

# Smart Karyotyping Image Selection based on Commonsense Knowledge Reasoning

Yufeng Xu <sup>a</sup>, Zhe Ding <sup>a</sup>, Lei Shi <sup>a</sup>, Juan Wang <sup>a,b</sup>, Linfeng Yu <sup>a</sup>, Haoxi Zhang <sup>a,b</sup>,  
Edward Szczerbicki <sup>c</sup>

<sup>a</sup> School of Cybersecurity, Chengdu University of Information Technology,  
Chengdu, China; <sup>b</sup> Advanced Cryptography and System Security Key Laboratory of  
Sichuan Province, Chengdu, China; <sup>c</sup> Faculty of Management and Economics, Gdansk  
University of Technology, Gdansk, Poland

**Abstract.** Karyotyping requires chromosome instances to be segmented and classified from the metaphase images. One of the difficulties in chromosome segmentation is that the chromosomes are randomly positioned in the image, and there is a great chance for chromosomes to be touched or overlap with others. It is always much easier for operators and automatic programs to tackle images without overlapping chromosomes than ones with largely overlapped chromosomes. In order to reduce the processing difficulty, adding a smart image selection procedure ahead of segmentation is practical and necessary. In this paper, we introduce the Smart Karyotyping Image Selection (SKIS) based on Commonsense Knowledge Reasoning. The initial experiment demonstrates that the proposed approach can select the expected images based on reasoning and benefit following karyotyping processes.

---

Address correspondence to Haoxi Zhang, School of Cybersecurity, Chengdu University of Information Technology, No. 24 Block 1, Xuefu Road, Chengdu, China, 610225. E-mail: [haoxi@cuit.edu.cn](mailto:haoxi@cuit.edu.cn)

**Keywords:** Commonsense knowledge, knowledge reasoning, karyotyping, image selection

## INTRODUCTION

Karyotyping (Clare 2008) is a critical technique in chromosome analysis to detect genetic abnormalities, such as chronic myelogenous leukaemia, which is usually performed by culturing the cells and separating the chromosomes from the nucleus during the metaphase stage of cell division and then staining them on slides for microphotography and analysis (Sharma et al. 2017). It requires chromosome instances in the microscopy to be segmented, classified, and chronologically arranged. Among these steps, the segmentation of chromosomes in metaphase images is one of the most challenging problems for machine learning models. Many researchers have attempted to address it, and various image-processing algorithms have been proposed to separate over-lapping chromosomes (Cao et al. 2011; Ji. 1994; Li et al. 2016). Nonetheless, chromosome segmentation remains difficult for human operators and automatic programs (Arora and Dhir. 2016; Shen et al. 2019).

Generally, one of the difficulties in chromosome segmentation is that the chromosomes are randomly positioned in the image, and there is a great chance for chromosomes to be touched or overlap with others. Consequently, these touched or overlapped chromosomes significantly challenge operators and automatic programs in segmenting them individually from metaphase images. It is always much easier for operators and automatic programs to tackle metaphase images without overlapping chromosomes than ones with largely overlapped chromosomes. In order to reduce the processing difficulty, adding a smart image selection procedure ahead of segmentation is practical and necessary.

Traditional methods like image quality assessment (IQA) do not necessarily equal image selection assessment. Fuderer (1988) proposed an objective measure of Magnetic Resonance (MR) image quality using Shannon's theoretical information content of the

MR images where he took into account the scan resolution, the Contrast-to-Noise Ratio (CNR), and the pixel size related to the Field of View (FOV). The scan resolution of the karyotyping image we evaluate is identical; hence we do not need to assess the image based on the scan resolution. Moreover, CNR cannot be used to assess the distribution of chromosomes. The International Telecommunication Union of the television industry proposed Double-Stimulus Continuous-Quality Scale (Series. 2012), which was used as an assessment tool for fast imaging and cannot help identify chromosome overlapping or touching.

The Learning based Blind Image Quality (LBIQ) uses a regression algorithm to incorporate numerous image quality features, *i.e.*, natural image statistics, distortion texture statistics, and blur/noise statistics (Tang et al. 2011). Even though blur/noise measurement is necessary for image quality, it is not the critical feature for metaphase image selection. In addition, LBIQ is based on a regression algorithm that lacks interpretability. Dutta et al. (2013) use a quantitative statistical method with closed-form analytical expressions to measure medical image quality on resolution and covariance. The covariance of images cannot effectively assess the dispersion of chromosomes in images. The Perceptual Difference Model (PDM) (Shiao et al. 2007; Daly et al. 1992) is proposed to calculate the visual difference between a test image and a standard reference image. Huo et al. (2006) used PDM to evaluate the images produced by different values of regularization parameters in parallel MR imaging. However, the semantic of less overlap or less touch can barely be measured by computing the general feature distance between two images. Moreover, due to the variety of chromosomes, it is hard to set an image as the standard reference.

To succeed in image selection, we argue that the capabilities of evaluating the degrees of dispersion, overlapping, and touching of the chromosomes in metaphase images are critical. Therefore, we introduce the Smart Karyotyping Image Selection (SKIS) based on commonsense knowledge reasoning, a method using commonsense knowledge and reasoning to evaluate images, making it explainable and very friendly to clinical operators.



## COMMONSENSE KNOWLEDGE AND REASONING

Commonsense knowledge (CSK) is information that humans usually have that helps them make sense of situations in daily life (Ilievski et al. 2021). It has been predominately created directly from human input or extracted from the text (Lenat et al. 1990; Liu and Singh. 2004; Carlson et al. 2010). CSK can generally be considered to be possessed by most people, and, according to the Gricean maxim (Grice. 1975), it is usually omitted in (written or oral) communication. People take CSK for granted since they understand CSK naturally. CSK has an exceedingly large scale in both amount and diversity. Based on these characteristics, CSK is defined as a tremendous amount and variety of knowledge of default assumptions about the world, which is shared by (possibly a group of) people and seems so fundamental and obvious that it usually does not explicitly appear in people's communications (Zang et al. 2013). Commonsense knowledge differs from encyclopaedic knowledge as it deals with general knowledge rather than the details of specific entities (Tandon et al. 2018). Most regular knowledge bases (KBs) contribute millions of facts about entities such as geopolitical entities or people but fail to provide fundamental knowledge, such as the notion that a child is likely too young to have a master's degree in mathematics. The fact that commonsense knowledge is often implicit presents a challenge for question-answering (QA) approaches and automated natural language processing (NLP) in that the extraction and learning algorithms cannot rely on the commonsense knowledge being available directly in the text (Ilievski et al. 2021). Commonsense is elusive because it is scarcely and often only implicitly expressed, it is affected by reporting bias (Gordon and Durme. 2013), and it may require considering multiple modalities.

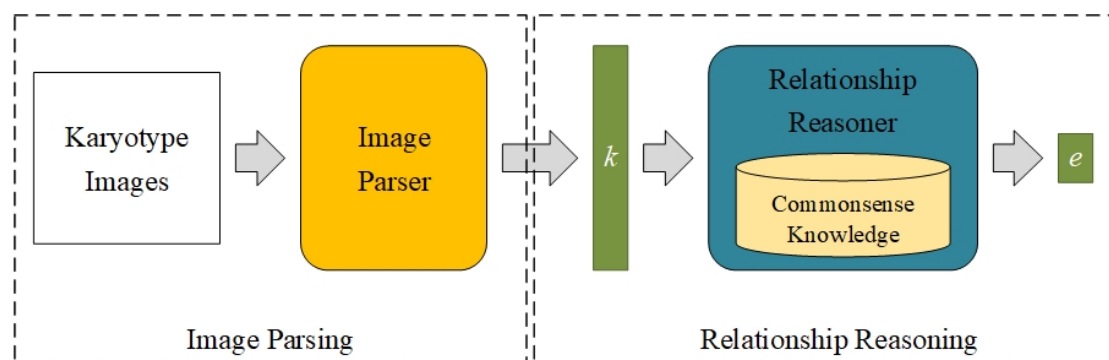
Commonsense reasoning is making deductions based on everyday knowledge and reusing it as background knowledge during inference. For instance, we can predict that if a person enters a room, the person will be in the room afterwards. Because we make such inferences so quickly, we might assume that commonsense reasoning is simple. But it is not easy, requiring a large amount of knowledge about the world and the ability to use it to reason about the world (Mueller 2014). There are a few different methods



for commonsense reasoning in AI. One common method is using a reasoning engine that can logically reason information. Another popular method is to use a knowledge base, which is a collection of facts and information about the world. In this work, we combine the two methods by creating a reasoner and a knowledge base. This approach allows more flexible reasoning and rule management as the reasoner and knowledge base are independent.

## SMART KARYOTYPING IMAGE SELECTION

### A. The System Overview



**Figure 1.** The structure of the Smart Karyotyping Image Selection based on Commonsense Knowledge Reasoning.

The proposed Smart Karyotyping Image Selection (SKIS) based on commonsense knowledge reasoning consists of two stages: image parsing and relationship reasoning (see Figure 1). The image parsing acquires the image representation for a given metaphase image. It takes a two-step and detection-based approach (Yi et al. 2018) for de-rendering: it first generates a set of chromosome object proposals and then predicts the positions of each chromosome via these proposals. As a result, the image representation contains information about chromosome instances and their positions in the image. The relationship reasoning maps an image representation to a latent program based on commonsense knowledge and integrates it with a fully transparent and interpretable reasoner. The program has a hierarchy of functional knowledge modules, each operating independently on the image representation. The reasoner applies these

functional knowledge modules to the image representation and generates a final evaluation score from processing the output sequence of these knowledge modules.

### B. Image Parsing

We use a pre-trained object detection network called DINO (Zhang et al. 2022) to parse karyotyping images. It consists of the Swin-Large (Liu et al. 2021), a multi-layer Transformer (Vaswani et al. 2017) as its backbone encoder, a multi-layer Transformer decoder, and multiple prediction heads. Given a chromosome metaphase image, we extract multi-scale chromosome features with the Swin-Large Transformer backbone and feed them into the Transformer encoder with corresponding positional embeddings. After feature enhancement with the encoder layers, the mixed query selection strategy is used to initialize anchors as positional queries for the decoder. With the initialized anchors and the learnable content queries, deformable attention (Zhu et al. 2020) is used to update the queries layer by layer and combine the features of the encoder outputs. Finally, bounding boxes and classification results predicted by refined content features comprise the image representation  $k$  as the output of the image parsing process.

Specifically, in the image representation  $k$ , each bounding box is defined by two coordinates:

$$box_i = [(x_0, y_0), (x_1, y_1)] \quad (1)$$

where  $(x_0, y_0)$  represents the lower left corner of the box, and  $(x_1, y_1)$  is the coordinate of the upper right corner. Based on this representation, we can have information about the box's centre point, width, and height using (2) and (3), respectively:

$$box\_width_i = x_1 - x_0, \quad box\_height_i = y_1 - y_0, \quad (2)$$

$$box\_centerpoint_i = \left( \frac{x_0+x_1}{2}, \frac{y_0+y_1}{2} \right). \quad (3)$$

### C. Chromosome Relationship Reasoning

This stage has two primary components: the commonsense knowledge base and the relationship reasoner.

- 1) Commonsense knowledge base (CSK base)

In our CSK base, commonsense knowledge (CSK) is represented as a latent program

consists a set of relations  $s_i = \{(op) r_a, (op, r_b)\}$ , where  $r$  is an object-object relation (Yatskar et al. 2016) and  $op$  is a logic operation. The CSK base is used to manage the CSKs  $S = \{s_1, \dots, s_n\}$ . To reason the chromosome relationships, we utilize the method of Region Connection Calculus (RCC) (Randell et al. 1992). For example, we can create a piece of CSK about *good images* as:

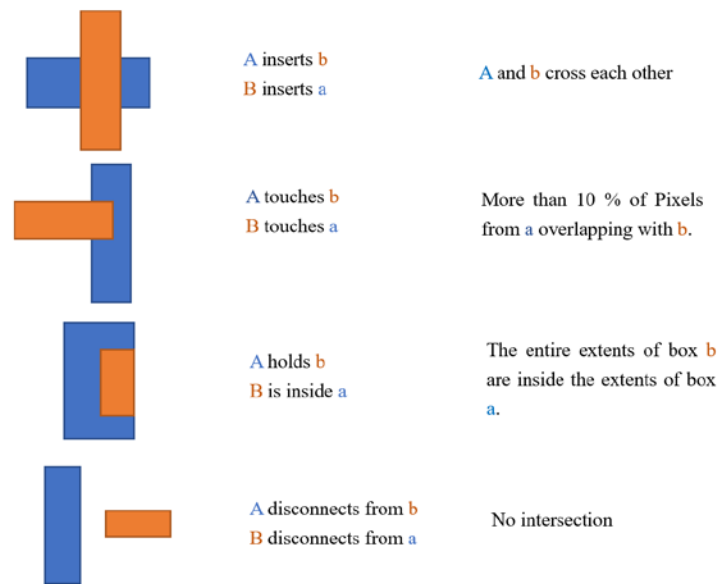
$$s = \{\neg insert \wedge \neg touch \wedge \neg hold\}, \quad (4)$$

where *insert*, *touch*, and *hold* are object-object relations, which can be further defined as three functional knowledge modules for pairs of chromosomes. For instance, we can determine whether two chromosomes *insert* each other by using the following module:

$$\max(Ax_0, Bx_0) \leq \min(Ax_1, Bx_1) \quad \&\& \quad \max(Ay_0, By_0) \leq \min(Ay_1, By_1), \quad (5)$$

where  $A(x_0, y_0)$  represents the lower left corner of box  $A$ , and  $B(x_1, y_1)$  is box  $B$ 's upper right corner. Our commonsense knowledge is inspired by RCC and the Visual Dependency Grammar of (Elliott et al. 2014; Elliott and de Vries. 2015)(see Figure 2).

In this work, we define four types of relations: *insert*, *touch*, *hold*, and *disconnect*.



**Figure 2.** We define four types of unique relations: {insert, touch, hold, disconnect}.

## 2) Relationship Reasoner

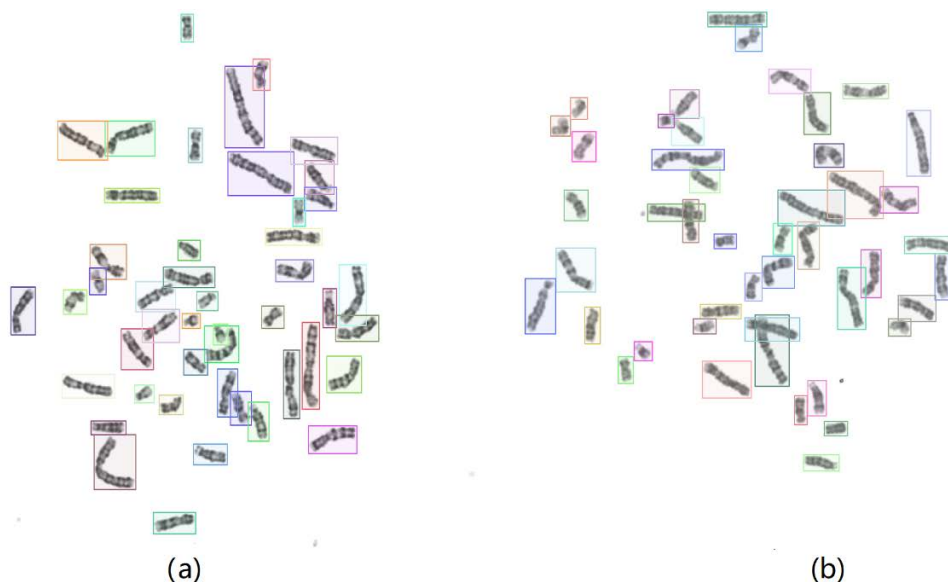
The relationship reasoner is a program executor based on commonsense knowledge as programs consisting of collections of functional modules and is designed to host all logic operations behind the relationship contained in functional modules.

Each functional module is in one-to-one correspondence with the image representation  $k$  from the input. The modules equipped with commonsense knowledge share the same input/output interface and can be arranged in any length and order.

A typical program begins with a relation token in the CSK and ends with the last relation token. Each functional module of the relation is sequentially performed and returns results to the reasoner. After the last module returns, the reasoner calculates the final score based on the CSK.

### THE INITIAL EXPERIMENT

To evaluate the performance of our SKIS, we compare the performance with the similarity-based method. Our dataset is created from metaphase images of daily clinic cases with overlapping and adherent chromosomes collected at the Department of Medical Genetics/Prenatal Diagnostic Center, West China Second Hospital. The resolution of images is 1280 x 1024, and they are annotated using the annotation tool LabelMe.



**Figure 3.** Metaphase images for evaluation. Left (a) with fewer insertions. Right (b) with more insertions and touches.

We choose two typical images (see Figure 3) to demonstrate the effectiveness of



our method. In Figure 3 (a), the chromosomes are evenly distributed with fewer touches and insertions, while there are more touches and insertions in Figure 3 (b). Images with fewer insertions are better than ones with more, hence the higher scores. Table 1 shows the evaluation results of the proposed method (*i.e.*, SKIS) and the similarity-based method.

**Table 1 comparison on two images and two evaluation methods**

Method	The score of Figure 3 (a)	The score of Figure 3 (b)
Similarity	116.77	119.67
SKIS	85.28	70.14

As shown in Table 1, SKIS gives a higher score (*i.e.*, 85.28) on (a) and a lower score on (b), which is correct and reasonable because (a) has fewer insertions and touches. On the contrary, the similarity-based method suggests (b) over (a) by predicting a higher score on (b). The method based on similarity cannot measure the chromosome distribution of metaphase images robustly and precisely due to its inability to reason the content of the images.

## CONCLUSIONS AND FUTURE WORK

In this paper, we introduce Smart Karyotyping Image Selection based on commonsense knowledge reasoning. The proposed method evaluates metaphase images based on inference with commonsense knowledge represented as functional modules. Unlike similarity-based methods, our method can assess images according to the object-object relations in images' content rather than calculating the images' overall similarities. The initial experiment shows that our method can correctly evaluate the metaphase images.

Since this research is at its early stage, we plan to do further research and refinement in:

- Knowledge representation and relations of chromosomes.



- Further development and refinement of the reasoning algorithm.
- Further design and development of the image representation.

## **ACKNOWLEDGEMENT**

The authors would like to thank the editors and anonymous reviewers for their valuable comments and suggestions on this paper. This work was supported by the Kemoshen Science Research Program.

## **REFERENCES**

Arora, T., & Dhir, R. 2016. A review of metaphase chromosome image selection techniques for automatic karyotype generation. *Medical & biological engineering & computing*, 54(8), 1147-1157.

Cao, H., Deng, H. W., & Wang, Y. P. 2011. Segmentation of M-FISH images for improved classification of chromosomes with an adaptive fuzzy C-means clustering algorithm. *IEEE Transactions on fuzzy systems*, 20(1), 1-8.

Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka, E. R., & Mitchell, T. M. 2010. Toward an architecture for never-ending language learning. In *Twenty-Fourth AAAI conference on artificial intelligence*.

Clare O'Connor. 2008. Karyotyping for chromosomal abnormalities. *Nature Education*, 1(1):27.

Daly, S. J. 1992. Visible differences predictor: an algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing, and Digital Display III* (Vol. 1666, pp. 2-15). SPIE.

Dutta, J., Ahn, S., & Li, Q. 2013. Quantitative statistical methods for image quality assessment. *Theranostics*, 3(10), 741.

Elliott, D., Lavrenko, V., & Keller, F. 2014. Query-by-example image retrieval using visual dependency representations. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 109-120).

Elliott, D., & de Vries, A. 2015. Describing images using inferred visual dependency representations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 42-52).



Fuderer, M. 1988. The information content of MR images. *IEEE Transactions on Medical Imaging*, 7(4), 368-380.

Gordon, J., & Van Durme, B. 2013. Reporting bias and knowledge acquisition. In *Proceedings of the 2013 workshop on Automated knowledge base construction* (pp. 25-30).

Grice, H. P. 1975. Logic and conversation. In *Speech acts* (pp. 41-58). Brill.

Huo, D., Xu, D., Liang, Z. P., & Wilson, D. 2006. Application of perceptual difference model on regularization techniques of parallel MR imaging. *Magnetic Resonance Imaging*, 24(2), 123-132.

Ilievski, F., Oltramari, A., Ma, K., Zhang, B., McGuinness, D. L., & Szekely, P. 2021. Dimensions of commonsense knowledge. *Knowledge-Based Systems*, 229, 107347.

Ji, L. 1994. Fully automatic chromosome segmentation. *Cytometry: The Journal of the International Society for Analytical Cytology*, 17(3), 196-208.

Johnson, J., Hariharan, B., Van Der Maaten, L., Hoffman, J., Fei-Fei, L., Lawrence Zitnick, C., & Girshick, R. 2017. Inferring and executing programs for visual reasoning. In *Proceedings of the IEEE international conference on computer vision* (pp. 2989-2998).

Lenat, D. B., Guha, R. V., Pittman, K., Pratt, D., & Shepherd, M. 1990. Cyc: toward programs with common sense. *Communications of the ACM*, 33(8), 30-49.

Li, Y., Knoll, J. H., Wilkins, R. C. Flegal, F. N., & Rogan, P. K. 2016. Automated discrimination of dicentric and monocentric chromosomes by machine learning-based image processing. *Microscopy research and technique*, 79(5), 393-402.

Liu, H., & Singh, P. 2004. ConceptNet—a practical commonsense reasoning toolkit. *BT technology journal*, 22(4), 211-226.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10012-10022).

Mueller, E. T. 2014. Commonsense reasoning: an event calculus based approach. Morgan Kaufmann.

Randell, D. A., Cui, Z., & Cohn, A. G. 1992. A spatial logic based on regions and connection. *KR*, 92, 165-176.

Series, B. T. 2012. Methodology for the subjective assessment of the quality of television pictures. *Recommendation ITU-R BT*, 500-13.

Sharma, M., Saha, O., Sriraman, A., Hebbalaguppe, R., Vig, L., & Karande, S. 2017. Crowdsourcing for chromosome segmentation and deep classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 34-41).

Shen, X., Qi, Y., Ma, T., & Zhou, Z. 2019. A dicentric chromosome identification method based on clustering and watershed algorithm. *Scientific reports*, 9(1), 1-11.

Shiao, Y. H., Chen, T. J., Chuang, K. S., Lin, C. H., & Chuang, C. C. 2007. Quality of compressed medical images. *Journal of Digital Imaging*, 20(2), 149-159.

Tandon, N., Varde, A. S., & de Melo, G. 2018. Commonsense knowledge in machine intelligence. *ACM SIGMOD Record*, 46(4), 49-52.

Tang, H., Joshi, N., & Kapoor, A. 2011. Learning a blind measure of perceptual image quality. In *CVPR 2011* (pp. 305-312). IEEE.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yatskar, M., Ordonez, V., & Farhadi, A. 2016. Stating the obvious: Extracting visual common sense knowledge. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 193-198).

Yi, K., Wu, J., Gan, C., Torralba, A., Kohli, P., & Tenenbaum, J. 2018. Neural-symbolic vqa: Disentangling reasoning from vision and language understanding. *Advances in neural information processing systems*, 31.

Zang, L. J., Cao, C., Cao, Y. N., Wu, Y. M., & Cao, C. G. 2013. A survey of commonsense knowledge acquisition. *Journal of Computer Science and Technology*, 28(4), 689-719.

Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., & Shum, H. Y. 2022. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*.

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. 2020. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*.