

Computer Vision Metrics

1 Introduction

Here we present a number of tasks in computer vision and detail the metrics applicable to them. The tasks are organized into categories. Within each metric-applied-to-task there are four points that should be covered, as shown in the example in Section 1.1.1. Comments can be made using the `\textcomment{}` command like so: **Comment: Comments can be used for ideas or references to papers.**

1.1 Example CV Task

Basic task description.

1.1.1 Metric

Details of the metric as applied to the given CV task. Points that should be covered:

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2 Image Analysis Tasks

2.1 Object localisation

Basic task description.

2.1.1 Center-cover Criterion

- a. *Metric definition and/or formula:* 1 (hit) if the center of the ground truth reference object is within the bounding box or mask of the predicted detection; 0 (miss) otherwise.

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:* Requires the calculation of the center point for the ground truth reference object, which is simple for a bounding box but more complicated for a mask.
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.1.2 Center-hit criteria

- a. *Metric definition and/or formula:* 1 (hit) if the center of the predicted object is within the bounding box or mask of the ground truth reference object; 0 (miss) otherwise.
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:* Requires the calculation of the center point for the predicted object, which is simple for a bounding box but more complicated for a mask.
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.1.3 Distance-based Hit Criterion

- a. *Metric definition and/or formula:* 1 (hit) if the distance d between the bounding box or mask centers of the ground truth reference object and the predicted object are within distance threshold τ ; 0 (miss) otherwise.
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:* Requires the calculation of the center points for the ground truth reference object and the predicted object, which is simple for a bounding box but more complicated for a mask.
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.1.4 Intersection-over-Union

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.1.5 Box Intersection-over-Union

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.1.6 Mask Intersection-over-Union

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.1.7 Boundary Intersection-over-Union

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.1.8 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2 Object classification

Assigning a label to an object in an image.

2.2.1 Accuracy

- a. *Metric definition and/or formula:*

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.2 Precision

- a. *Metric definition and/or formula:*

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.3 Recall

Also known as: Sensitivity, True Positive Rate.

- a. *Metric definition and/or formula:*

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.4 Specificity

Also known as: True Negative Rate.

- a. *Metric definition and/or formula:*

$$\text{Recall} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.5 F_1 -score

- a. *Metric definition and/or formula:*

$$F_1 = \frac{2}{\text{Recall}^{-1} + \text{Precision}^{-1}} \quad (5)$$

$$= 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

$$= \frac{2 \text{ TP}}{2 \text{ TP} + \text{FP} + \text{FN}} \quad (7)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.6 F_β -score

- a. *Metric definition and/or formula:*

$$F_1 = \frac{\beta^2 + 1}{\beta^2 \cdot \text{Recall}^{-1} + \text{Precision}^{-1}} \quad (8)$$

$$= \frac{(1 + \beta^2) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \quad (9)$$

$$= \frac{(1 + \beta^2) \cdot \text{TP}}{(1 + \beta^2) \cdot \text{TP} + \text{FP} + \beta^2 \cdot \text{FN}} \quad (10)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:* Hyperparameter β which controls the relative importance of Recall versus Precision. $\beta = 1$ weights them equally and reduces F_β to F_1 -score as above; $\beta > 1$ gives more importance to Recall; and $\beta < 1$ gives more importance to Precision.

- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.7 Balanced Accuracy

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.8 Matthew's Correlation Coefficient

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.9 Expected Cost

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.10 AUROC

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.11 Negative Predictive Value

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.12 Positive Likelihood Ratio

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.13 Youden's Index J, aka Bookmaker Informedness

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.14 Cohen's Kappa κ

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.15 Weighted Cohen's Kappa κ

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.16 Sensitivity@Specificity

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.17 Expected Calibration Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.18 Maximum Calibration Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.2.19 Brier Score

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3 Object detection

Detecting one or multiple instances of a given object class in an image. Requires both localizing the instance within the image, and determining the label of the instance. When applying metrics each predicted detection must be paired with a ground truth object via an assignment strategy.

2.3.1 Accuracy

- a. *Metric definition and/or formula:*

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (11)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.2 Precision

- a. *Metric definition and/or formula:*

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.3 Recall

- a. *Metric definition and/or formula:*

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.4 Specificity

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.5 F_1 -score

a. *Metric definition and/or formula:*

$$F_1 = \frac{2}{\text{Recall}^{-1} + \text{Precision}^{-1}} \quad (14)$$

$$= 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

$$= \frac{2 \text{ TP}}{2 \text{ TP} + \text{FP} + \text{FN}} \quad (16)$$

b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*

c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*

d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.6 F_β -score

a. *Metric definition and/or formula:*

$$F_1 = \frac{\beta^2 + 1}{\beta^2 \cdot \text{Recall}^{-1} + \text{Precision}^{-1}} \quad (17)$$

$$= \frac{(1 + \beta^2) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \quad (18)$$

$$= \frac{(1 + \beta^2) \cdot \text{TP}}{(1 + \beta^2) \cdot \text{TP} + \text{FP} + \beta^2 \cdot \text{FN}} \quad (19)$$

b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:* Hyperparameter β which controls the relative importance of Recall versus Precision. $\beta = 1$ weights them equally and reduces F_β to F_1 -score as above; $\beta > 1$ gives more importance to Recall; and $\beta < 1$ gives more importance to Precision.

c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*

d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.7 Balanced Accuracy

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.8 Matthew's Correlation Coefficient

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.9 Expected Cost

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.10 AUROC

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.11 Negative Predictive Value

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.12 Positive Likelihood Ratio

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.13 Youden's Index J, aka Bookmaker Informedness

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.14 Cohen's Kappa κ

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.15 Weighted Cohen's Kappa κ

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.16 Sensitivity@Specificity

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.17 Expected Calibration Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.18 Maximum Calibration Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.19 Brier Score

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.20 Mean Average Precision

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.3.21 Free-Response Receiver Operating Characteristic (FROC), FROC Score

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4 Semantic segmentation

Basic task description.

2.4.1 Dice Similarity Coefficient

Also known as: Sørensen-Dice Index; Sørensen Index; Dice’s Coefficient; F_1 -score when applied to Boolean data.

a. *Metric definition and/or formula:*

$$DSC = \frac{2 |X \cap Y|}{|X| + |Y|} \quad (20)$$

where X and Y are sets with cardinalities (i.e. number of elements) $|X|$ and $|Y|$ respectively.

For Boolean data the DSC is identical to the F_1 score:

$$DSC = \frac{2 \text{ TP}}{2 \text{ TP} + \text{FP} + \text{FN}} \quad (21)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.2 Jaccard Index

Also known as: Intersection over Union (IoU)

a. *Metric definition and/or formula:*

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (22)$$

$$= \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (23)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.3 Pixel Accuracy

- a. *Metric definition and/or formula:*

$$\text{Pixel Accuracy} = \frac{\text{Number of correctly labeled pixels}}{\text{Total number of pixels}} \quad (24)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.4 Hausdorff Distance aka Maximum Symmetric Surface Distance

- a. *Metric definition and/or formula:* (M, d) is a metric space, $X \subset M$ and $Y \subset M$ are two non-empty subsets of M . The Hausdorff distance between X and Y is defined as:

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} d(x, Y), \sup_{y \in Y} d(X, y) \right\} \quad (25)$$

where \sup is the supremum operator, $d(a, B) = \inf_{b \in B} d(a, b)$, and \inf is the infimum operator.

Alternatively,

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\} \quad (26)$$

- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.5 Hausdorff Distance 95% Percentile

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.6 Average Symmetric Surface Distance

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.7 Mean Average Surface Distance

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.8 Boundary IoU

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.9 Absolute/Relative Volume Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.10 Symmetric Relative Volume Difference

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.11 Centerline Coefficient

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.12 Topology Precision

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.13 Topology Sensitivity

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.14 Surface Overlap

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.15 Surface Dice

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.4.16 Volumetric Dice

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.5 Instance segmentation

Basic task description.

2.5.1 Panoptic Quality

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6 Image Regression

Basic task description.

2.6.1 Mean Absolute Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.2 Mean Squared Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.3 Root Mean Squared Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.4 R^2 Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.5 Mean Absolute Percentage Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.6 Root Mean Squared Log Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.7 Symmetric Mean Absolute Percentage Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.8 Mean Directional Accuracy

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.9 Median Absolute Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.6.10 Explained Variance Score

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.7 Panoptic segmentation

Basic task description.

2.7.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.8 Face analysis

Basic task description.

2.8.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.9 Saliency detection

Basic task description.

2.9.1 Increase in Confidence (IIC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.9.2 Average Drop (AD)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.9.3 Average Drop in Deletion

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.9.4 Deletion Area Under Curve (DAUC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.9.5 Insertion Area Under Curve (IAUC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.9.6 Deletion Correlation (DC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.9.7 Insertion Correlation (IC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.10 Scene graph generation

Basic task description.

2.10.1 Recall@K

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.10.2 Mean Recall@K

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.10.3 No Graph Constraint Recall@K

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.11 Instance identification

Basic task description.

2.11.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.12 Feature extraction

Basic task description.

2.12.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.13 Geometric feature detection

Basic task description.

2.13.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.14 Image registration

Basic task description.

2.14.1 Mean Squared Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*

- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.14.2 Mutual Information

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.14.3 Normalized Mutual Information (NMI)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.14.4 Normalized cross correlation (NCC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.14.5 Modality Independent Neighborhood Descriptor

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.15 Image stitching

Basic task description.

2.15.1 Peak Signal-to-Noise Ratio (PSNR)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.15.2 Structure Similarity Index Measurement (SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16 Image similarity

Basic task description.

2.16.1 Mean Squared Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.2 Root Mean Squared Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.3 Normalized Mean Square Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.4 Mean Absolute Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.5 Peak Signal-to-Noise Ratio (PSNR)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.6 Structure Similarity Index Measurement (SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.7 Multi-Scale Structural Similarity index (MS-SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.8 Complex Wavelet Structural Similarity index (CW-SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.9 Feature Similarity

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.10 Universal Image Quality Index

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.11 Learned Perceptual Image Patch Similarity (LPIPS)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.12 Deep Image Structure and Texture Similarity (DISTS)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.13 Information theoretic-based Statistic Similarity Measure (ISSM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.14 Signal to reconstruction error ratio (SRE)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.15 Pearson Correlation Coefficient (PCC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.16 Mutual Information

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.17 Normalized Mutual Information (NMI)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.16.18 Spectral Angle Mapper

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

2.17 Image retrieval (including cross modal retrieval)

Basic task description.

2.17.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3 Spatial Analysis Tasks

3.1 Depth estimation

Basic task description.

3.1.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.2 Object / 3D reconstruction

Basic task description.

3.2.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.3 Object pose estimation

Basic task description.

3.3.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.4 Camera pose estimation

Basic task description.

3.4.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.5 Structure from Motion

Basic task description.

3.5.1 Absolute Trajectory Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*

- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.5.2 Relative Pose Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.6 SLAM

Simultaneously constructing or updating a map of an environment and estimating the robot's state (position, orientation, velocity) within said environment.

3.6.1 Metric

3.6.2 Absolute Trajectory Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.6.3 Relative Pose Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.6.4 Loop Closure Precision and Recall

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.6.5 Map Consistency

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

3.6.6 Map Density and Coverage

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4 Temporal analysis

4.1 Optical flow estimation

Basic task description.

4.1.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.2 Motion detection

Basic task description.

4.2.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.3 Single object tracking

Basic task description.

4.3.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.4 Multiple object tracking

Basic task description.

4.4.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.5 Action recognition

Basic task description.

4.5.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*

- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.6 Event detection

Basic task description.

4.6.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.7 (Multiple) Object re-identification

Basic task description.

4.7.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.8 Visual Odometry

Estimate the state (position, orientation, velocity) of a robot within an environment using camera measurements. Can be used as part of Visual Simultaneous Location and Mapping.

4.8.1 Absolute Trajectory Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.8.2 Relative Pose Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.9 Change detection

Basic task description.

4.9.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.10 Deep fake detection

Basic task description.

4.10.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

4.11 Video summarisation

Basic task description.

4.11.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5 Image Generation Tasks

5.1 Image creation

Basic task description.

5.1.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2 Image denoising

Basic task description.

5.2.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.2 Mean Squared Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.3 Root Mean Squared Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.4 Peak Signal-to-Noise Ratio (PSNR)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.5 Structure Similarity Index Measurement (SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.6 Multi-Scale Structural Similarity index (MS-SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.7 Feature Similarity

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.8 Universal Image Quality Index

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.9 Deep Image Structure and Texture Similarity (DISTS)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.10 Information theoretic-based Statistic Similarity Measure (ISSM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.11 Signal to reconstruction error ratio (SRE)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.12 Visual information fidelity (VIF)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.13 Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.14 Natural Image Quality Evaluator (NIQE)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.15 Perception-based Image Quality Evaluator (PIQE)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.2.16 Model Specialization Metric (MSM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.3 Super-resolution

Basic task description.

5.3.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.4 Image inpainting

Basic task description.

5.4.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5 Image-to-image translation

Basic task description.

5.5.1 Mean Squared Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.2 Normalized Mean Square Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.3 Mean Absolute Error

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.4 Peak Signal-to-Noise Ratio (PSNR)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.5 Structure Similarity Index Measurement (SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.6 Multi-Scale Structural Similarity index (MS-SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.7 Complex Wavelet Structural Similarity index (CW-SSIM)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.8 Learned Perceptual Image Patch Similarity (LPIPS)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.9 Deep Image Structure and Texture Similarity (DISTS)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.10 Pearson Correlation Coefficient (PCC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.11 Normalized Mutual Information (NMI)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.12 FCNScore

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.13 SAMScore

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.14 Mean Blur (MB)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.15 Blur Ratio (BR)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.16 Natural Image Quality Evaluator (NIQE)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.17 Blur Effect (BE)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.18 Variance of Laplacian

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.19 Blurred Edge Widths (BEW)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.20 Cumulative Probability of Blur Detection (CPBD)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.21 Mean Line Correlation (MLC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.22 Mean Shifted Line Correlation (MSLC)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.5.23 Just Noticeable Blur (JNB)

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.6 Style transfer

Basic task description.

5.6.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.7 Object swapping

Basic task description.

5.7.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

5.8 Novel view synthesis

Basic task description.

5.8.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

6 Video Generation Tasks

6.1 Video enhancement

Basic task description.

6.1.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

6.2 Video stabilisation

Basic task description.

6.2.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

6.3 Talking head generation

Basic task description.

6.3.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

7 Multi-modal Tasks

7.1 Visual question answering

Basic task description.

7.1.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

7.2 Image captioning

Basic task description.

7.2.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

7.3 Visual grounding

Basic task description.

7.3.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*

7.4 RGB-depth fusion

Basic task description.

7.4.1 Metric

- a. *Metric definition and/or formula:*
- b. *Any further details, hyperparameters, processing steps, underspecification of the formula, context dependencies, or other ambiguities in the metric itself that can affect reproducibility:*
- c. *Typical range of values obtained with this metric on state-of-the-art systems for this particular task:*
- d. *Any other issues or subtleties that should be mentioned regarding this metric when applied to this task:*