TOMOGRAPHY IMAGE QUALITY ASSESSMENT

Non confidential
(Online available)

Tutor: Hammami Omar (ENSTA)
Arnaud Demortière (UPJV)

Student: Zhang kai
08/31/2020

Abstract

Images perception plays a fundamental role in the tomography-based approaches for microstructure characterization and has a profound impact on all subsequent image processing steps (segmentation and analysis). However, the enhancement of image perception frequently involves the observer-dependence, which translate into user-to-user dispersion and uncertainties in the calculated parameters. This work presents an objective quantitative method, which utilizes convolutional neural networks, for tomographic image quality assessment. Compared to most existing data-driven methods, our method requires less annotations and is more appropriate for tomographic images applications. Different metrics were employed to evaluate the correlation of our predicted scores with the subjective human opinion as well as the segmentation accuracy. The evaluation results from this work demonstrate that our method can be a direct tool that guides the enhancement process and conduct to a reliable segmentation results in respect to the subjective human opinion. As a result, the image processing can turn into a very robust, observer-independent process.

Keywords:

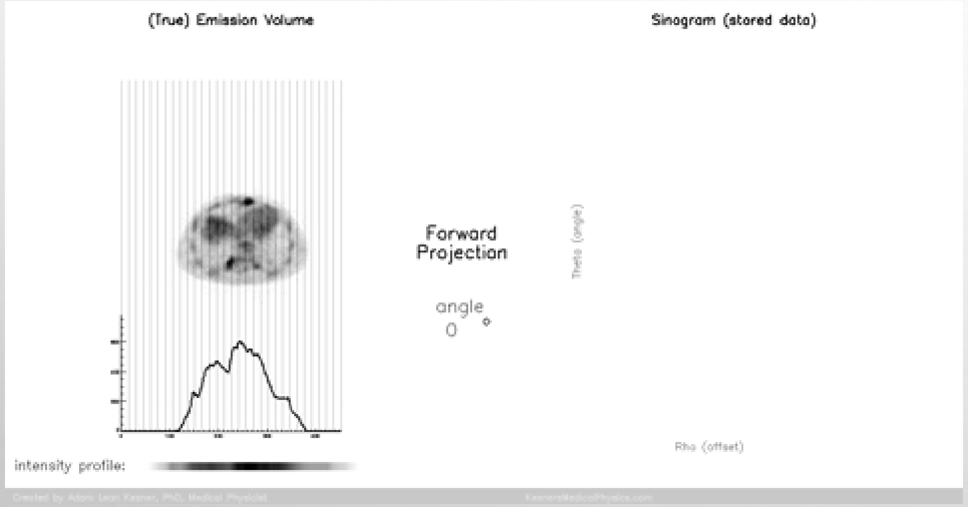
Image quality assessment, Semantic segmentation, Tomography images

Contents

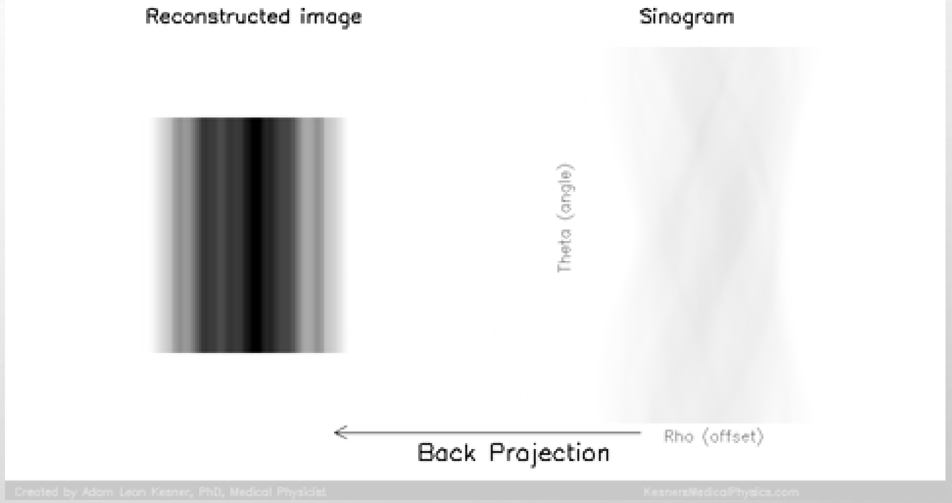
1. INTRODUCTION
2. METHOD
3. RESULTS
4. CONCLUSION

Tomography images: generation workflow

Data Acquisition (2D projections)



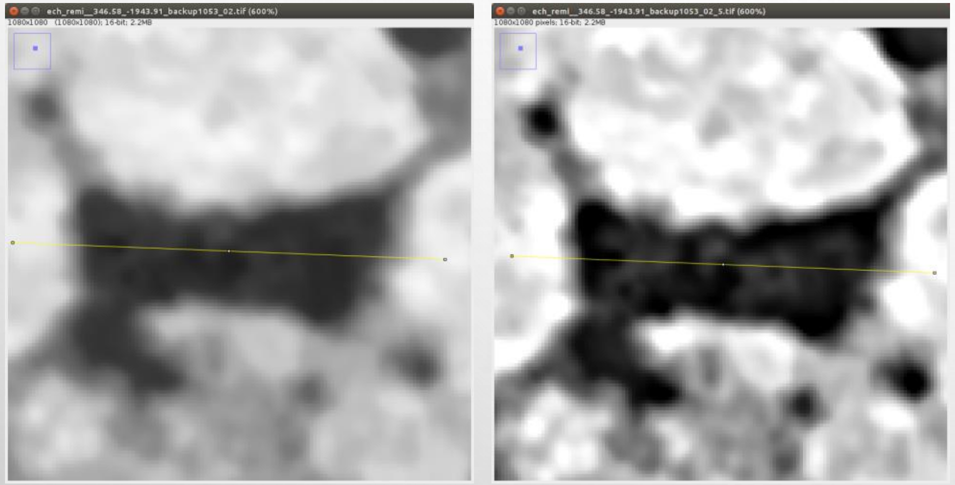
Data Reconstruction (3D volume)



Patrick Pietsch and Vanessa Wood*, Annual Review of Materials Research 2017 47:1, 451-479*

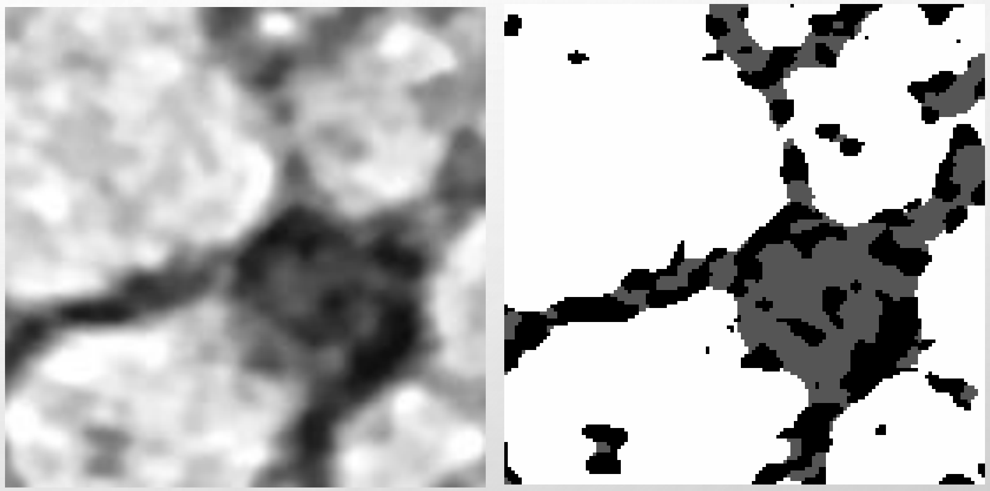
Tomography images: analysis workflow

Image Preprocessing



Intensity adjustment
Distortion removal

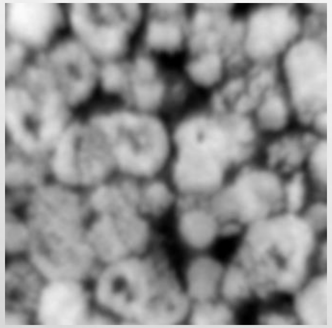
Data Segmentation



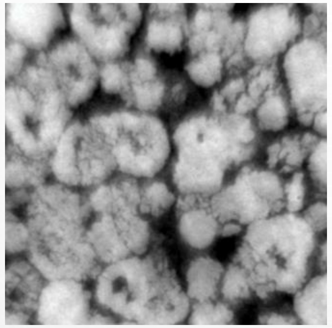
Quantitative analysis

Assess image quality and select the best image for segmentation

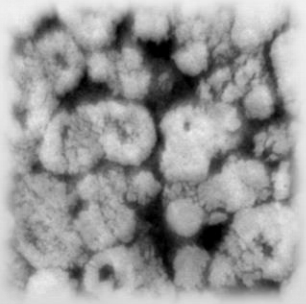
Image quality assessment



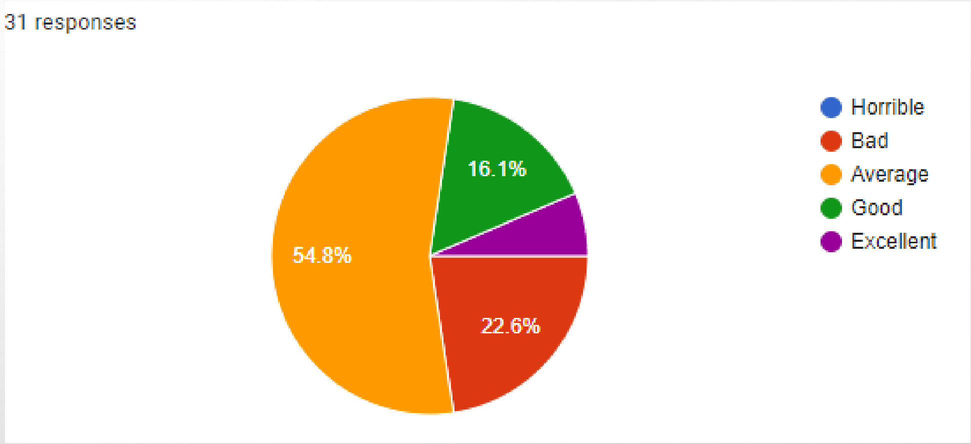
Distorted image



Reference image



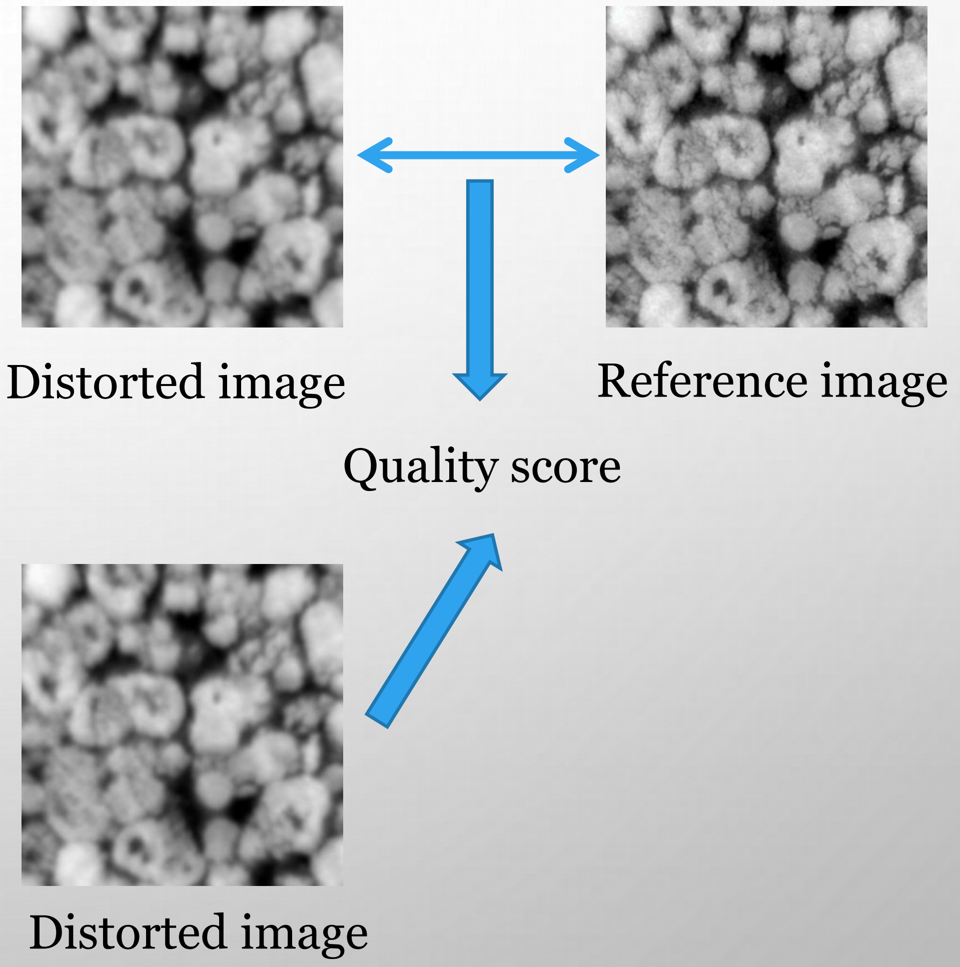
Hallucinated scene



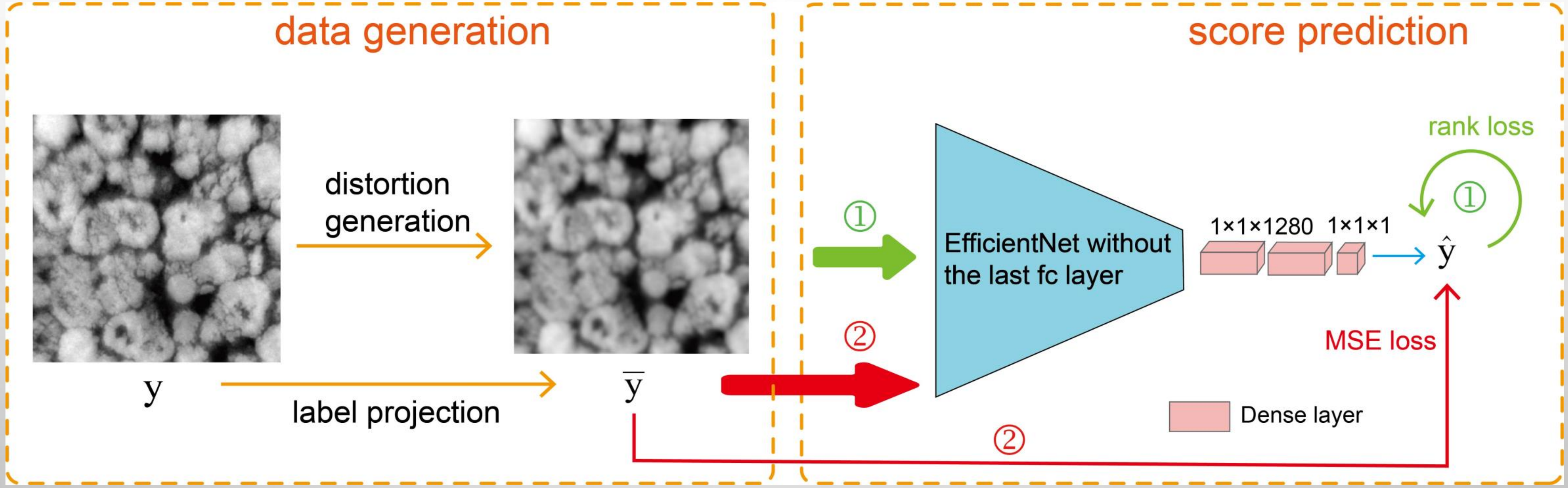
Trust-worthy quantitative metric

Image quality assessment

- **FR-IQA**: Full reference image quality assessment.
- **RR-IQA**: Reduced reference image quality assessment.
- **NR-IQA**: No reference image quality assessment.

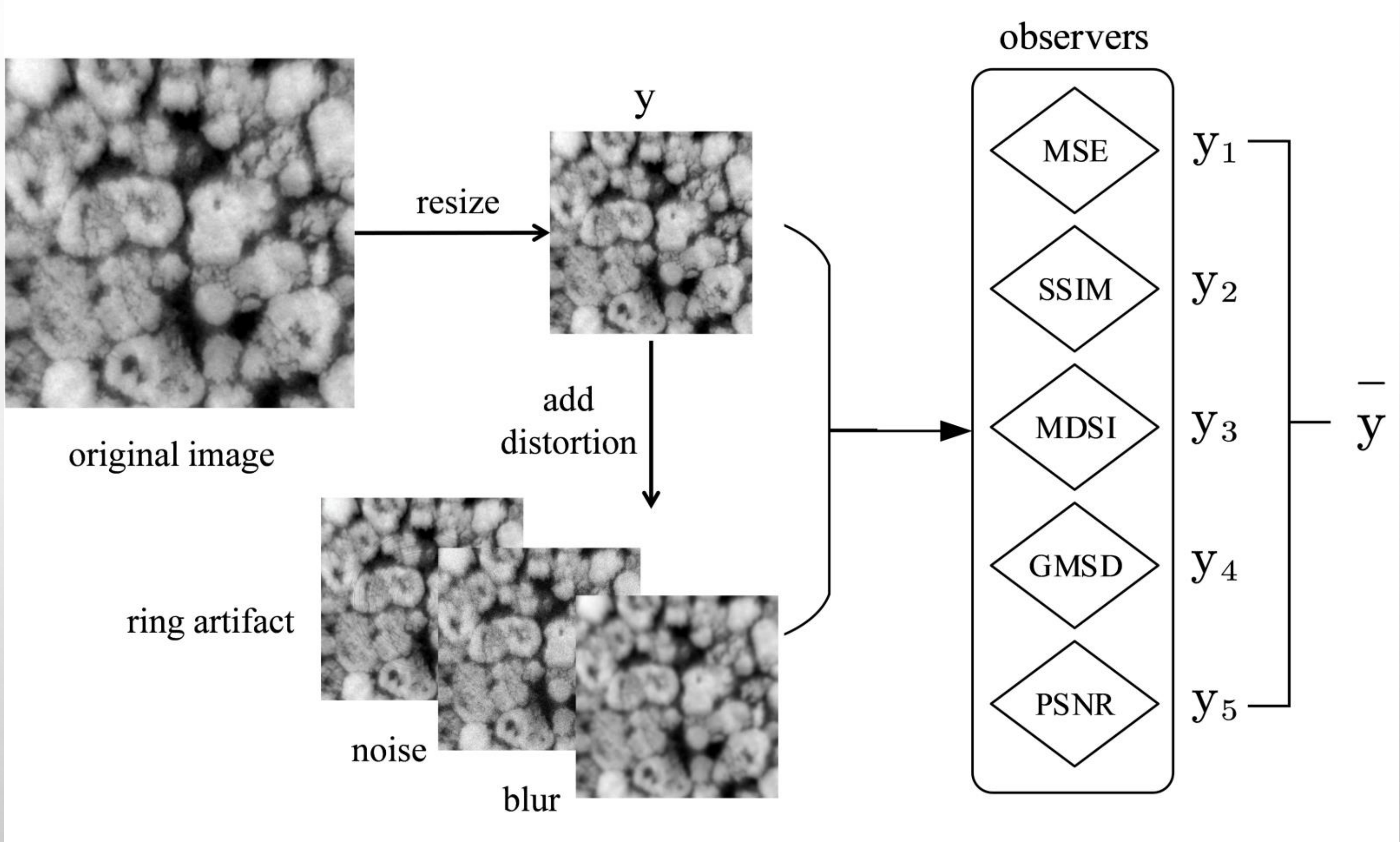


Method



Pipeline of tomography image quality assessment method

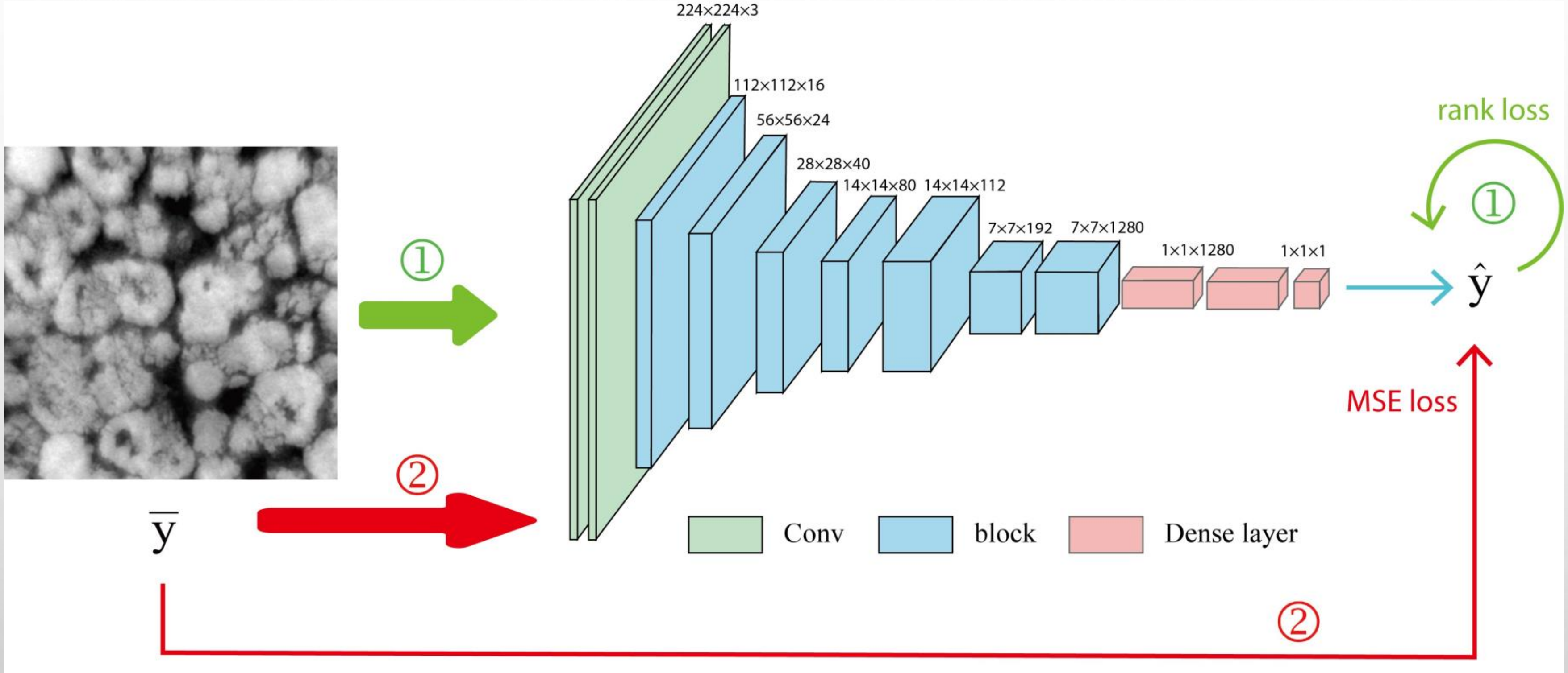
Data generation



Detailed structure of data generation.

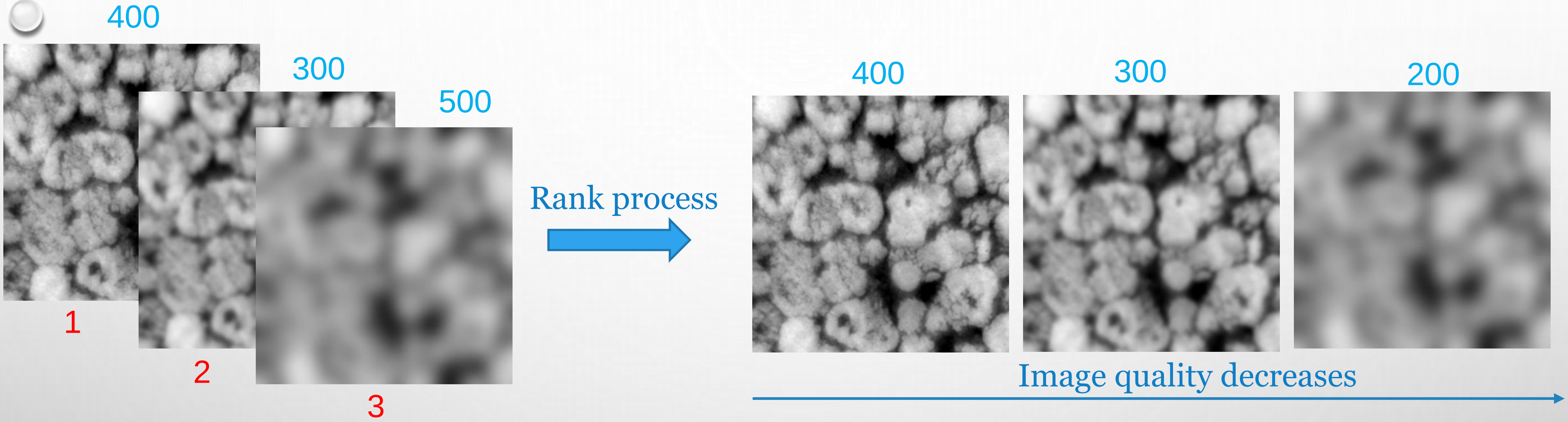
- Tackle the insufficient dataset problem
- Reduce the cost for human annotations

Score prediction



Structure of score prediction module. $\bar{y}$ is the human annotated score or the generated score by label projection. $\hat{y}$ is the score predicted by the network.

Score prediction - ranking

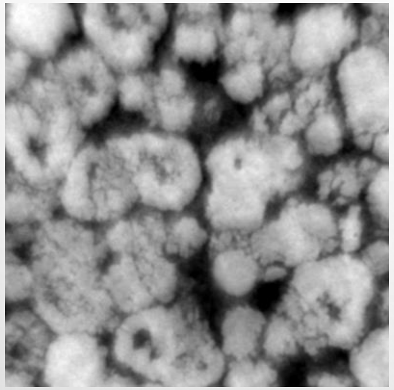


① Rank loss

$$L(\hat{y}_i, \hat{y}_j) = \max(0, m + \hat{y}_j - \hat{y}_i)$$

Higher score means better quality

Score prediction - regression



300 → 2.47

② MSE loss

$$L(\hat{y}, y) = (\hat{y} - y)^2$$

Regress the ranking information to a specific score, in the range of 1 to 5 (terrible, bad, average, good, excellent)

Evaluation method - two metrics

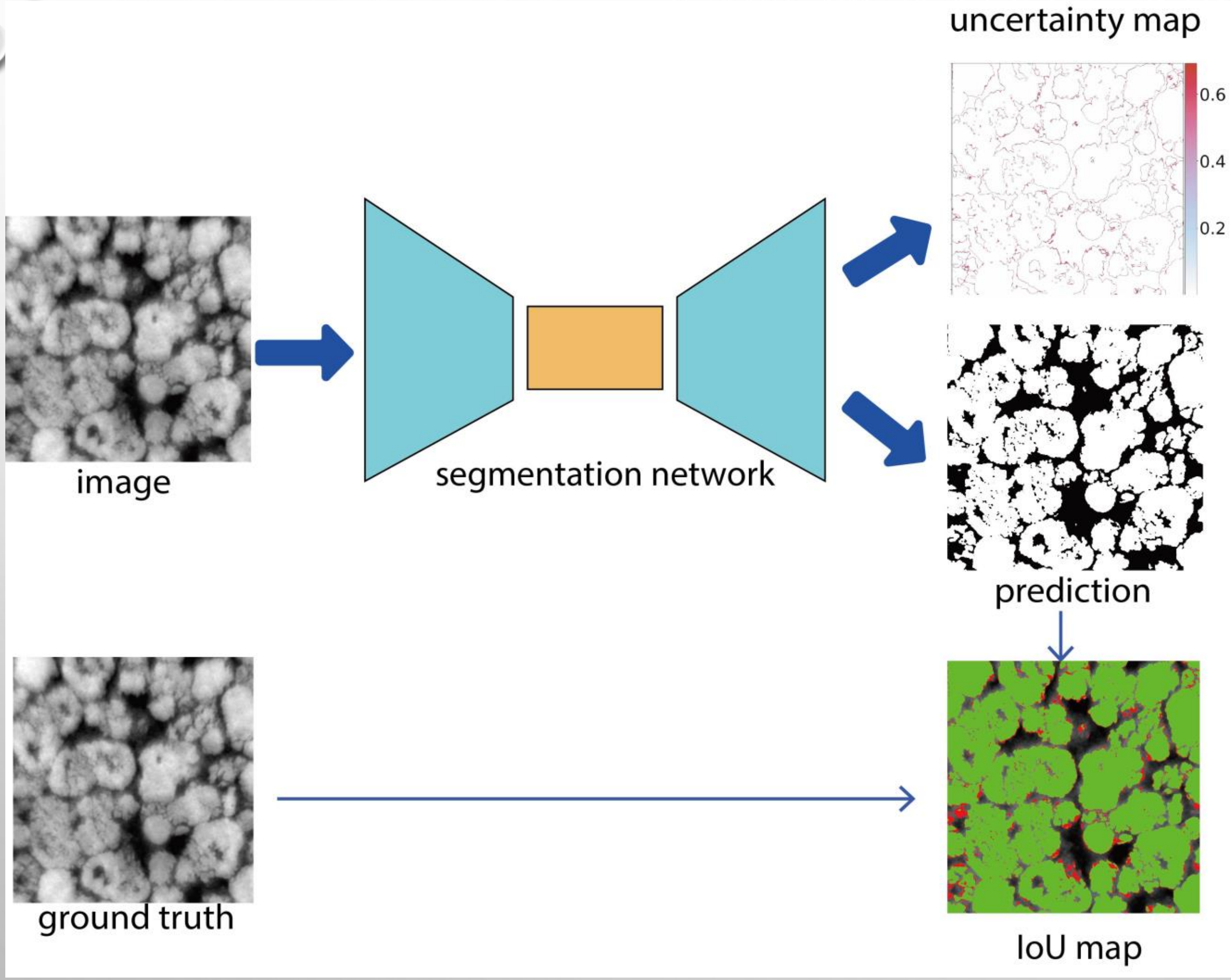
Pearson's linear correlation coefficient (PLCC). It measures the capability of the metric to predict the subjective scores with low error.

$$\mathrm{PLCC} = \frac{\sum_{i=1}^{M_d} \left(\widehat{y}_i - \widehat{y}_{avg}\right)\left(y_i - y_{avg}\right)}{\left(\sum_{i=1}^{M_d} \left(\widehat{y}_i - \widehat{y}_{avg}\right)^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{M_d} \left(y_i - y_{avg}\right)^2\right)^{\frac{1}{2}}}$$

Spearman's rank ordered correlation coefficient (SROCC). It compares the monotonicity of the prediction performance.

$$\mathrm{SROCC} = 1 - \frac{6 \sum_{i=1}^{M_d} (d_i)^2}{M_d (M_d^2 - 1)}$$

Evaluation method - segmentation



High uncertainty

Low uncertainty

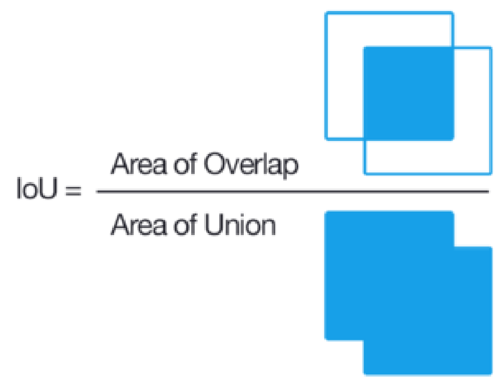
True condition

Predicted condition

	Phase 1 (w)	Phase 2 (b)
Phase 1 (w)	green	blue
Phase 2 (b)	red	black

Evaluation method - segmentation

Intersection-Over-Union (IoU)



IoU calculation visualized. Source: Wikipedia

Dice Coefficient (F1 Score)

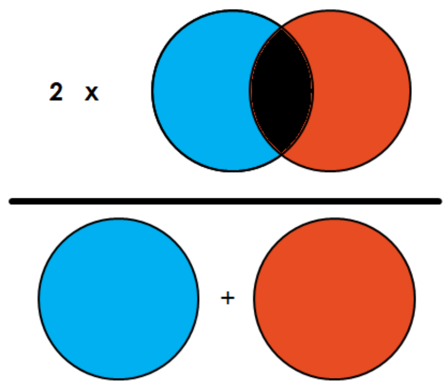
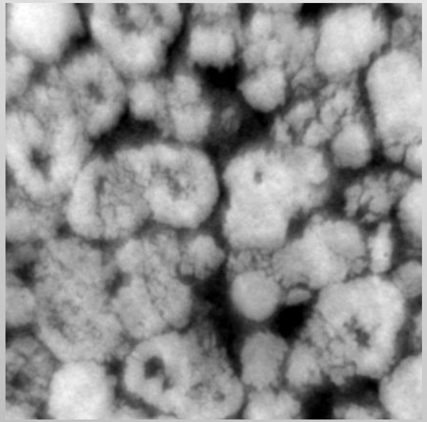


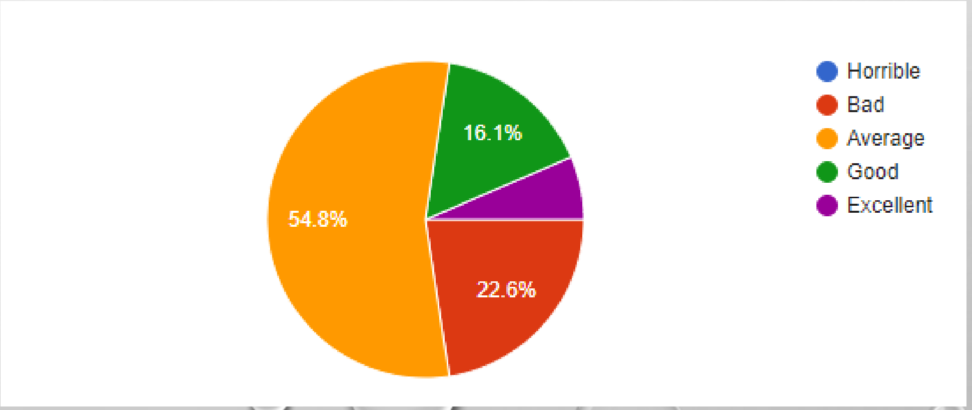
Illustration of Dice Coefficient. 2xOverlap/Total number of pixels

wikipedia

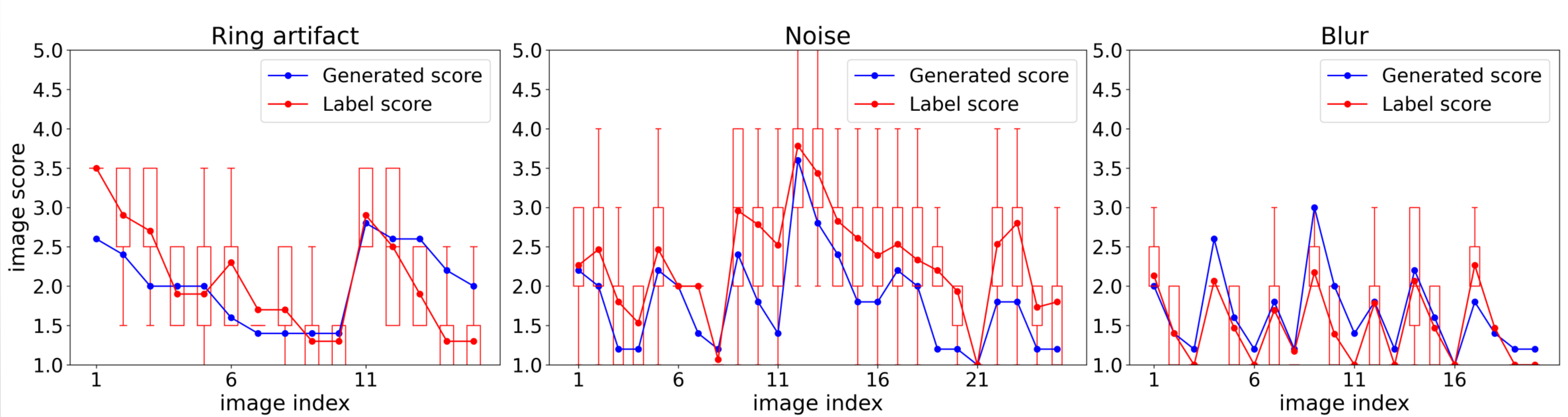
Results – Dataset creation

attributes	description
Total number of images	111 images
Image size	224×224
Image bits	8-bit
Distortion types	Ring artifact, noise, blur
Annotation label	terrible, bad, average, good, excellent
Number of observers for each image	15(averagely)





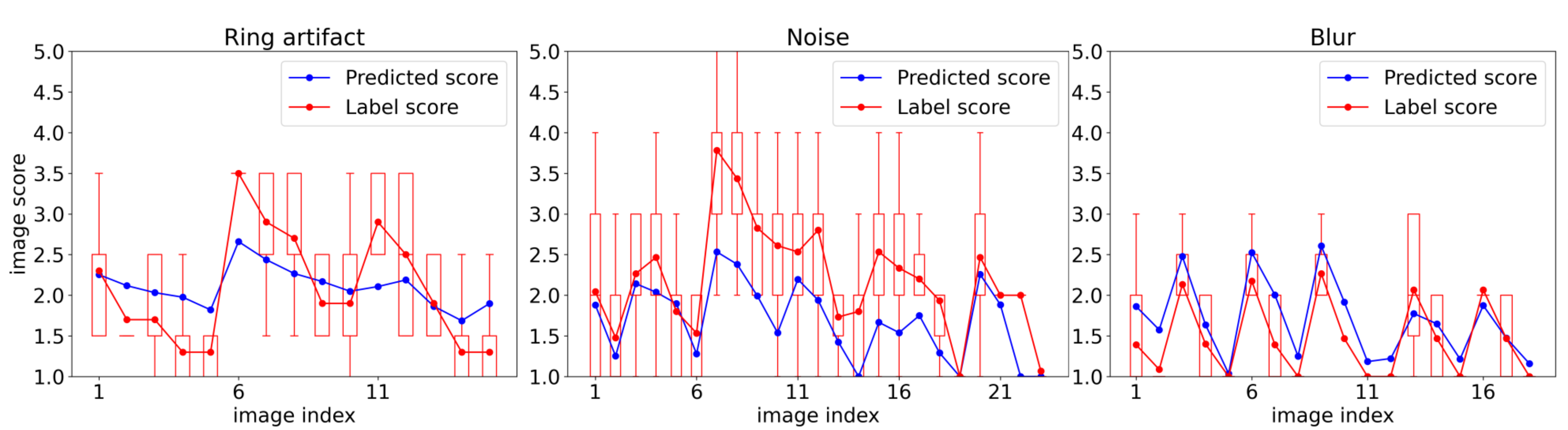
Results – Data generation



The qualitative results of label projection. The red box means the human labels with 95% confidence. The red dot in each line is the average value of labels.

metric / distortion	**Num of images**	**SROCC**	**PLCC**
Ring artifact	15	0.655	0.641
Noise	25	0.813	0.858
Blur	20	0.852	0.846

Results - Score prediction



Evaluation results of score prediction module. The figures from left to right are the evaluation of images with ring artifact distortion, noise and blur, respectively. The red box represents the scores with confidence

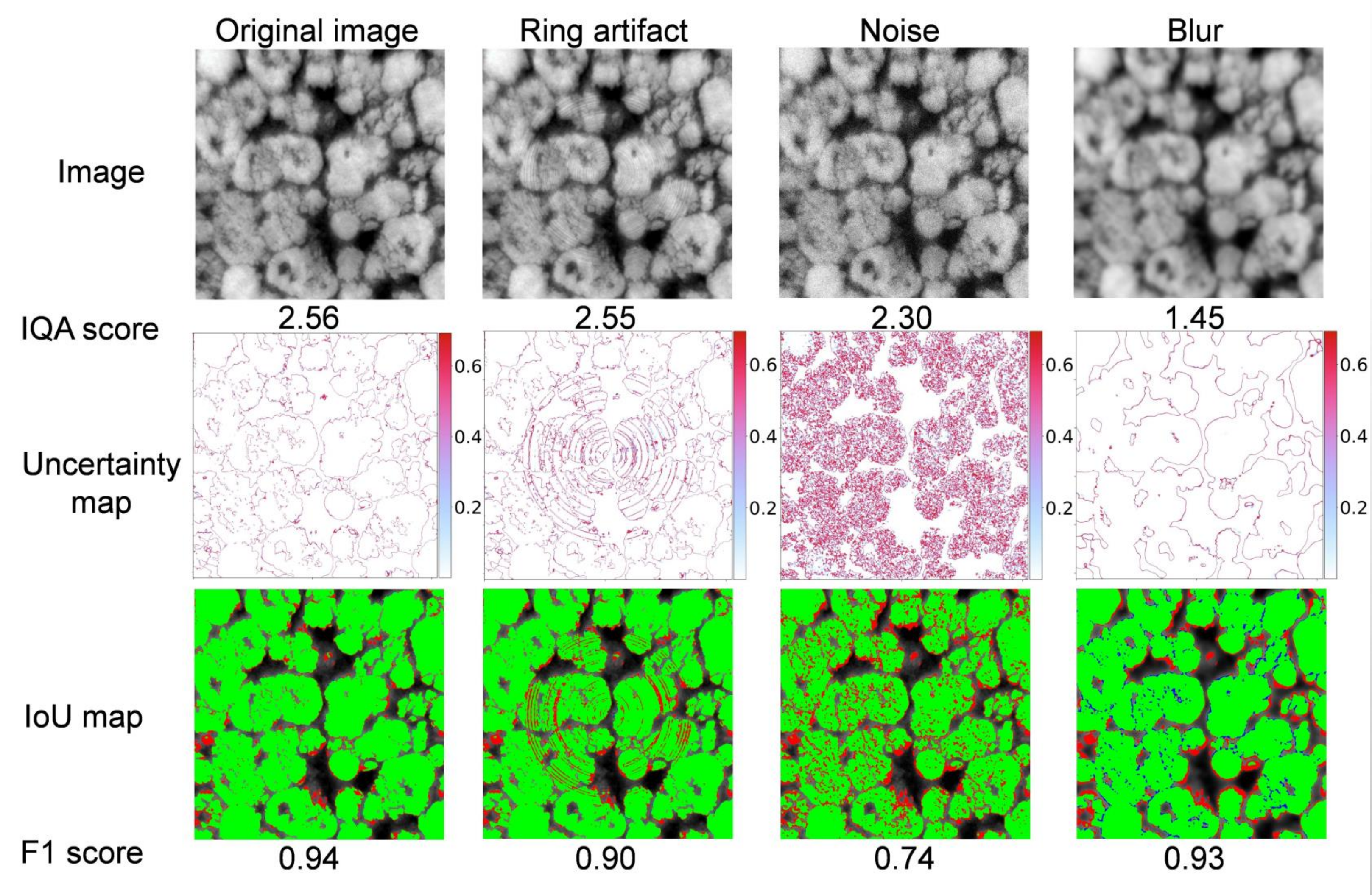
Metric	Method	Ring Artifact	Noise	Blur	ALL
SROCC	BRISQUE [1]	0.785	0.737	0.861	0.794
	RankIQA [2]	0.761	0.769	**0.897**	0.809
	TIQA(ours)	**0.839**	**0.778**	0.895	**0.837**
PLCC	BRISQUE[1]	0.450	0.757	0.849	0.685
	RankIQA[2]	0.758	0.768	**0.879**	0.801
	TIQA(ours)	**0.871**	**0.799**	0.854	**0.841**

138 millions

9 millions

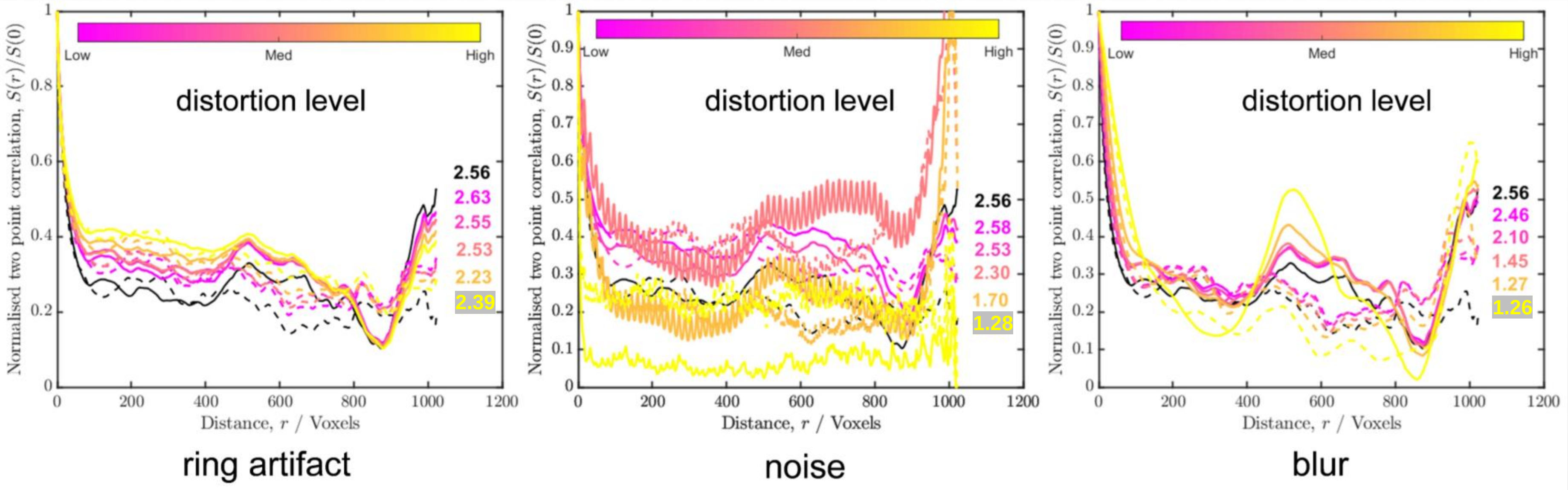
29 images for training;
56 images for testing

Results – Correlation



Results of different distorted images evaluated by TIQA and segmentation. For F1 score, it is in the range of 0 (the worst) and 1 (the best).

Results – Correlation



Point correlation between predicted segmentation mask and ground truth for phase 2. The figures from left to right demonstrate the correlation results from images with different types of distortion. The color bar shows the distortion at different level, from little distortion to severe distortion. The solid line means the point correlation at X direction while the dash line indicates the relation at Y direction.

metric / distortion	**SROCC**	**PLCC**
Ring artifact	0.925	0.887
Noise	0.829	0.877
Blur	0.928	0.952

Quantitative results of the correlation between predicted quality score and segmentation accuracy

Conclusion

- A NR-IQA method is proposed for TIQA which requires a small number of annotations and achieves outperformed results.
- A data generation method is developed by imitating the human observers to label the distorted images automatically for the purpose of addressing the insufficient data problem.
- The correlation between our predicted quality scores and the segmentation performance is investigated.

Reference

1. Mittal, A., Moorthy, A. K. & Bovik, A. C. Blind/referenceless image spatial quality evaluator. in *Conference Record - Asilomar Conference on Signals, Systems and Computers* 723–727 (2011). doi:10.1109/ACSSC.2011.6190099.
2. Liu, X., Weijer, J. Van De & Bagdanov, A. D. RankIQA: Learning from Rankings for No-Reference Image Quality Assessment. in *Proceedings of the IEEE International Conference on Computer Vision* vols 2017-Octob 1040–1049 (Institute of Electrical and Electronics Engineers Inc., 2017).
3. Tan, M. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *36th Int. Conf. Mach. Learn. ICML 2019* 2019-June, 10691–10700 (2019).
4. Zhou, L., Zhang, C. & Wu, M. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* 2018-June, 192–196 (2018).

Thanks