Supplemental Document of
Ord2Seq: Regarding Ordinal Regression as Label Sequence Prediction

Jinhong Wang[1*], **Yi Cheng**[1*], **Jintai Chen**[1†], **TingTing Chen**[1], **Danny Chen**[2], **Jian Wu**[1†]

[1]Zhejiang University [2]University of Notre Dame

{wangjinhong, chengy1, wujian2000}@zju.edu.cn

{jtchen721, ttchen0603}@gmail.com dchen@nd.edu

1. Global Hyper-parameter Settings

For both of the Transformer encoder and decoder, we apply the standard settings as in Table 1.

Hyper-parameter	d_model	nhead	dim_forward	dropout	activation	normalization	num_layers
Value	512	8	2048	0.1	Relu	LN	6

Table 1: Hyper-parameter settings of the Transformer encoder and decoder. The hyper-parameter "d_model" is the token size. The hyper-parameter "nhead" defines the amount of heads in computing multi-head attention. The hyper-parameter "dim_forward" is the hidden dimension size of each token. The hyper-parameter "num_layers" defines the layer numbers of both Transformer encoder and decoder.

2. Age Estimation

2.1. 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 2).

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 2: The amount of samples for each category in the Adience dataset.

2.2. Dichotomy Tree Construction

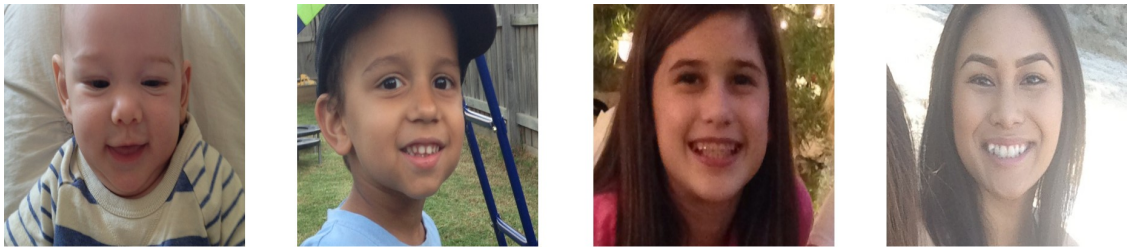
Since the Adience dataset contains eight categories, we construct a complete dichotomy tree as in Fig. 2 for better viewing. According to the tree, we obtain the corresponding binary label sequence and multi-hot sequence for each category, as presented in Table 3.

2.3. Specific Experimental Settings

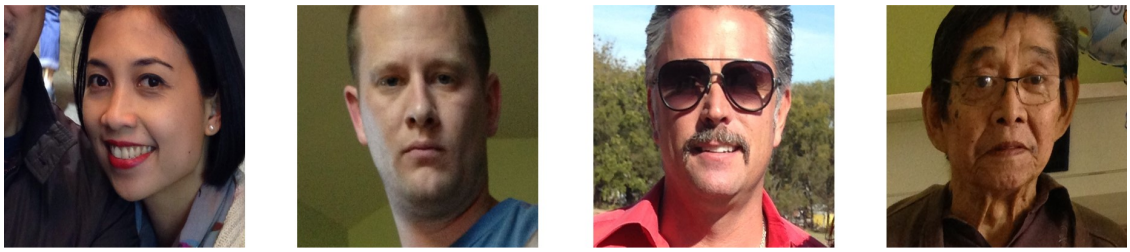
For the training iterations, the warm-up iteration number is 1000 and the max iteration number is 15000. For optimization, the Adam optimizer is utilized with a learning rate of 10^{-4} and the learning rate decays at iteration 7500.

*Equal contribution.

†Corresponding authors.



(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: A visualization of the Adience dataset for age estimation. Some examples of the ordinal categories are shown.

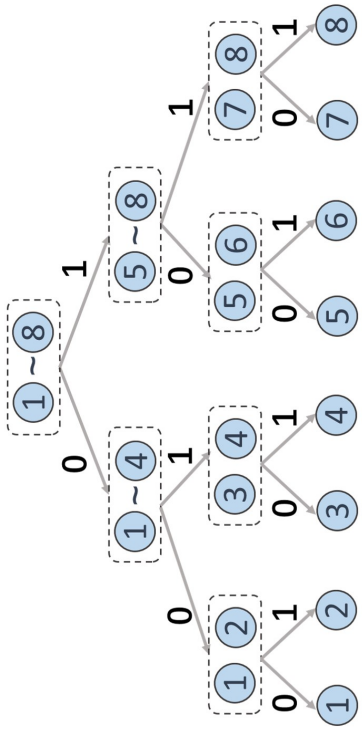


Figure 2: A dichotomy tree for the Adience dataset.

Category	Binary Label Sequence	Multi-hot Label Sequence
1	[0, 0, 0]	[[1, 1, 1, 1, 0, 0, 0, 0], [1, 1, 0, 0, 0, 0, 0, 0], [1, 0, 0, 0, 0, 0, 0, 0]]
2	[0, 0, 1]	[[1, 1, 1, 1, 0, 0, 0, 0], [1, 1, 0, 0, 0, 0, 0, 0], [0, 1, 0, 0, 0, 0, 0, 0]]
3	[0, 1, 0]	[[1, 1, 1, 1, 0, 0, 0, 0], [0, 0, 1, 1, 0, 0, 0, 0], [0, 0, 1, 0, 0, 0, 0, 0]]
4	[0, 1 ,1]	[[1, 1, 1, 1, 0, 0, 0, 0], [0, 0, 1, 1, 0, 0, 0, 0], [0, 0, 0, 1, 0, 0, 0, 0]]
5	[1, 0, 0]	[[0, 0, 0, 0, 1, 1, 1, 1], [0, 0, 0, 0, 1, 1, 0, 0], [0, 0, 0, 0, 1, 0, 0, 0]]
6	[1, 0, 1]	[[0, 0, 0, 0, 1, 1, 1, 1], [0, 0, 0, 0, 1, 1, 0, 0], [0, 0, 0, 0, 0, 1, 0, 0]]
7	[1, 1 ,0]	[[0, 0, 0, 0, 1, 1, 1, 1], [0, 0, 0, 0, 0, 0, 1, 1], [0, 0, 0, 0, 0, 0, 1, 0]]
8	[1, 1, 1]	[[0, 0, 0, 0, 1, 1, 1, 1], [0, 0, 0, 0, 0, 0, 1, 1], [0, 0, 0, 0, 0, 0, 0, 1]]

Table 3: Binary label and Multi-hot label sequences of each category.

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. 3) and the amount of samples of each category (see Table 4).

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

Figure 3: A visualization of the Aesthetics dataset for image aesthetics. Some examples of the ordinal categories are shown.

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

Table 4: The amount of samples for each category in the Aesthetics dataset.

3.2. Dichotomy Tree Construction

Since the Aesthetics dataset contains five categories, we construct an incomplete dichotomy tree for it as in Fig. 4. Thus, we obtain the corresponding binary label sequence and multi-hot sequence for each category, as presented in Table 5.

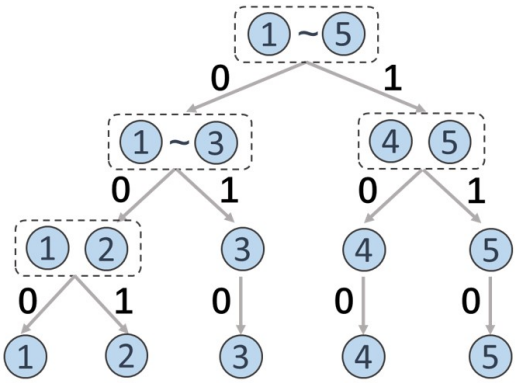


Figure 4: A dichotomy tree for the Aesthetics dataset.

Category	Binary Label Sequence	Multi-hot Label Sequence
1	[0, 0, 0]	[[1, 1, 1, 0, 0], [1, 1, 0, 0, 0], [1, 0, 0, 0, 0]]
2	[0, 0, 1]	[[1, 1, 1, 0, 0], [1, 1, 0, 0, 0], [0, 1, 0, 0, 0]]
3	[0, 1, 0]	[[1, 1, 1, 0, 0], [0, 0, 1, 0, 0], [0, 0, 1, 0, 0]]
4	[1, 0 ,0]	[[0, 0, 0, 1, 1], [0, 0, 0, 1, 0], [0, 0, 0, 1, 0]]
5	[1, 1, 0]	[[0, 0, 0, 1, 1], [0, 0, 0, 0, 1], [0, 0, 0, 0, 1]]

Table 5: The binary label and multi-hot label sequences of each category.

3.3. Specific Experimental Settings

For the training iterations, the warm-up iteration number is 1000 and the max iteration number is 20000. For optimization, the Adam optimizer is utilized with a learning rate of 10^{-4} with no learning rate decay.

4. Historical Image Dating

4.1. 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. 5).

4.2. Dichotomy Tree Construction

The dichotomy tree construction is the same as that of the Image Aesthetics.

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

Figure 5: A visualization of the HCI dataset for the historical image dating task. Some examples with the ordinal category labels are shown.

4.3. Specific Experimental Settings

For the training iterations, the warm-up iteration number is 1 and the max iteration number is 1700. For optimization, the Adam optimizer is utilized with a learning rate of 10^{-4} with no learning rate decay.

5. Diabetic Retinopathy Grading

5.1. Dichotomy Tree Construction

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 an incomplete dichotomy tree for it as in Fig. 6 to better recognize categories 2 and 3. Then, we obtain the corresponding binary label sequence and multi-hot sequence for each category, as presented in Table 6.

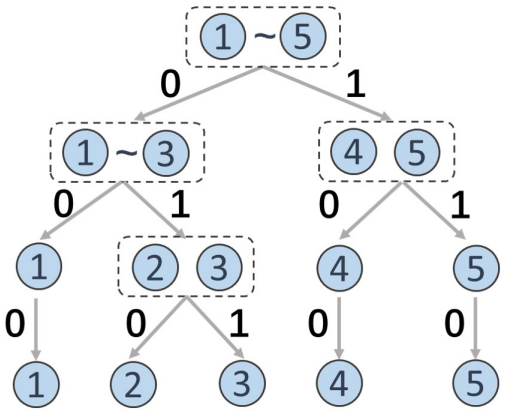


Figure 6: A dichotomy tree for the DR dataset.

Category	Binary Label Sequence	Multi-hot Label Sequence
1	[0, 0, 0]	[[1, 1, 1, 0, 0], [1, 0, 0, 0, 0], [1, 0, 0, 0, 0]]
2	[0, 1, 0]	[[1, 1, 1, 0, 0], [0, 1, 1, 0, 0], [0, 1, 0, 0, 0]]
3	[0, 1, 1]	[[1, 1, 1, 0, 0], [0, 1, 1, 0, 0], [0, 0, 1, 0, 0]]
4	[1, 0 ,0]	[[0, 0, 0, 1, 1], [0, 0, 0, 1, 0], [0, 0, 0, 1, 0]]
5	[1, 1, 0]	[[0, 0, 0, 1, 1], [0, 0, 0, 0, 1], [0, 0, 0, 0, 1]]

Table 6: The binary label and multi-hot label sequences of each category.

5.2. Specific Experimental Settings

For the training iterations, the warm-up iteration number is 1000 and the max iteration number is 50000. For optimization, the Adam optimizer is utilized with a learning rate of 10^{-4} with no learning rate decay.

6. Additional Experiment Results on Age Estimation

To further validate the performances of our Ord2Seq on the age estimation task, we conduct additional experiments on the UTK and CACD datasets and compare the results with the current state-of-the-art method MWR. The MAE scores are reported in Table. 7. It is observed that our Ord2Seq outperforms MWR by a large margin on MAE (↓) on these two facial age estimation datasets with a large number of categories, demonstrating the superiority of our method.

Method	UTK	CACD (train)	CACD (valid)
MWR [4]	4.37	4.41	5.68
Ord2Seq (VGG)	3.19	3.48	4.63

Table 7: The performance of MWR and our Ord2Seq on UTK and CACD datasets.

7. Potential Application Scenario Discussion

Rather than the scenarios mentioned in the main paper, our proposed method also has diverse potential applications. For example, it can be used in monocular depth estimation where the distance can be discretized to make monocular depth estimation become an ordinal regression task. Thus, the number of categories can be large since the distance interval can be chosen to be very small. In such a situation, we think our method can offer the potential of attaining much more accurate results than the known methods. And we leave a detailed discussion of these scenarios for future exploration.

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. 3

[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*, volume 9, pages 397–406, 2015. 2

[4] Nyeong-Ho Shin, Seon-Ho Lee, and Chang-Su Kim. Moving window regression: A novel approach to ordinal regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18760–18769, 2022. 5