Supplemental Document For Paper 6007
OrderChain: Towards General Instruct-Tuning for Stimulating the Ordinal Understanding Ability of MLLM

Jinhong Wang[1,2,3*] Shuo Tong[2*] Jian Liu[3] Dongqi Tang[3] Weiqiang Wang[3] Wentong Li[1]
Hongxia Xu[2] Danny Z. Chen[4] Jintai Chen[5†] Jian Wu[2†]

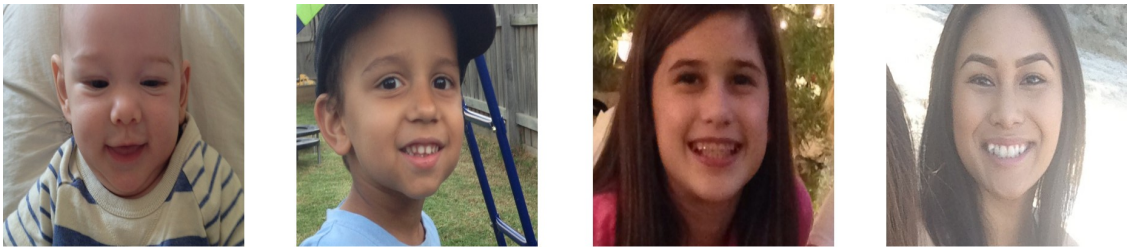
[1]College of Computer Science & Technology, Zhejiang University
[2]Transvascular Implantation Devices Research Institute and Liangzhu Laboratory
[3]Ant Group [4]University of Notre Dame [5]HKUST (Guangzhou)

[*]Contribute equally [†]Corresponding Authors

1. Age Estimation

1.1. Sample Distribution and Visualization

The basic properties of the Adience dataset [1] for age estimation have been presented in the main text. Here we show some samples of different categories (see Fig. 1) and the amount of samples for each category (see Table 1).



(a) 0-2 (b) 4-6 (c) 8-13 (d) 15-20

(e) 25-32 (f) 38-43 (g) 48-53 (h) 60+

Figure 1. Age estimation visualization. Examples of the ordinal categories in the Adience dataset.

Category	1 (0-2)	2 (4-6)	3 (8-13)	4 (15-20)	5 (25-32)	6 (38-43)	7 (48-53)	8 (60+)
Amount	2488	2140	2124	1642	4932	2293	830	872

Table 1. The amount of samples for each category in the Adience dataset.

1.2. Category Recursive Division Details

Since the Adience dataset contains eight categories, we construct a balanced binary tree as in Fig. 2. The corresponding range optimization Chain-of-Thought is presented in the main paper.

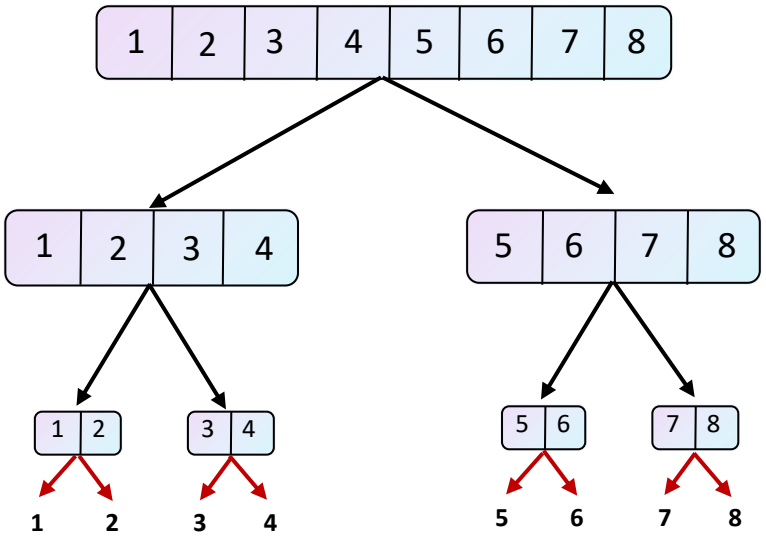


Figure 2. The Category Recursive Division process of the Adience dataset.

1.3. Specific Experimental Settings

For the Adience dataset, all the images are divided into 5 subject-exclusive folds for cross-validation. For training epochs, the maximum epoch number is 3.

2. Historical Image Dating

2.1. Sample Distribution and Visualization

The basic properties of the historical color image (HCI) dataset [2] for historical image dating have been presented in the main text. Here we only show some samples of different categories (see Fig. 3).

(a) 1930s (b) 1940s (c) 1950s (d) 1960s (e) 1970s

Figure 3. Historical image dating visualization. Examples of the ordinal categories in the HCI dataset.

2.2. Category Recursive Division Details

Since the historical color image dataset contains five categories, we construct an incomplete binary tree for it as in Fig. 4.

2.3. Specific Experimental Settings

For the historical color image (HCI) dataset, we randomly split the 265 images of each decade into three subsets: 210 for training, 5 for validation, and 50 for testing. 10-fold cross-validation is performed and the mean values of the results are recorded. For training epochs, the maximum epoch number is 1.

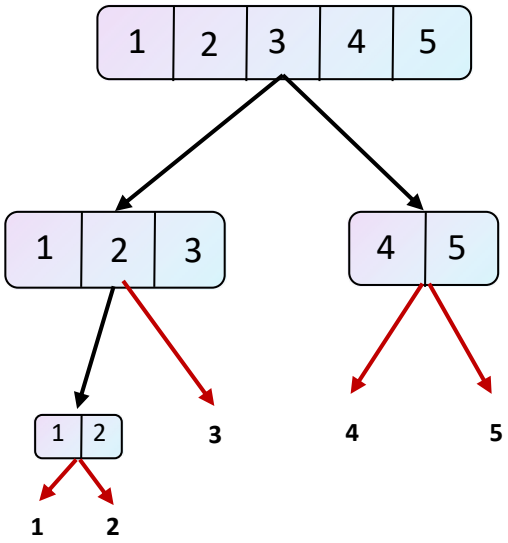


Figure 4. The Category Recursive Division process of the HCI dataset.

3. Image Aesthetics

3.1. Sample Distribution and Visualization

The basic properties of the Aesthetics dataset [3] for image aesthetics grading have been presented in the main text. Here we show some samples of different categories (see Fig. 5) and the amount of samples of each category (see Table 2).

3.2. Category Recursive Division Details

Since this dataset also has five categories, the Category Recursive Division process is the same as that of the historical color image dataset.

3.3. Specific Experimental Settings

For the Image Aesthetics dataset, the images are randomly divided into 75%, 5%, and 20% for training, validation, and testing, respectively. For training epochs, the maximum epoch number is 2.

(a) unacceptable (b) flawed (c) ordinary (d) professional (e) exceptional

Figure 5. Image Aesthetics visualization. Examples of the ordinal categories in the Aesthetics dataset.

Category	1 (unacceptable)	2 (flawed)	3 (ordinary)	4 (professional)	5 (exceptional)
Amount	248	3315	9002	1116	25

Table 2. The amount of samples for each category in the Aesthetics dataset.

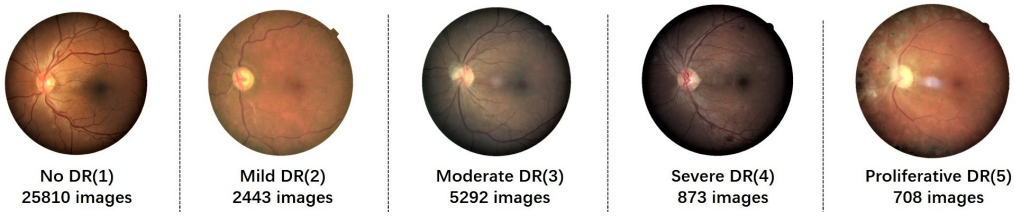


Figure 6. Some sample fundus images with different diabetic retinopathy levels in the DR dataset.

4. Diabetic Retinopathy Grading

4.1. Sample Distribution and Visualization

The Diabetic Retinopathy (DR) dataset is available at https://www.kaggle.com/c/diabetic-retinopathy-detection. Some sample images are shown in Fig. 6.

4.2. Category Recursive Division Details

The Diabetic Retinopathy (DR) dataset contains five categories, but the amuont of samples for category 1 is much larger than those of the other categories. Thus, we construct a customized binary tree opposite to the tree of the above two datasets to balance samples, as shown in Fig. 7.

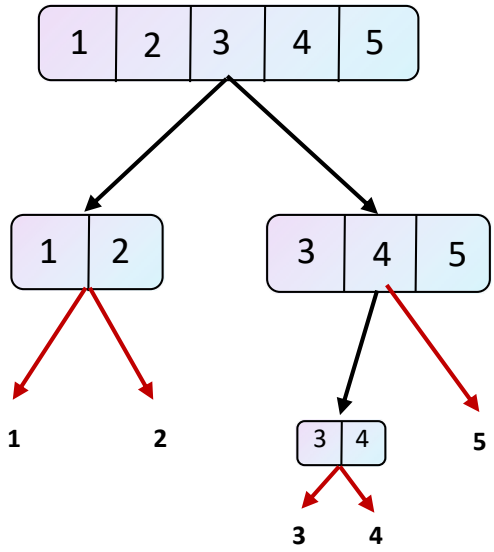


Figure 7. The Category Recursive Division process of the DR dataset.

4.3. Specific Experimental Settings

For the Diabetic Retinopathy (DR) dataset, we apply the subject-independent 10-fold cross-validation and report the mean values of the results. For training epochs, the maximum epoch number is 3.

References

[1] Gil Levi and Tal Hassner. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–42, 2015. 1

[2] Frank Palermo, James Hays, and Alexei A Efros. Dating historical color images. In *European Conference on Computer Vision*, pages 499–512. Springer, 2012. 2

[3] Rossano Schifanella, Miriam Redi, and Luca Maria Aiello. An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 397–406, 2015. 3