Error Classification System Manual

Our error classification system is designed to offer an effective, fair, and quantitative method for evaluating human errors in AI-generated images. The system focuses on identifying, locating, and measuring errors based on two key dimensions: error categories and body parts. Additionally, we introduce three levels of error severity, which, when combined with error proportion, provide a robust framework for calculating the overall error score for each image. By minimizing subjectivity, this system ensures a fair and consistent approach to assessing human errors in AI-generated imagery.

Error Categories

1. **Missing Errors**
    - **Definition:** Absence of body parts, such as a hand with only two fingers or a person with only one arm. Body parts that are reasonably blocked due to viewing angle or position are not considered missing.
    - **Example:**

Fig 1. images/stable_cascade/couple_hugging/SC_hug_08.png

2. **Extra Errors**
    - **Definition:** Presence of additional body parts, such as a foot with six toes or a person with three hands.
    - **Example:**

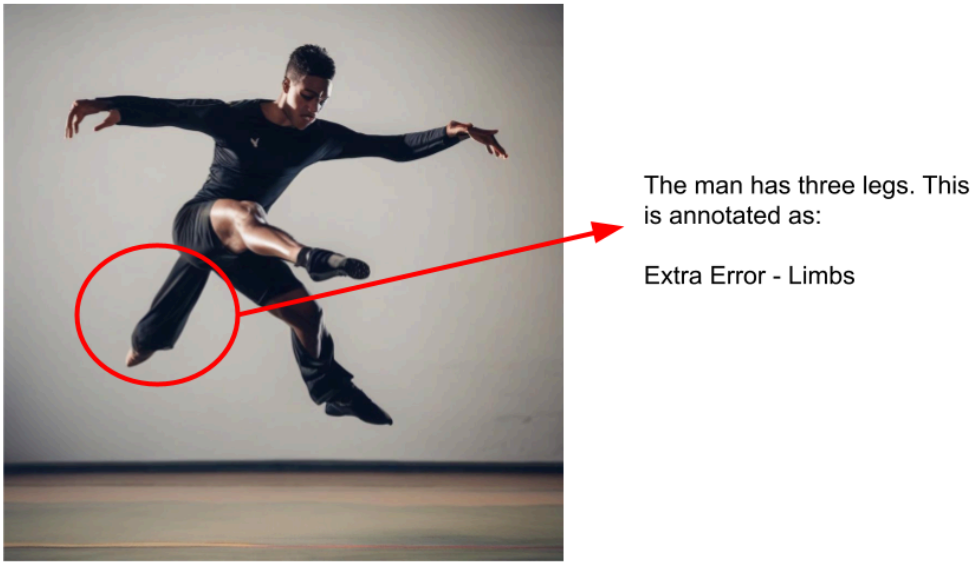

Fig 2. images/sdXL/athlete_performing_salto/sdXL_athlete_04.jpg

3. **Configuration Errors**
    - **Definition:** Disconnection, displacement, or fusion of body parts, either with other body parts or with objects/backgrounds, creating unnatural appearances. Examples include a hand not connected to the arm, an arm growing out of the chest, or fingers fused with a cup.
    - **Example:**

Fig 3. images/dall-e3/people_eating_pizza/dall_e3_pizza_02.png

4. **Orientation Errors**
    - **Definition:** Incorrect or impossible orientation of body parts, such as feet and knees facing in completely opposite directions.
    - **Example:**

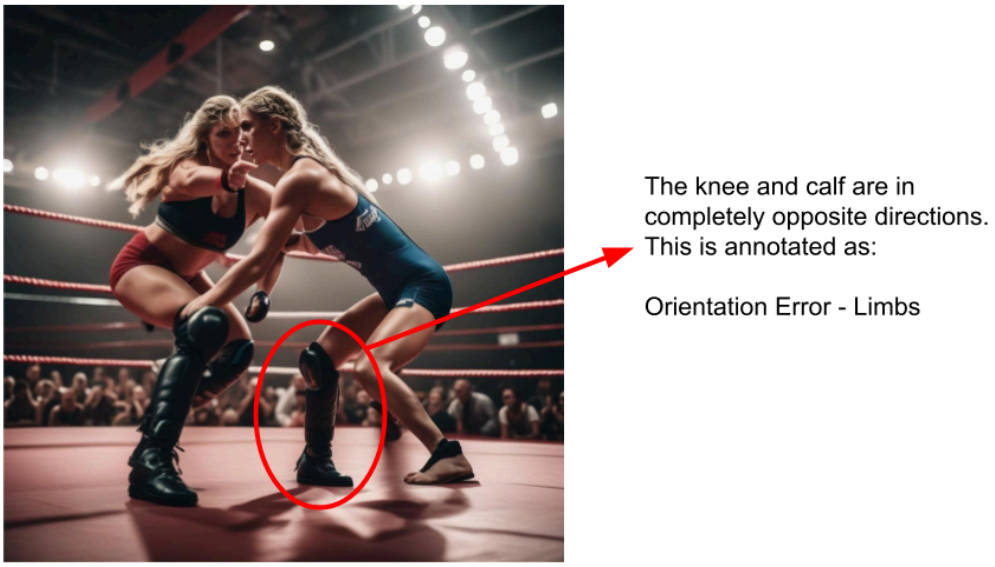

Fig 4. images/sdXL/wrestling_in_arena/sdXL_wrestling_02.jpg

5. **Proportion Errors**
    - **Definition:** Incorrect proportions of body parts, such as calves much shorter than thighs, toes that are too long, or a waist that is disproportionately thin compared to the upper body. Extreme muscle mass or thinness, unless typical for special cases like athletes, also falls under this category. Note that disproportion due to the incomplete generation of body parts is categorized as a Missing Error.
    - **Example:**

Fig 5. images/stable_cascade/person_jogging/SC_jogging_03.png

Body Parts:

1. **Torso**

    Definition: The central axis of the human body, excluding limbs, hands, feet, and face. The head and neck are considered part of the torso.

2. **Limbs**

    Definition: Upper and lower limbs (arms and legs), excluding hands and feet.

3. **Feet**

    Definition: From ankles to toes.

4. **Hands**

   Definition: From wrist joints to fingers.

5. **Face**

   Definition: Forehead to jaw, including eyes, ears, nose, mouth, and other facial features.

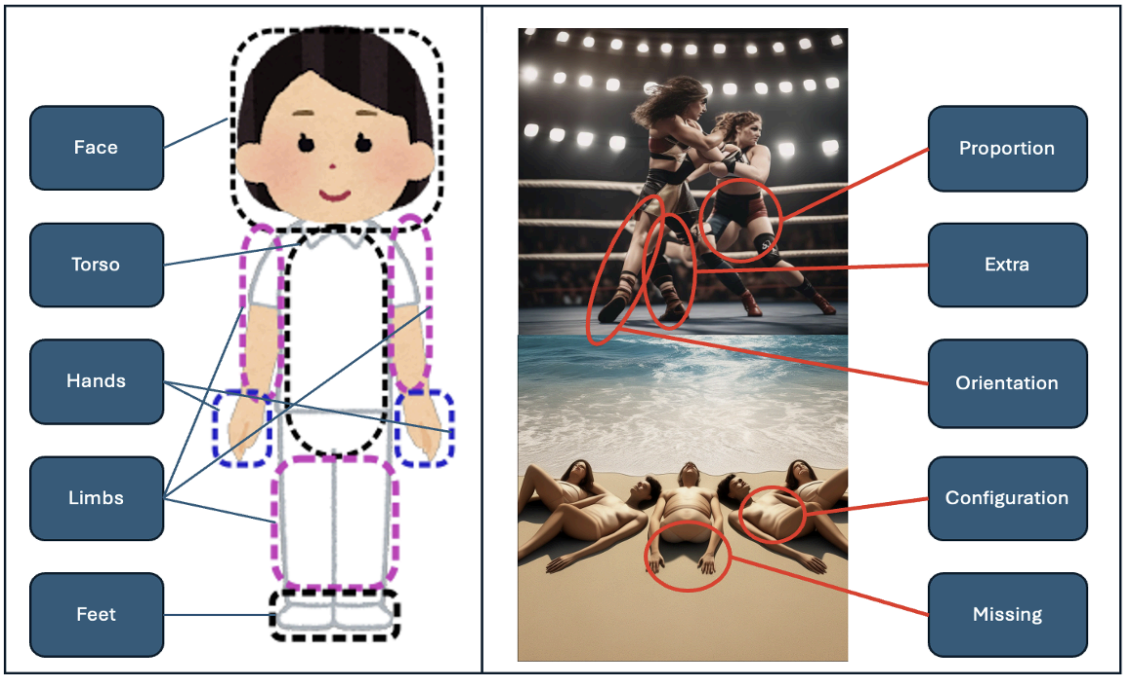

Fig 6. classification_system.png

Error Severity:

To quantify the severity of errors in AI-generated human images, we categorize errors by severity:

- **A:** Mild error, assigned as 0.2 error severity score.
- **B:** Moderate error, assigned as 0.5 error severity score.
- **C:** Severe error, assigned as 1.0 error severity score.

Example:

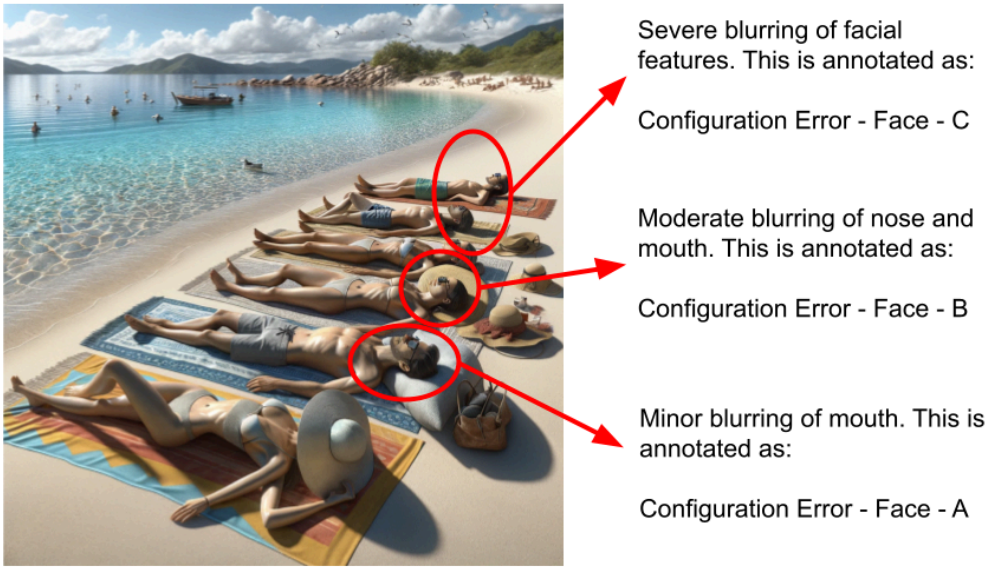

Fig 7. images/dall-e3/five_people_sunbathing_on_beach/dall_e3_sunbathing_02.png

Error Annotation Process:

For each image, annotators must label errors across five error types and five body parts, totaling 25 labeling spaces. Each label is calculated using the formula:
(A error fraction * A error severity score) + (B error fraction * B error severity score) + (C error fraction * C error severity score)

The error fraction is determined as follows:

- For Configuration, Orientation, and Proportion Errors, the denominator is the total number of body parts in the image, and the numerator is the number of body parts with errors of that type and severity.
- For Missing and Extra Errors, the denominator is the number of body parts that should appear in the image, and the numerator is the number of body parts with errors of that type and severity. The "number of body parts that should appear" is based on the image content, not the prompt. For example, if the prompt describes five people but the image shows six, errors should be assessed according to the six people present.

Example:

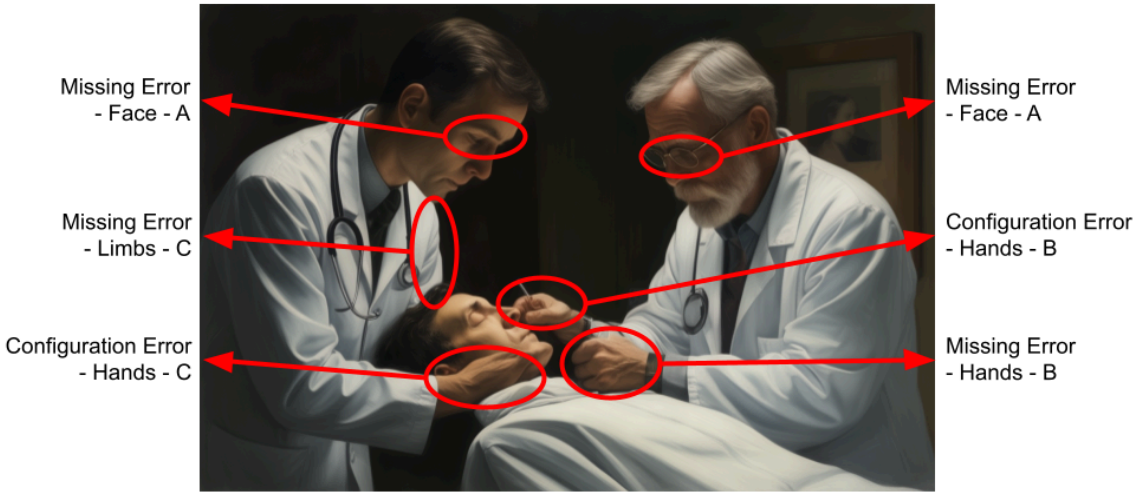

Fig 8. images/stable_cascade/physician_examining_patient/SC_physician_03.png

In the above image, from left to right, we can identify the following errors:

- The left doctor is missing eyeballs, though the eye contours are still present, marked as *Missing Error - Face - A*.
- The left doctor has one arm with only the upper arm present, missing the lower arm, marked as *Missing Error - Limbs - C*. Note that since the absence of the lower arm naturally leads to the hand being missing, we do not count the missing hand as a separate error for this arm.
- The left doctor's other hand is completely fused with the patient's face, marked as *Configuration Error - Hands - C*.
- The right doctor is also missing eyeball details, but the overall contours are still present, marked as *Missing Error - Face - A*.
- One of the right doctor's hands is fused with a tool, though not very obviously, marked as *Configuration Error - Hands - B*.
- The right doctor's other hand is missing two fingers, but compared to the more severe error of an entirely missing hand, it is marked as *Missing Error - Hands - B*.

Therefore, the error annotations for this image are as follows:

- *Missing Error - Face: 2/2 A*, meaning there are 2 faces in total, and both have A-level errors.

- *Missing Error - Limbs: 1/4 C*, meaning there should be 4 limbs in total (excluding the lower limbs due to the viewing angle, so the image should show four upper limbs), with one limb having a C-level error.
- *Configuration Error - Hands: 1/3 C, 1/3 B*, meaning there are 3 hands in total, with one having a C-level error and another having a B-level error.
- *Missing Error - Hands: 1/3 B*, meaning there are 3 hands in total, with one having a B-level error.

By applying the assigned weights, the error score for this image can be calculated as follows:

$$\frac{2}{2} \times 0.2 + \frac{1}{4} \times 1.0 + \frac{1}{3} \times 0.5 + \frac{1}{3} \times 1.0 + \frac{1}{3} \times 0.5 \approx 1.117$$

General Guidelines for Annotation:

- Focus on significant errors that are noticeable without zooming in.
- Prioritize errors that are immediately apparent upon viewing the image.
- When multiple errors are present, give precedence to missing or additional errors.

Notes:

- Errors involving background or objects are only considered if they impact body parts. For example, a bicycle with only one wheel is not an error, but if a person's feet are fused with the bicycle pedals, it is categorized as *Configuration Errors - Feet*.
- Do not consider issues such as unrealistic skin textures. The focus is on structural errors in human anatomy.
- Annotations should be based on the image itself, not the prompt. If the prompt describes five people but the image shows six, annotate according to six people.