitative analysis can be seen in the next section.

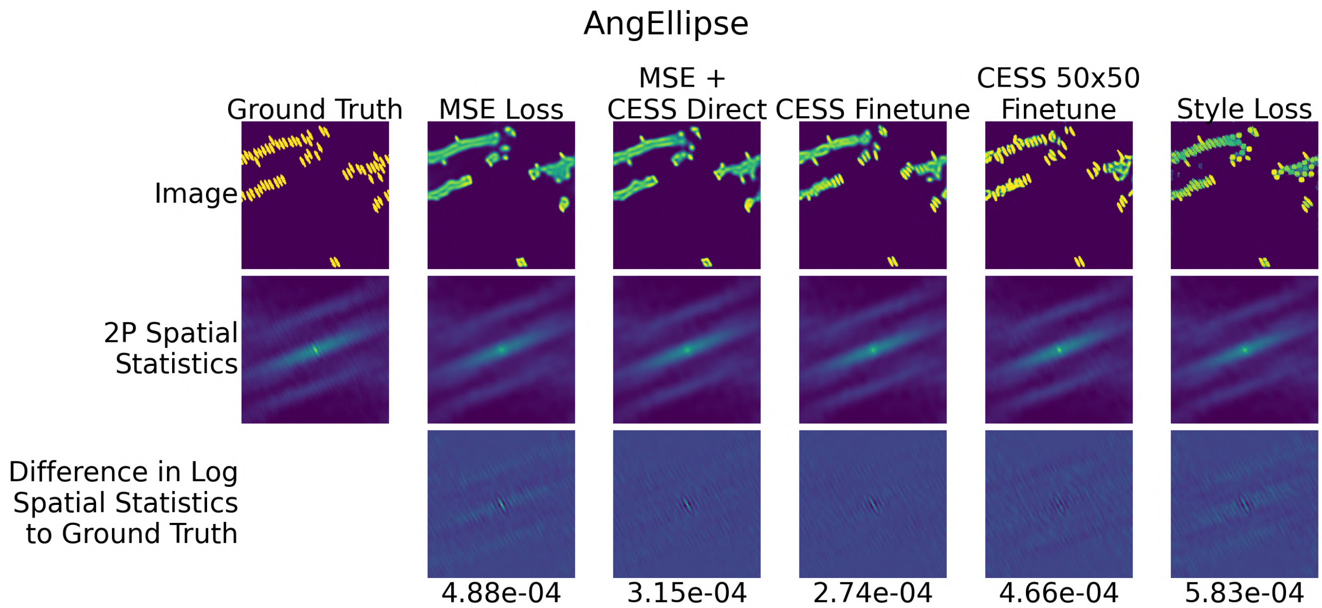


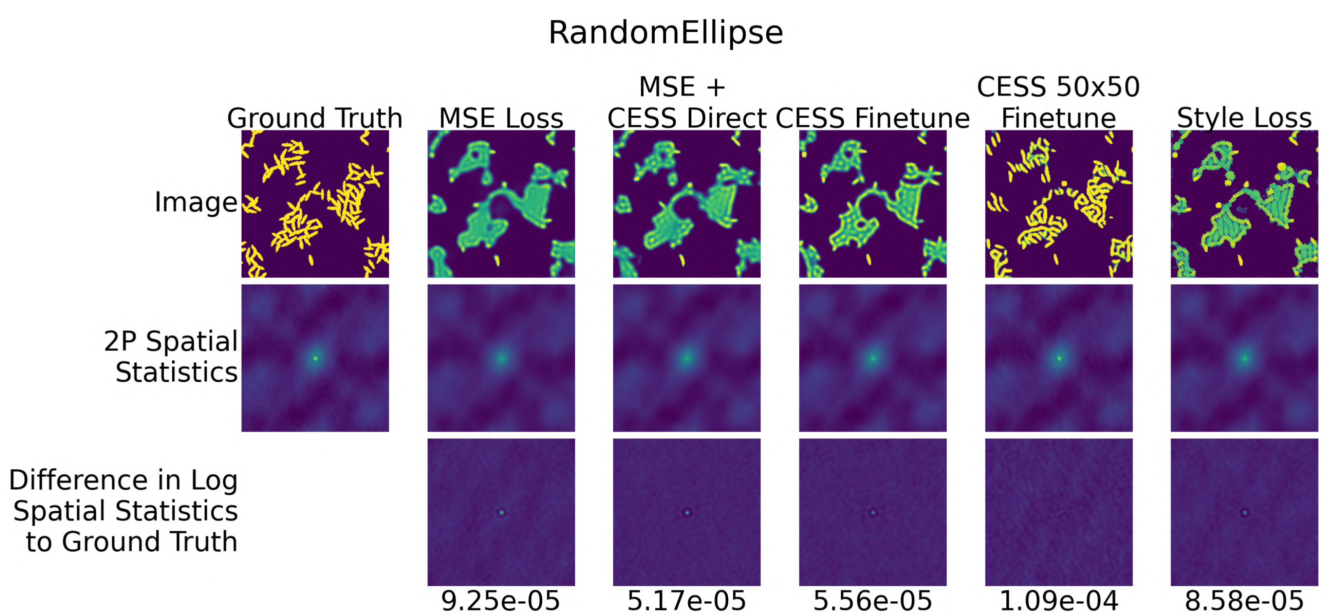

GRF
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
2.10e-03
1.99e-03
1.14e-03
1.74e-03
4.76e-03

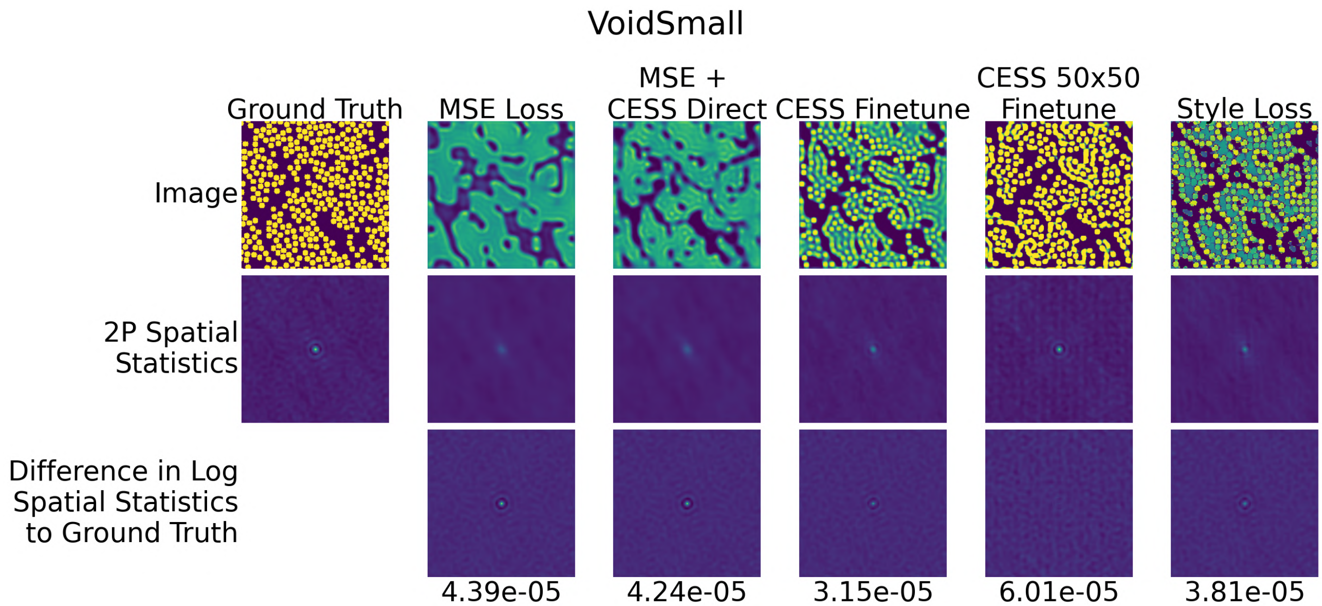

VoronoiMediumSpaced
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
1.82e-04
1.26e-04
3.98e-05
1.50e-04
2.01e-04

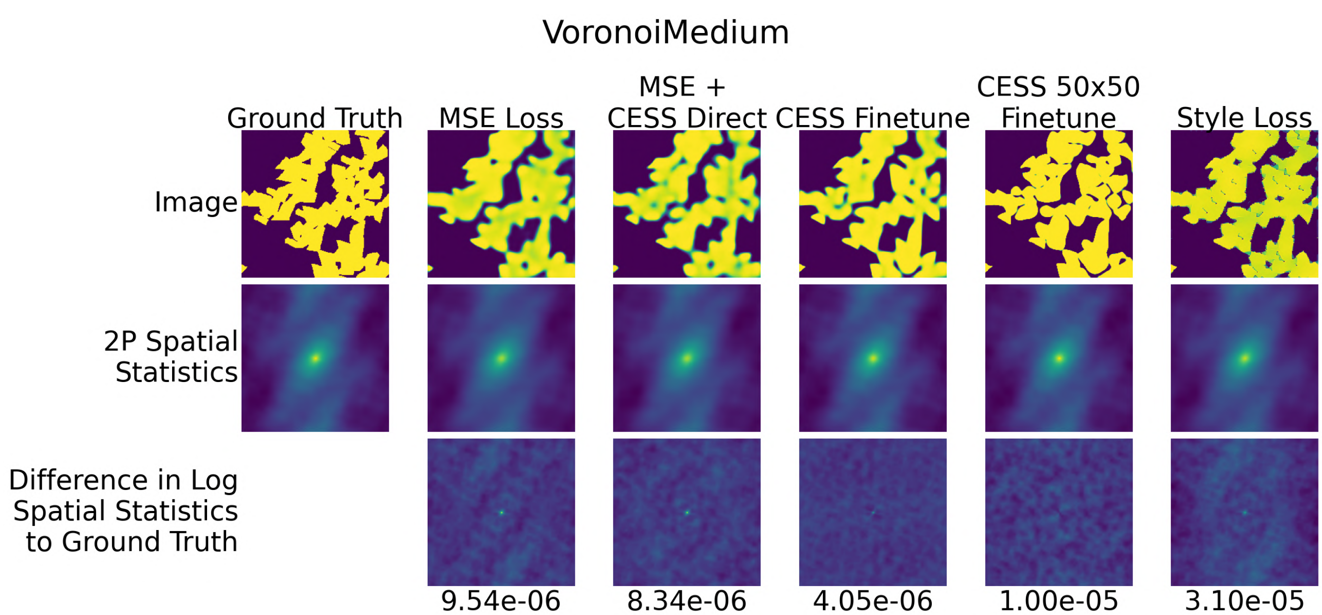

VoidSmallBig
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
8.11e-05
5.36e-05
4.05e-05
6.25e-05
6.70e-05

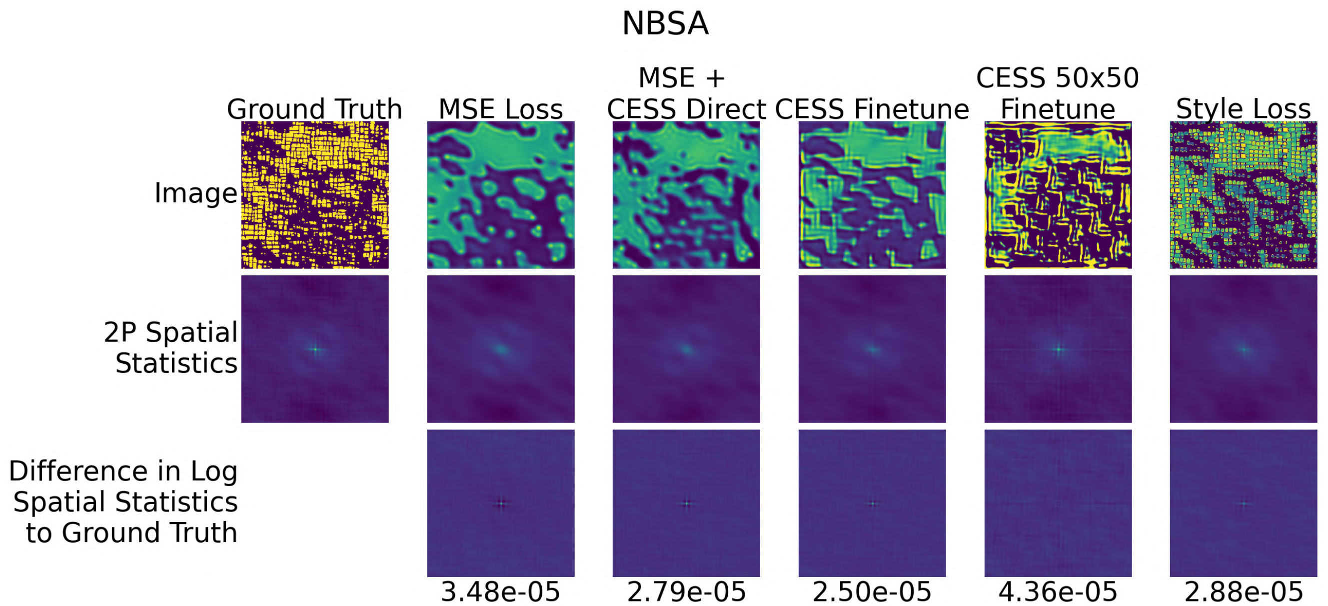

VoronoiLarge
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
1.45e-05
5.72e-06
2.15e-06
4.29e-06
1.81e-05


AngEllipse
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
4.88e-04
3.15e-04
2.74e-04
4.66e-04
5.83e-04


RandomEllipse
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
9.25e-05
5.17e-05
5.56e-05
1.09e-04
8.58e-05

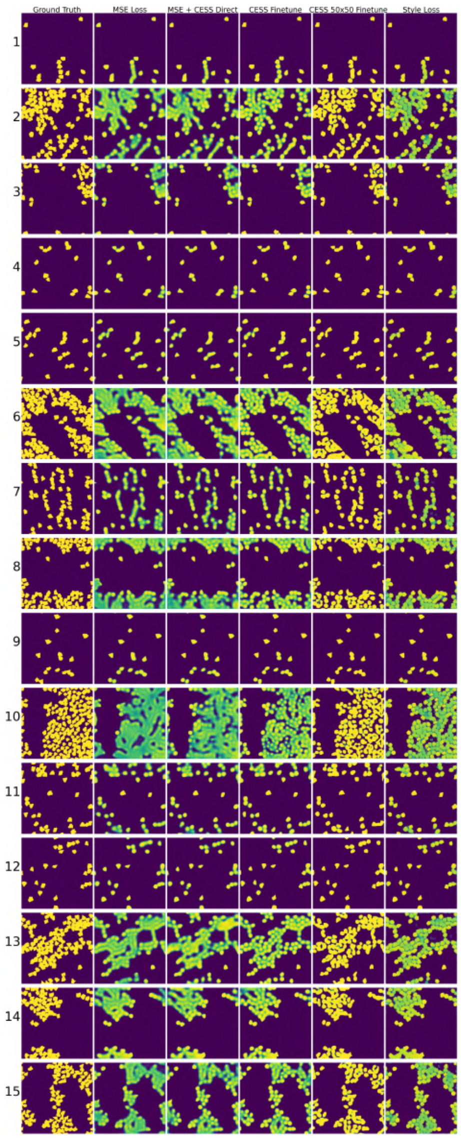

VoidSmall
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
4.39e-05
4.24e-05
3.15e-05
6.01e-05
3.81e-05


VoronoiMedium
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
9.54e-06
8.34e-06
4.05e-06
1.00e-05
3.10e-05


NBSA
Ground Truth
MSE Loss
MSE +
CESS Direct
CESS Finetune
CESS 50x50
Finetune
Style Loss
Image
2P Spatial
Statistics
Difference in Log
Spatial Statistics
to Ground Truth
3.48e-05
2.79e-05
2.50e-05
4.36e-05
2.88e-05

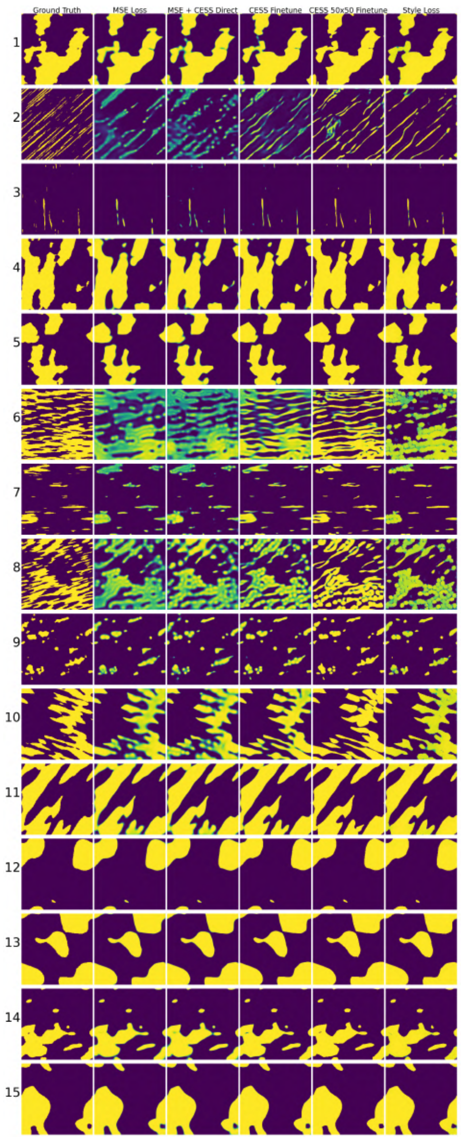
1.3. Qualitative assessment of reconstructions of samples from each class

1.3.1. VoronoiSmall



MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Good reconstructions for relatively sparse microstructures, but grains still often rounded into blurry, circular blobs.	Slightly sharper reconstructions compared to MSE, but similar problems with grain detail.	Similar to MSE + CESS; has sharper reconstructions compared to MSE but still often reconstructs grains as circular blobs rather than polygons	Reconstructions are much sharper, but polygonal nature of grains is not well preserved, with grains in reconstruction essentially completely rounded.	Much better at preserving polygonal nature of grains, but still has issues with rounding their corners.

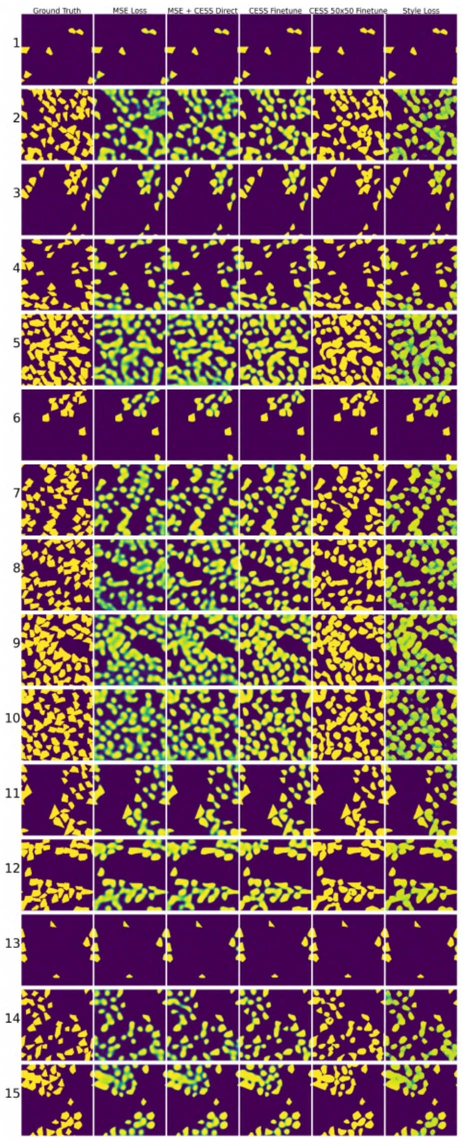
1.3.2. GRF



Microstructures in this class are somewhat diverse, as the classification refers to the generation technique (Gaussian Random Field) and not inherently to any visual similarity.

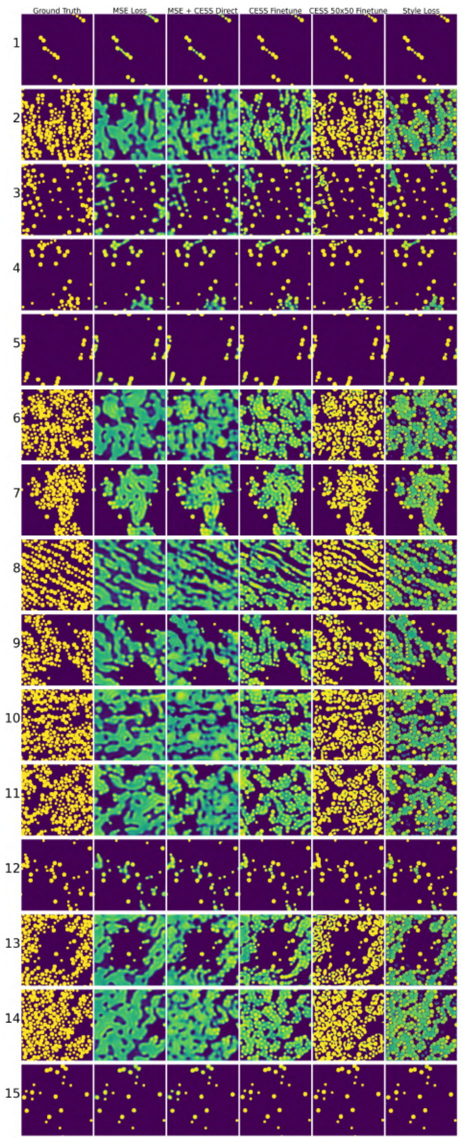
MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Reconstructs images with large, smooth blobs well. Does poorly on images with fine detail, generally reconstructing them blurrily.	Similar to MSE; slightly better resolution of fine detail, although sometimes reconstructs lines as series of blobs (Image 2)	Slightly better reconstructions for some images (Image 2). Sometimes hallucinates existence of grains (Image 8).	Better reconstructions, and image is almost entirely binary. Does not hallucinate grains as strongly (Image 8). Reconstructions sometimes introduce artifacts (Image 2), and line like microstructures do not contain all streaks.	Reconstructions are generally good, but often hallucinate existence of polygonal grains (Images 6, 7, and 8). Line like microstructures still lack some detail in streaks (Image 2).

1.3.3. VoronoiMediumSpaced



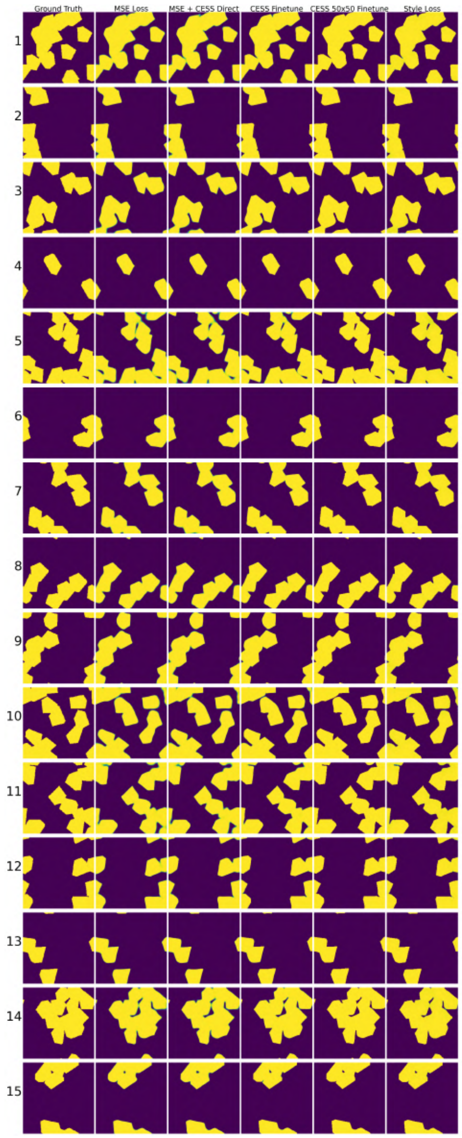
MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Good reconstructions for relatively sparse microstructures. Denser microstructures have blurry reconstructions, and grain detail is lost, often represented instead as blurry circles.	Slightly sharper reconstructions compared to MSE, but similar problems with grain detail.	Similar to MSE + CESS; has sharper reconstructions compared to MSE but still sometimes reconstructs grains as blurry blobs	Reconstructions are much sharper, but polygonal nature of grains is not well preserved, with grains in reconstruction generally rounded.	Slightly better at preserving polygonal nature of grains, but still has issues with rounding their corners.

1.3.4. VoidSmallBig



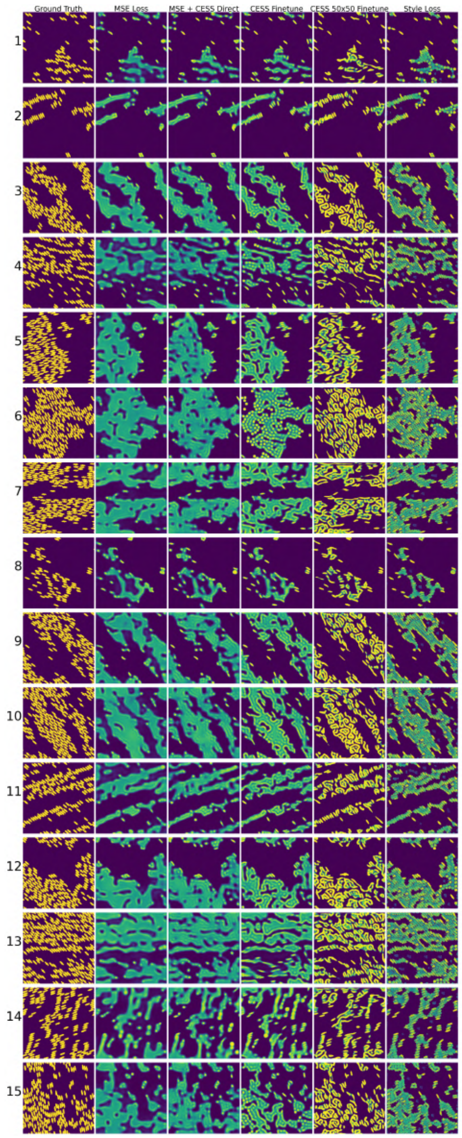
MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Good reconstructions for sparse microstructures. Denser microstructures have blurry reconstructions, with groups of grains blurring into large blobs.	Slightly better grain reconstruction compared to MSE, but similar problems with blurring.	Much better grain resolution compared to MSE and MSE+CESS, but still has some issues with blurring for dense microstructures.	Much sharper reconstructions. Sometimes incorrectly reconstructs groups of grains as streaks (Image 8). Some circular grains also become ellipses (Images 3, 4)	Preserves grain structure the best, but image is not binary

1.3.5. VoronoiLarge



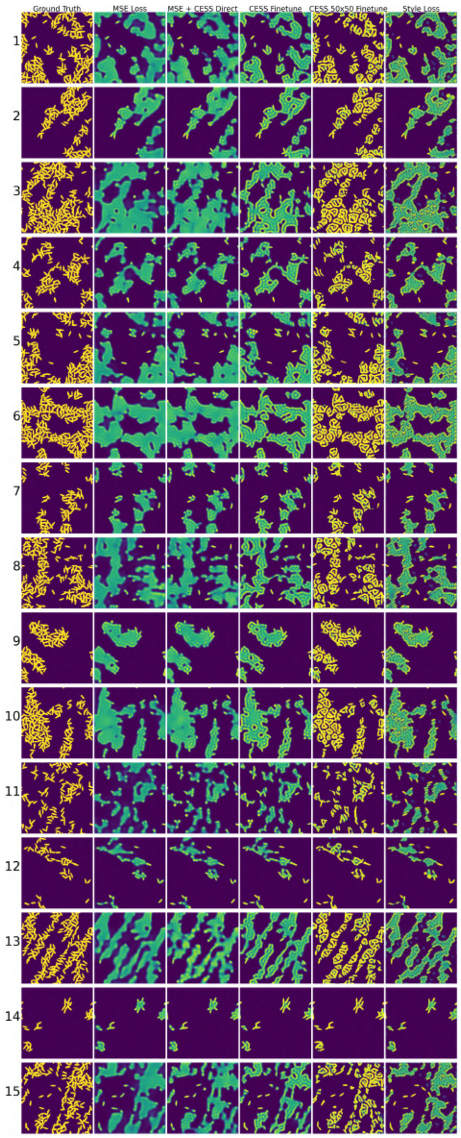
MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Good reconstructions, but slight blurring near edges of grains.	Very similar to MSE, with same blurring issue near grain edges.	Much less blurring near edges compared to MSE and MSE+CESS.	No blurring, but corners of polygonal grains are sometimes rounded (Image 10)	Very similar to CESS finetuned on 50x50 crop; has similar problems with rounding of polygonal grains.

1.3.6. AngEllipse



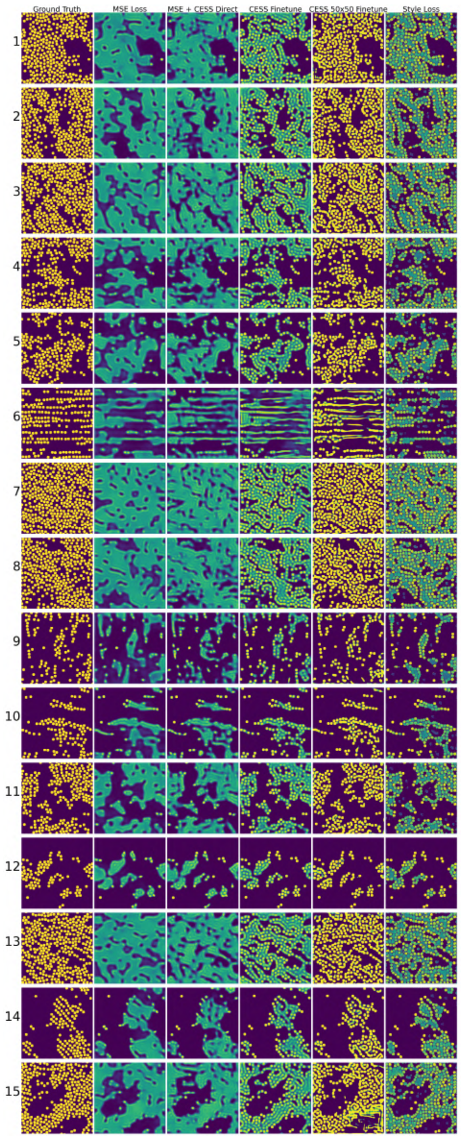
MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Reconstructions are poor, generally blurring large areas of grains into a single blob.	Similar to MSE, although slightly better resolution of grains in blobs, although elliptical nature of grains is not evident.	Slightly clearer resolution of grains, but elliptical nature is still not prominent and sometimes hallucinates existence of circular grains (Image 6)	Much clearer resolution of elliptical grains, but not consistently oriented in one direction	Very clear resolution of elliptical grains, with consistent orientation. However, often hallucinates existence of circular grains (Images 1, 2, 6, 12, 13)

1.3.7. RandomEllipse



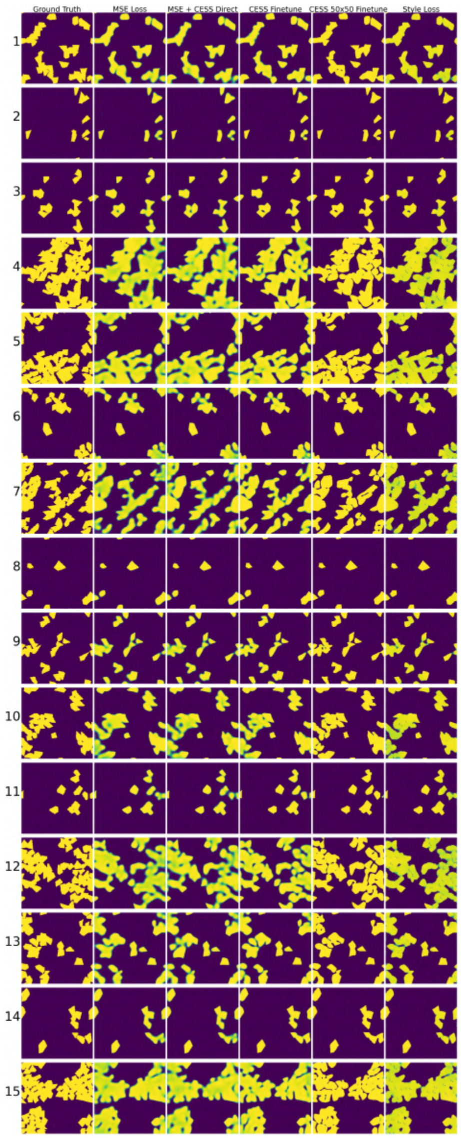
MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Reconstructions are poor, generally blurring large areas of grains into a single blob.	Similar to MSE. Slight resolution of grains in blobs, but often as circular grains and not ellipses.	Slightly clearer resolution of grains, but elliptical nature is still not prominent and sometimes hallucinates existence of circular grains (Images 5, 11)	Much clearer resolution of elliptical grains, but dense sections of grains sometimes merge.	Unclear resolution of elliptical grains, instead producing more polygonal like grain structures.

1.3.8. VoidSmall



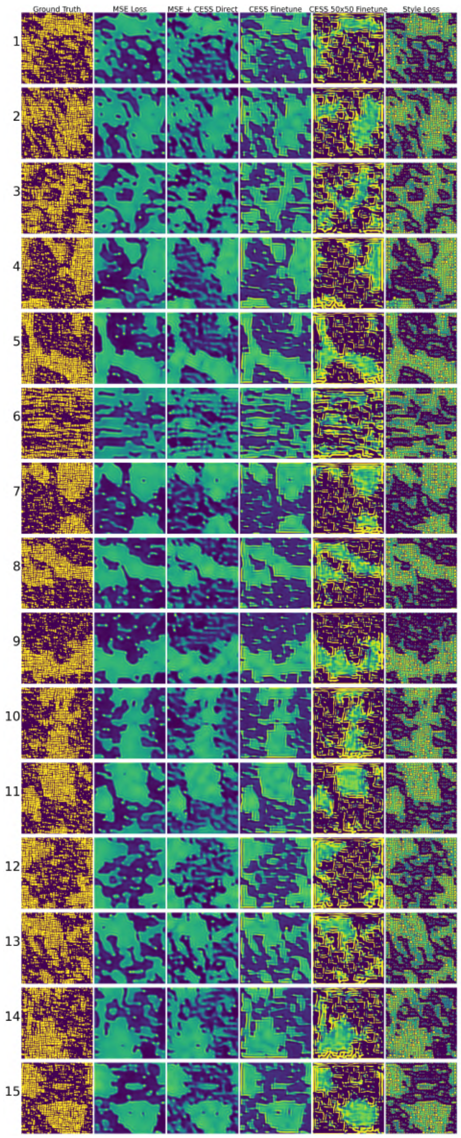
MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Generally unrealistic reconstructions; closely packed grains are reconstructed as a blurry blob.	Generally unrealistic, but some grain structure is now visible in the blobs.	Somewhat realistic. Grains are mostly resolved in blobs, although some blurriness is still present. Lines of grains are sometimes incorrectly merged in the reconstruction into linear streaks (Image 6).	Grains are essentially completely resolved and image is almost entirely binary. Lines of grains are still merged into linear streaks (Image 6)	Reconstructions completely preserve grain structure, but image is not binary and has variable weight for each grain.

1.3.9. VoronoiMedium



MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Sparse regions are reconstructed well, but dense regions of grains are merged into blurry blobs	Slightly better resolution in blurry blob regions, but still faces essentially the same issues as MSE.	Slightly better resolution than MSE+CESS, but polygonal nature of grains is sometimes not present in reconstructions, which have rounded corners	Image is much sharper but faces the same problem as CESS Finetune in preserving polygonal nature of grains.	Best preserves the polygonal structure of grains.

1.3.10. NBSA



MSE	MSE + CESS	CESS Finetune	CESS 50x50 Finetune	Style Loss
Poor reconstructions; grain structure not present, with regions being represented as blurry blobs.	Slightly better resolution in blurry regions, but grains resolved as circular grains rather than square-like grains.	Slight resolution of grains at corners as square-like, but for the most part grains are merged into lines.	Poor reconstructions, with few grains preserving square-like nature and instead mostly merging into lines.	Preserves the square like nature of grains very well.

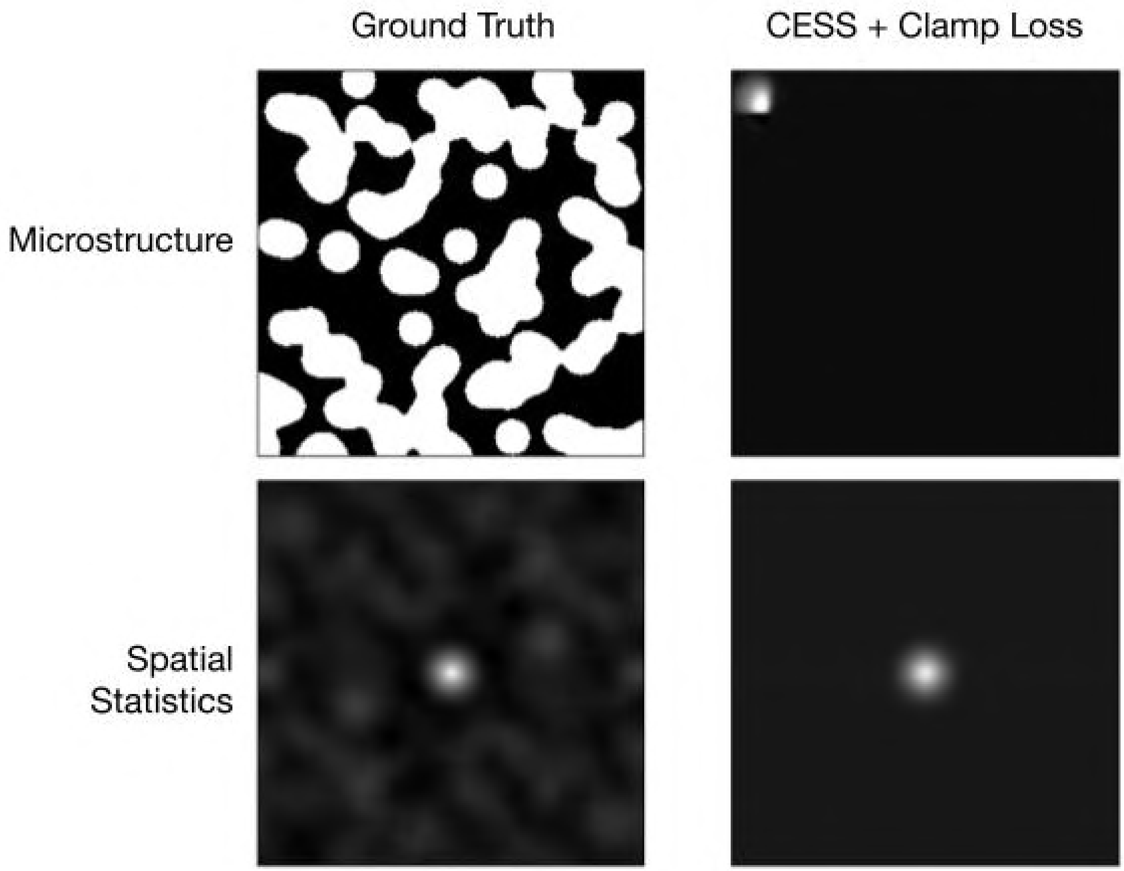


Figure 1. Results from training with only CESS and clamp loss.

1.4. Hypothetical explanation for poor training behaviour of CESS

We first visualized an approximate loss landscape. Two synthetic circle microstructures with 5 circles were first generated; a smooth path was then generated between them by interpolating the circle locations. This roughly represents the shortest path between these microstructures in the "ground truth" latent space. The loss between the first microstructure and evenly spaced interpolants from the path were then computed and plotted (Fig 2).

The generated landscapes led us to hypothesize that the CESS loss has many local minima when far from the original microstructure. We additionally hypothesize that this is because there are many microstructures that satisfy the binary constraint and have similar spatial statistics (Fig. 3).

Three main variants of the loss and training procedure were tried to solve this issue, with some justification:

- Training a model with a linear combination of MSE on the microstructure and CESS; MSE dominates when far from the original microstructure, so this has the effect of smoothing out the loss in those regions.
- Finetuning a model trained with MSE on CESS; this would hopefully mean that reconstructions are already close to the original, making CESS more stable
- Finetuning a model trained with MSE on a variant of CESS that only looked at the center 50x50 crop of spatial statistics; this would ideally make the loss focus more on local properties (e.g. grain size, shape) and be less sensitive to global changes.

Models were also trained with MSE and with a mixture of MSE and Style Loss to be used as baselines.

The approximate impact of these methods to the loss landscape is visualized in Fig. 4; overall, the results show that these methods would theoretically remove local extrema and generally smooth the loss landscape.

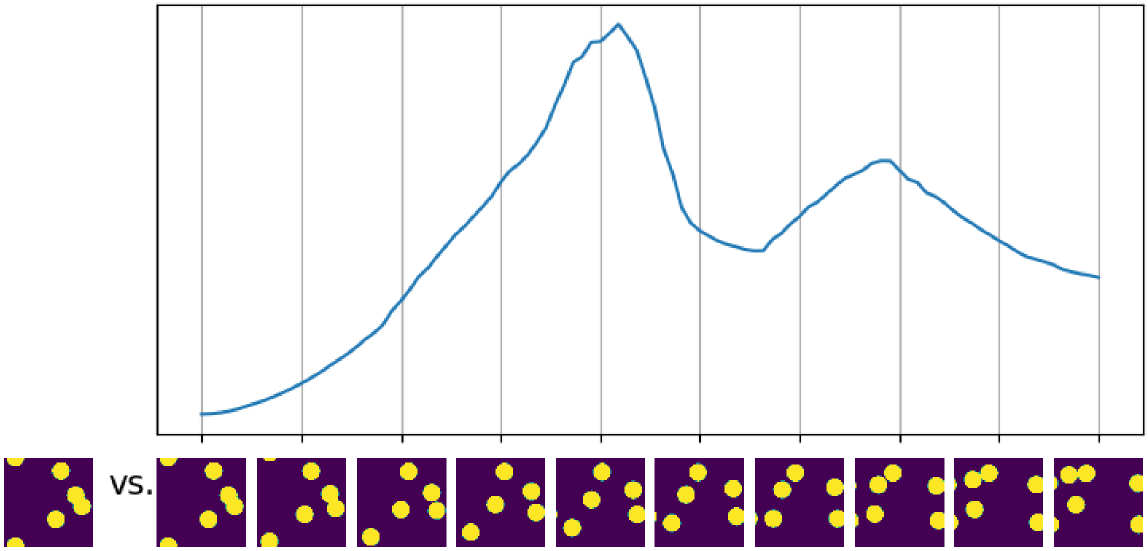


Figure 2. Approximate loss landscape of CESS. Loss is computed vs. the image on the far left of the x-axis.

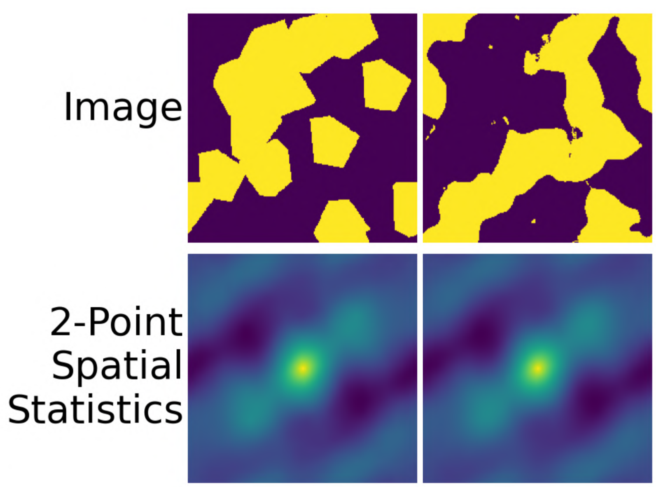


Figure 3. Two different binary images with highly similar spatial statistics.

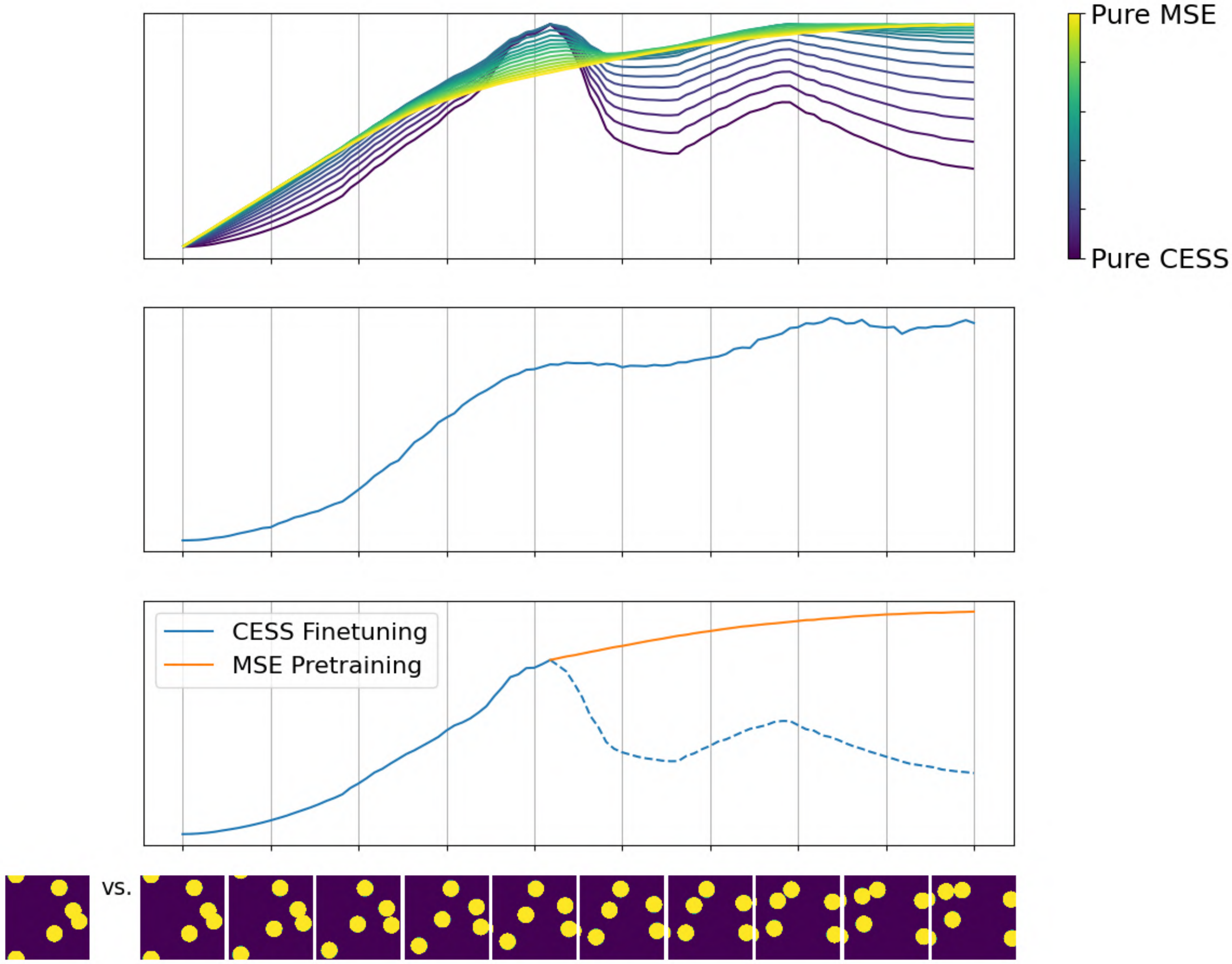


Figure 4. Top: Loss landscape with different weights of MSE. Middle: Loss landscape with CESS on a 50x50 crop. Bottom: Idealized loss landscape using finetuning.

1.5. Network architecture

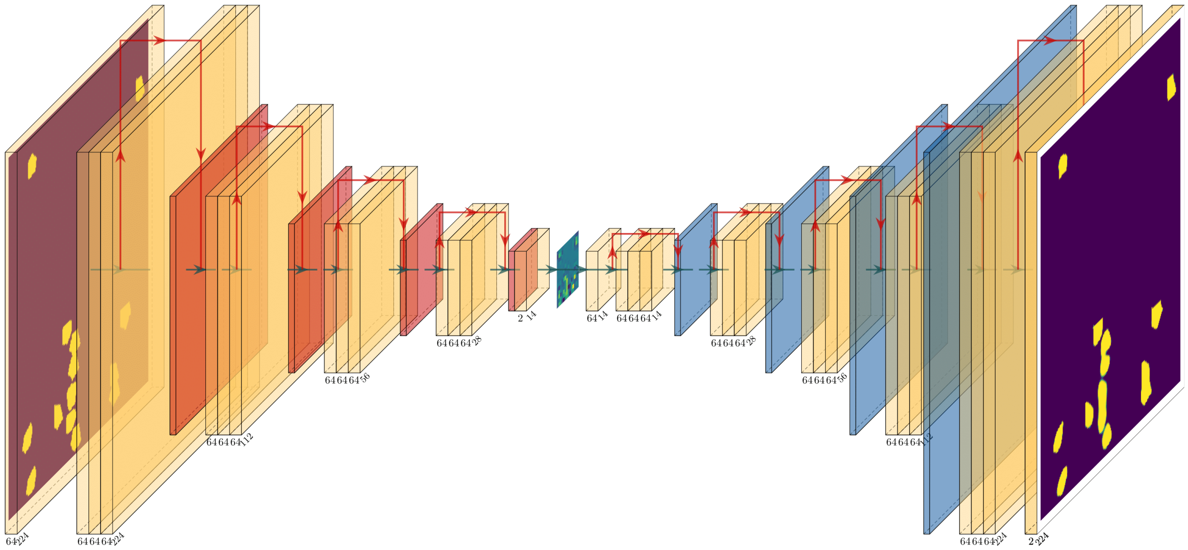


Figure 5. Architecture diagram for model with depth 1. Yellow layers are convolutions, red layers are pooling, and blue layers are upsampling. A ReLU activation is assumed to follow every convolution layer.