


Measuring joint attention in co-creation through automatic human activity recognition

Tao Shen ¹, Yanyi Li², Yonqqi Lou¹, Chun Liu², Danwen Ji¹, Man Zhang³ and Ying Li¹

¹College of Design and Innovation, Tongji University, Shanghai, China

²College of Surveying and Geo-Informatics, Tongji University, Shanghai, China

³Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai, China

Abstract

Within the broad context of design research, joint attention within co-creation represents a critical component, linking cognitive actors through dynamic interactions. This study introduces a novel approach employing deep learning algorithms to objectively quantify joint attention, offering a significant advancement over traditional subjective methods. We developed an optimized deep learning algorithm, YOLO-TP, to identify participants' engagement in design workshops accurately. Our research methodology involved video recording of design workshops and subsequent analysis using the YOLO-TP algorithm to track and measure joint attention instances. Key findings demonstrate that the algorithm effectively quantifies joint attention with high reliability and correlates well with known measures of intersubjectivity and co-creation effectiveness. This approach not only provides a more objective measure of joint attention but also allows for the real-time analysis of collaborative interactions. The implications of this study are profound, suggesting that the integration of automated human activity recognition in co-creation can significantly enhance the understanding and facilitation of collaborative design processes, potentially leading to more effective design outcomes.

Keywords: Joint attention, Co-creation, Human activity recognition, Deep learning, Data-driven innovation

1. Introduction

In the intricate tapestry of design research, the phenomenon of joint attention within co-creation emerges as a pivotal thread, weaving together cognitive actors and their dynamic interactions (Falck-Ytter *et al.* 2023; Sani-Bozkurt & Bozkus-Genc 2023). Rooted in the broader paradigm of intersubjectivity, joint attention epitomizes the confluence of individual cognitive processes within a shared, collaborative space. The overarching concept of intersubjectivity, as postulated by Fuchs & De Jaegher (2009) and Racine & Carpendale (2007), encapsulates a shared intellectual and emotional state. Yet, the specific manifestation of joint attention within co-creation remains an underexplored niche, warranting rigorous academic scrutiny.

T.S. and Y.L. have made equal contributions to the work of this paper.

Received 27 December 2024

Revised 04 July 2025

Accepted 10 July 2025

Corresponding authors

Yonqqi Lou and Chun Liu; Emails:

louyongqi@tongji.edu.cn;

liuchun@tongji.edu.cn

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

Des. Sci., vol. 11, e33

journals.cambridge.org/dsj

DOI: [10.1017/dsj.2025.10021](https://doi.org/10.1017/dsj.2025.10021)



Traditional approaches to assessing co-creation effectiveness have relied heavily on subjective observation methods, including self-report questionnaires, expert evaluations and post-hoc interviews (Shalley, Zhou & Oldham 2004; Heiss & Kokshagina 2021). These subjective methods, while valuable for capturing experiential dimensions of collaboration, present several critical limitations that constrain the advancement of co-creation research. First, subjective observations are inherently susceptible to observer bias, where researchers' theoretical preconceptions and expectations can influence their interpretation of collaborative behaviors (Nguyen & Mougenot 2022). Second, self-report measures suffer from retrospective bias, as participants may struggle to accurately recall or articulate the nuanced dynamics that occurred during co-creation activities (Cash, Dekoninck & Ahmed-Kristensen 2020). Third, the temporal granularity of traditional methods is insufficient for capturing the micro-level interactions that constitute joint attention, as these methods typically rely on broad, summary assessments rather than moment-by-moment behavioral tracking (Behoora & Tucker 2015). Fourth, the scalability of subjective methods is limited, as expert observation becomes resource-intensive when applied across multiple sessions or large numbers of participants (Kassner, Patera & Bulling 2014; Spagnolli *et al.* 2014).

Quantitative research methodologies offer complementary advantages that can address these limitations while preserving the valuable insights from qualitative approaches. Objective measurement systems provide consistent, reproducible data that is independent of observer interpretation, enabling more reliable cross-study comparisons and meta-analyses (Kent *et al.* 2022). The temporal precision of computer vision-based approaches allows for the capture of brief, ephemeral moments of joint attention that might be missed by human observers (Erichsen *et al.* 2021). Furthermore, quantitative methods enable the analysis of large datasets, facilitating the identification of patterns and relationships that emerge across multiple co-creation sessions (Hansen & Özkil 2020). Most importantly, the integration of quantitative and qualitative approaches creates a more comprehensive understanding of co-creation dynamics, where objective behavioral indicators can validate and complement subjective experiential reports (Kleinsmann, Valkenburg & Sluijs 2017).

Co-creation, as a distinctive form of collaborative activity, represents more than mere coordination of efforts among participants. As Cash *et al.* (2021) demonstrate in their Design Science research, it involves complex intersubjective dynamics where participants not only work together but also actively construct shared meaning through joint attentional processes. This stands in contrast to general collaboration, which may involve coordinated action without necessarily sharing cognitive and emotional states. The distinction is crucial for understanding how design knowledge emerges through collective processes rather than individual contributions alone.

Historically, the realm of intersubjectivity has been predominantly assessed through the lens of self-descriptions and expert observations, as elucidated by Shalley *et al.* (2004). While these traditional methodologies offer invaluable insights, they are not devoid of limitations. The inherent subjectivity of self-descriptions, coupled with the potential biases of expert observations, often culminates in data that may be riddled with discrepancies (So *et al.* 2023). This lacuna underscores the exigency for a more objective, data-driven approach to deciphering joint attention in co-creation.

Recent advancements in design research methodology have begun to address this gap through computational approaches to measuring design activity. Kent *et al.* (2022) demonstrate how network analysis can reveal patterns in prototyping activities that remain invisible to traditional observation methods. Similarly, Erichsen *et al.* (2021) have pioneered digital approaches to capturing physical design artifacts, providing more objective measures of design processes. These methodological innovations align with what Cash *et al.* (2020) identify as a broader trend toward more rigorous, quantitative approaches to understanding design cognition and collaboration.

Enter the realm of human activity recognition, underpinned by deep learning algorithms. The burgeoning advancements in this domain proffer the tantalizing possibility of objectively quantifying joint attention, transcending the constraints of subjective interpretations (Ozdemir, Akin-Bulbul & Yildiz 2024). This aligns with growing recognition in the design research community of the need for more robust measurement approaches to intersubjective phenomena. Our study addresses this methodological gap by developing a computational framework for measuring joint attention in co-creation contexts.

This study, therefore, embarks on an ambitious odyssey to harness the prowess of deep learning in elucidating the nuances of joint attention within co-creation. By developing and validating a quantitative measurement framework for joint attention in co-creation, our research contributes to what Hansen & Özkil (2020) identify as a critical need for more objective approaches to understanding design collaboration. This framework not only enables more precise assessment of co-creation effectiveness but also provides a foundation for evidence-based enhancement of co-creation processes across diverse domains.

The ensuing discourse is meticulously structured to offer a holistic overview of the research. Section 2 delves into a comprehensive literature review, elucidating the theoretical underpinnings of intersubjectivity and its manifestations in co-creation. Section 3 unveils the optimized deep learning algorithm tailored for human activity recognition. Section 4 delineates the research methodology, encompassing the procedural intricacies and data sources. Section 5 elucidates the empirical findings, accentuating the reliability and intercorrelations of joint attention measures. Finally, Section 6 culminates in a synthesis of the research insights, charting potential avenues for future exploration.

Our study makes three principal contributions to the field of design research. First, it develops a novel computational approach to measuring joint attention in co-creation contexts, addressing what Kleinsmann *et al.* (2017) identify as a significant methodological gap in design research. Second, it identifies and quantifies three dimensions of joint attention – empathic sharing, social context and key area – providing a structured framework for understanding intersubjectivity in co-creation. Third, it establishes weighted indicators that enable design researchers and practitioners to optimize co-creation environments for enhanced effectiveness. Collectively, these contributions advance the field beyond subjective assessments of co-creation toward more rigorous, evidence-based approaches to understanding and facilitating this complex form of collaborative design activity.

2. Literature review

2.1. Joint attention in the cocreation process

Co-creation is a multifaceted concept that has evolved significantly over time across various disciplines. At its core, co-creation refers to the joint creation of value by multiple stakeholders through collaborative processes (Prahalad & Ramaswamy 2004). In the management and marketing literature, co-creation has been conceptualized as “the joint creation of value by the company and the customer; allowing the customer to co-construct the service experience to suit their context.”

From a design perspective, Sanders & Stappers (2008) define co-creation as “any act of collective creativity, i.e., creativity that is shared by two or more people.” In this context, co-creation encompasses various participatory approaches for design and decision-making with diverse participants (Ramaswamy & Ozcan 2018), distinguished by assisted involvement in orchestrated multistakeholder interactions, such as formal workshops and self-organizing modes of engagement.

Importantly, across these diverse conceptualizations, several common elements emerge: co-creation involves multiple stakeholders working together, it is an interactive and collaborative process and it aims to create value that benefits all involved parties (Ind & Coates 2013; Ramaswamy & Ozcan 2018). These characteristics highlight the fundamentally social nature of co-creation processes.

The limitations of traditional co-creation assessment methods have been increasingly recognized in recent design research literature. Lloyd & Oak (2018) identified significant challenges in capturing the temporal dynamics of collaborative design processes through conventional observation methods, noting that “categories, stories, and value tensions” often emerge through micro-interactions that escape traditional documentation approaches. Similarly, Andersen & Mosleh (2021) demonstrated that conflicts and resolutions in co-design activities occur through subtle gestural and spatial interactions that require fine-grained temporal analysis to understand fully. These findings support the need for more sophisticated measurement approaches that can capture the nuanced behavioral indicators underlying effective co-creation (Cooper 2023). Furthermore, Devos & Loopmans (2022) emphasize the importance of embodied intersubjectivity in co-creation processes, arguing that traditional verbal and survey-based assessments fail to capture the full spectrum of collaborative engagement that occurs through physical presence and spatial interaction.

In recent years, a growing number of scholars have sought to investigate the social dimension of the co-creation process (Park *et al.* 2023). Analogous to the study of individual designers, codesign is categorized as social psychology, a type of design work that relies on ongoing, subtle social interactions and transformative work involving the design of artifacts (Button & Sharrock 1996). According to Devos and Loopmans, co-creation involves the enactment of creation through interactions that go beyond mere collaboration between two or more human actors, thereby revealing its inherently social nature (Devos & Loopmans 2022). Within this context, a multitude of studies have focused on examining social interactions and conflicts that emerge during the design process (Andersen & Mosleh 2021). This line of inquiry encompasses various facets, including how designers create artifacts or employ them to facilitate and promote collaborative

efforts (Andersen & Mosleh 2021; Christensen & Abildgaard 2021), as well as the gestures and sketches they produce during interactions. The primary objectives of these research endeavors are to foster collaboration and communication (Howard & Bevins 2022), resolve conflicts and discrepancies (Le Bail, Baker & D tienne 2022), make decisions (Cooper 2023) and ultimately examine the nature of joint and collaborative meaning-making, which is of paramount importance (Ind & Coates 2013).

The measurement of co-creation processes presents significant challenges due to its abstract and multifaceted nature. Various approaches have been proposed in the literature, focusing on different aspects of co-creation. Some studies have measured outcomes, such as innovation performance (Frow *et al.* 2015) or customer satisfaction (Gr nroos & Voima 2013), while others have examined process aspects like customer participation (Yi & Gong 2013) or collaboration quality (Ranjan & Read 2016). In the context of design, researchers have investigated the development of shared understanding during collaborative work (Cash *et al.* 2020) and significant “episodes” during the process (Lloyd & Oak 2018).

Therefore, this paper places a greater emphasis on the role of interaction in co-creation as a means of promoting mutual understanding and embodied cognition (Devos & Loopmans 2022), which in turn helps to create and reconstruct our own and others’ roles in the process of relating (Mosleh & Larsen 2021), thus fostering social innovation. To understand the positive impact of social interactions that emerge during a series of design processes, we must borrow relevant concepts from the fields of social psychology, cognitive science and human-computer interaction (Wang, Kim & Lin 2024). Additionally, we need to develop appropriate measures and evaluation methods that allow us to assess the quality and effectiveness of social interaction in the context of co-creation. Researchers have focused on individual behaviors by investigating the “episodes” during the process that are especially significant to participants (Lloyd & Oak 2018) and have presented the relations among collaborative design work and the development of shared understanding (Cash *et al.* 2020).

Most recent studies have measured these indices using qualitative or subjective questionnaires (Heiss & Kokshagina 2021). Quantitative approaches reported in the literature require wearing intrusive sensors for each participant (Kassner *et al.* 2014; Spagnolli *et al.* 2014). Behoora & Tucker (2015) examine interpersonal interactions in cocreation activities by using computer vision and machine learning methods to measure the emotional state of individuals. However, separately identifying motions and facial expressions to assess their “interaction” behavior during multiperson cocreation is inaccurate.

After thorough consideration of various potential measures, joint attention has been selected as the indicator for measuring co-creation in this study. This selection is based on several theoretical and practical considerations that align with the social and interactive nature of co-creation processes:

- (1) Fundamental to social interaction: Joint attention refers to the ability of two or more individuals to focus on the same object, event or person with the intention of interacting with each other (Tomasello 1995; Moore, Dunham & Dunham 2014). This directly corresponds to the interactive nature of co-creation, which involves multiple stakeholders working together toward shared goals (Pralhad & Ramaswamy 2004).

- (2) Indicator of shared understanding: Joint attention is closely linked to the development of shared understanding among participants (Carpenter, Nagell & Tomasello 1998), which is a crucial aspect of effective co-creation (Ind & Coates 2013). By measuring joint attention, we can assess the extent to which participants are developing a common ground for collaboration.
- (3) Observable and measurable: Unlike some abstract aspects of co-creation, joint attention can be observed and measured through behaviors, such as gaze direction, gestures and verbal references (Mundy, Sullivan & Mastergeorge 2007), making it a practical indicator for empirical research.
- (4) Associated with positive outcomes: Research has shown that joint attention is associated with improved communication, enhanced problem-solving and greater mutual understanding in collaborative contexts (Richardson, Dale & Kirkham 2007; Shteynberg & Galinsky 2011), all of which are essential for successful co-creation.

Joint attention necessitates the ability to gain, maintain and shift attention (Mundy *et al.* 2007), and it is located at the intersection of various complex abilities that facilitate our cognitive, emotional and action-oriented connections with other individuals (Rauschnabel *et al.* 2024).

The emergence of positive interactions in the context of co-creation is a complex process that is influenced by several factors. Understanding these underlying principles and processes is crucial in order to identify the measurable indicators that characterize these factors. Furthermore, it is important to explore how design strategies can be leveraged to enhance these indicators in the context of co-creation activities. In this regard, our focus is on research related to joint attention, which examines how individuals share subjective experiences and construct meaning through their interactions. By drawing on theories and concepts from this field, we aim to shed light on the underlying mechanisms that enable positive effects among interactors and to identify design strategies that can be used to promote effective social interaction in the context of co-creation.

2.2. Joint attention in intersubjectivity

Intersubjectivity has been recognized as a critical factor in small-group research, as mutual understanding among group members promotes productivity, dependability and flexibility (Weick & Roberts 1993). Given that the cocreation process is characterized by “social interaction” (Edvardsson, Tronvoll & Gruber 2011; Finsterwalder & Kuppelwieser 2011) and “meaning-making in dialogue” (De Jaegher, Peräkylä & Stevanovic 2016), assessing intersubjectivity in the cocreation process can provide valuable insights into the experience of the process and the effectiveness of its outcomes. Several design researchers have emphasized the role of intersubjectivity in understanding the design process, including its role in facilitating mutual understanding (Ma 2013) and establishing spaces for equal dialogue among participants (Ho & Lee 2012). In addition, social cognition situates the study of intersubjectivity within interaction theory (IT) (Gallagher 2001, 2009), providing further support for research on intersubjectivity in the context of social interaction in co-creation activities. These findings are highly relevant to our study and contribute to our understanding of the importance of intersubjectivity in the cocreation process.

Intersubjectivity has been theorized to entail a shared intellectual and emotional statement of social competence in those that they serve (Djenar, Ewing & Howard 2017) and a shared involvement in a reciprocal exchange (Loots & Devisé 2003). Shared involvement refers to concurrently observing or concentrating on the same aspect of the environment (Moore *et al.* 2014). Reciprocal exchange refers to the active and reciprocal involvement of both interaction partners, whether physically in coordinated behavior patterns and vitality affects; existentially in the sharing of intentions, feelings and objects of joint attention or symbolically in the creation of linguistic and symbolic meaning (Rochat, Passos-Ferreira & Salem 2009). Beebe & Lachmann (2002) proposed a systems model of interaction that illustrates that verbalizable symbolic narratives (dialogues), unconscious gaze, facial expressions, eye contact, spatial orientation and body posture momentarily influence intersubjectivity during interactions.

Several studies have begun to describe the impact of intersubjectivity on social interactions and applied intersubjectivity as a measure (Loots, Devisé & Jacquet 2005; Damen *et al.* 2015). Scholars have assessed children's joint attention, joint focus, shared meaning-making (Trevvarthen & Aitken 2001; Göncü, Patt & Kouba 2002), emotional attunement and social coordination (Bateman, Campbell & Fonagy 2021) during group interactions. Garte (2015) provides a method to capture the interactive social competence development process. This assessment approach can provide new insights into how intersubjectivity supports social cognition and competence. Matsumae captured each participant's emotional fluctuation during the cocreation process to assess the degree of qualitative coincidence of fluctuation as a state of intersubjectivity being formed among them (Matsumae & Nagai 2018) (Table 1).

Based on previous studies, joint attention has been identified as a critical dimension in measuring intersubjectivity (Garte 2015). It serves as a link between primary intersubjectivity and secondary intersubjectivity (Trevvarthen 1998; Trevvarthen 2012). As joint attention plays a critical role in the development of intersubjectivity, it is measured as an indicator of intersubjectivity in the co-creation process. Interaction turns have also been used as a representation of intersubjectivity in previous research, such as in Loots' study. However, the description of interaction turns in representations are similar to the method used to measure joint attention. Therefore, joint attention is considered a reliable and measurable indicator of co-creation in our research.

Computer vision was utilized to recognize behaviors and emotions among a group of individuals in videos. By computing metrics related to interactions in nonverbal and implicit modes, which are typically outside of conscious awareness, we were able to measure joint attention among participants in various cocreation scenarios. The subsequent table presents indicators of joint attention among groups of individuals in the cocreation process, inspired by existing research (Table 2).

3. Methods and technical principles

In design workshops, the accurate identification of participants' objectives serves as the foundation for subsequent joint attention analysis. Current methods predominantly encompass questionnaire surveys, artificial scene observations and on-site

Table 1. List of important references			
References	Category	Representation	Index
Matsumae & Nagai (2018)	Emotional fluctuation	Qualitative coincidence of fluctuation; Quantitative comparison of each participant's emotional wave	Categorization of wave pattern observed; Numerically self-described emotional waves
Loots <i>et al.</i> (2005)	Interaction turns	Interaction turns that are not followed by interaction behavior that is part of a moment of intersubjectivity; Interaction turns that initiate or continue moments of physical/existential intersubjectivity; Interaction turns that initiate or continue moments of symbolic intersubjectivity.	Number of intersubjectivity moments; Length of intersubjectivity moments; Number of symbolic intersubjectivity moments
Garte (2015)	Behaviors occurred simultaneously or reciprocally in social, joint attention, and conflict dimensions.	<i>Social dimension:</i> Touching; Eye contact; Mutual positive emotion; Joint attention verbal expression; Reciprocal conversation. <i>Joint attention dimension:</i> Joint attention; Joint attention materials; Mutual focus. <i>Conflict dimension:</i> Violation of property; Violation of space; Mutual negative emotion.	The episode duration of peer interactions in an episode was recoded to a four-point scale.

interviews, among others. However, these approaches are prone to measurement inaccuracies and may introduce varying degrees of interference for participants. Notably, there is a paucity of quantitative research on individuals' engagement levels in design workshops within the design field, and the factors influencing the extent of collaborative participation in such settings remain ambiguous. Consequently, it is imperative to incorporate non-contact observation techniques to gather information on design workshop participants without compromising their experience and to quantitatively evaluate the specific indicators that influence their engagement levels. To accomplish this research objective, we employed deep learning-based visual measurement technology to precisely identify workshop participants during the study. Given the intricate nature of diverse design workshop scenarios, we further optimized and enhanced this methodology, which will be elaborated upon in this chapter.

Table 2. Reference source comparison table of important indicators

Indicators	Interpretation	Reference	Measurement in this study
Number of People	Essential characteristics of different cocreation activities	Jaegher, Paolo & Gallagher (2010)	Number of people in the scene
Intensity of movement	Amount of movement among people in the scene	Trevarthen (2012)	Number of activity tracks
Interaction occurred	Eye contact and facial communication.	Beebe (2005); Garte (2015)	Frequency of eye contact
Joint attention times	Number of people in a scene with mutual focus or simultaneous attention to the same material/task	Garte (2015)	Number of people in key areas
Joint attention duration	Length of time that people pay attention to the same material/task/activity.	Garte (2015); Loots <i>et al.</i> (2005)	Time of appearance of people in key areas
Emotional fluctuation	Number of common changes in emotions of people in a scene	Garte (2015); Matsumae & Nagai (2018)	Frequency of common facial expressions
Frequency of joint attention	Number of occurrences of joint attention.	Garte (2015); Markus <i>et al.</i> (2000)	Frequency of common attention
Social distance	Social distance between two people in a scene	Beebe & Lachmann (2002)	Mutual social distance

3.1. YOLO-TP: design workshop person target recognition network

To measure the joint attention among design workshop scenes using computer vision, we must accurately identify personnel targets in various complex scenes. The accuracy of personnel target recognition will affect the reliability of subsequent joint attention analysis. Therefore, this study uses extensive sample data combined with the improved YOLO-TP (YOLO Transformer Person) deep learning network to accurately identify and extract personnel objectives in the design workshop.

For the optimization and improvement of the network model, this study mainly uses YOLO v5s as the primary network structure, as shown in Figure 1. The basic architecture of the network mainly consists of backbone, neck structure, and head structures. There are many problems in the actual space environment of design creativity workshops, such as multiple personnel goals, complex environmental backgrounds, different lighting conditions and large data volumes (Wang *et al.* 2021; Wu *et al.* 2021; Sharma *et al.* 2022). This research adopts several optimization methods to improve the traditional YOLO v5s network structure (Aziz *et al.* 2020) to form a YOLO-TP target recognition network that is dedicated to personnel recognition in the design creativity workshop space.

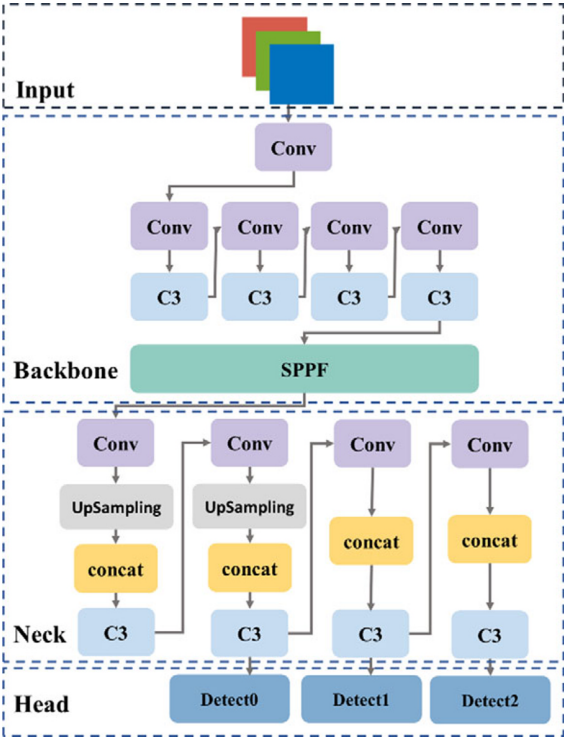


Figure 1. YOLO v5s network structure diagram.

3.2. Improvement and optimization of model structure

3.2.1. Introducing the TransformV2 module to optimize the backbone structure

First, this part optimizes the backbone part of the network. We replace CSP1_1 in Figure 1 with the latest Transformer structure. Compared with the traditional CSP structure (Zhang *et al.* 2022), the Transformer structure can better overcome bottlenecks when training numerous data and provide a more significant and stable detection model to stably identify the personnel target in the design workshop.

Here, we creatively introduce the structure layer of the new version of Transformer V2 to improve the problems of the traditional Transformer V1 structure layer. The improved part is reflected in the part marked in red in Figure 2. The traditional Transformer V1 structure layer faces three problems:

- (1) Increasing the visual model may create excellent training instability.
- (2) For many downstream tasks that require high resolution, there is no well-explored method to transfer the trained model with low resolution to a larger scale model.
- (3) In a complex background environment, a small number of pixels have considerable interference.

For the first problem of unstable training, we adopt the idea of the post norm, which is to move the layer norm layer in the Transformer block from the front of the attention layer to the back of the attention layer. The advantage of this idea is that after the attention is calculated, the output will be normalized to stabilize the output value.

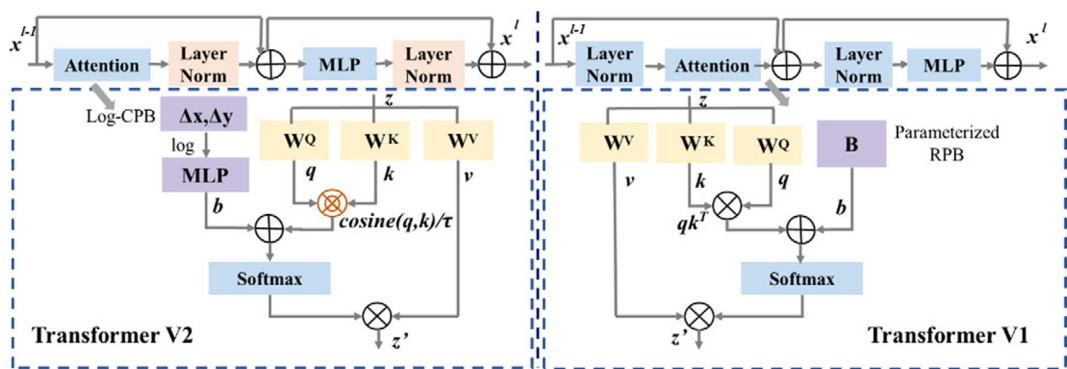


Figure 2. Improvement of V2 compared with V1.

For the second problem, the module uses log space continuous position offset technology to migrate the low-resolution pretraining model to the high-resolution pretraining model. The traditional processing step here adopts the continuous position offset method (Liu *et al.* 2022). The principle of this method is shown in the publicity (1), and the meta-network for relative coordinates is adopted.

$$B(\Delta x, \Delta y) = G(\Delta x, \Delta y) \tag{1}$$

In the above formula, G is a small network that generates offset parameters for any relative coordinates, so it can naturally migrate any variable window size. The log space continuous position offset technology mentioned here alleviates the problem that a large proportion of the relative coordinate range needs to be extrapolated when migrating across large windows. The publicity of this part of technology is shown in (2).

$$\begin{cases} \widehat{\Delta x} = \text{sign}(x) \cdot \log(1 + |\Delta x|) \\ \widehat{\Delta y} = \text{sign}(y) \cdot \log(1 + |\Delta y|) \end{cases} \tag{2}$$

Logarithmic operation is adopted here as the extrapolation ratio required for block resolution migration will be smaller using logarithmic space coordinates, which can lay the foundation for migration from a low-resolution pretraining model to a high-resolution pretraining model.

For the third problem, we discovered that in the V1 version of the self-attention calculation process, the pixel property of the pixel pair is calculated by the dot product of the query and key. However, in the workshop, a scene model with many data, the attention map of some modules and heads will be dominated by a small number of pixels. To alleviate this problem, the scaled cosine attention (SCA) method is applied in this part. The formula is shown in (3).

$$\text{Sim}(q_i, k_i) = \cos(q_i, k_i) / \tau + B_{ij} \tag{3}$$

3.2.2. Introducing the Asff module to optimize the head structure

Pyramid feature representation is a standard method to solve the problem of target scale change in target detection. However, the inconsistency between the two different feature scales is the main limitation of the target detector based on the feature pyramid. Here, we use a new and data-driven pyramid feature fusion

strategy that is referred to as adaptive spatial feature fusion (ASFF) in academia. This structure can effectively overcome the problem caused by different scales of data features due to the diversity of workshop scenarios, thus improving the scale invariance of features. Additionally, this structure does not require additional computing resources. In the actual design workshop, during the process of personnel recognition, the adaptive spatial feature fusion module effectively fuses pedestrian features in different scene backgrounds and further improves the robustness of the model. In the network, we replace the traditional Detect module with the Asff module discussed in this section. Figure 3 shows a schematic of the three-layer ASFF structure.

As shown in Figure 3, the module achieves feature fusion by setting weights α , β and γ . Note that the weight coefficient is automatically generated by a 1x1 convolution layer, softmax function and backpropagation. The principle of publicity is shown in (4).

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l} \quad (4)$$

where X is the input of each scale and y is the feature map output after scale fusion in space. We need to meet the following conditions: $\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1$,

$$\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l \in [0, 1] \text{ and } \alpha_{ij}^l = \frac{e^{\lambda_{ij}^l}}{e^{\lambda_{ij}^l} + e^{\lambda_{ij}^l} + e^{\lambda_{ij}^l}}.$$

3.2.3. Introducing SKAttention attention mechanism to optimize neck structure

It is undeniable that the introduction of an attention mechanism has a vital role in the accuracy of the personnel identification model. In the previous section, we reduced the computational overhead of some GPUs by optimizing Transformer V2. In this section, we consider that we can reasonably use the computational

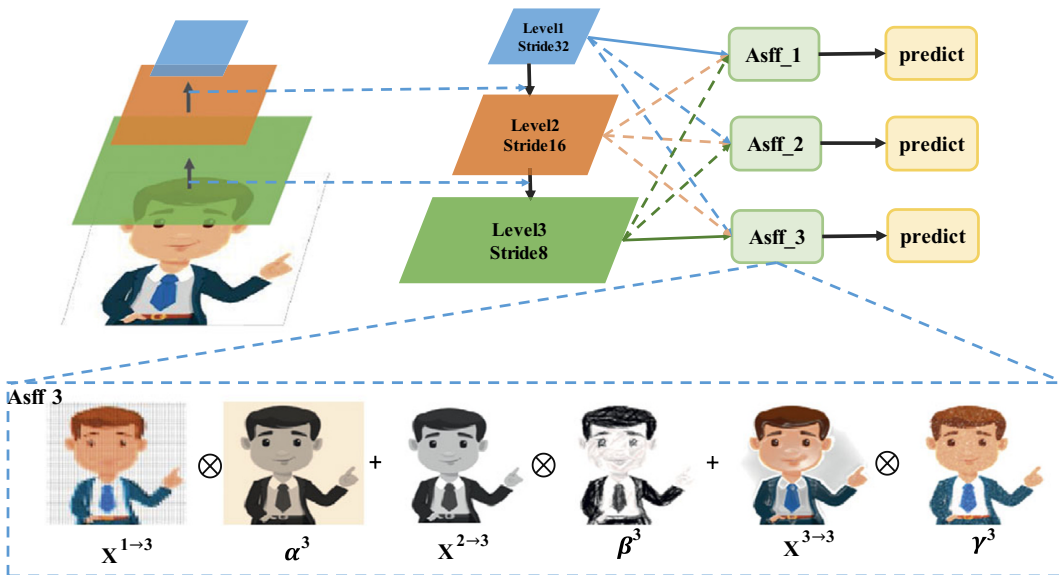


Figure 3. ASFF module structure diagram.

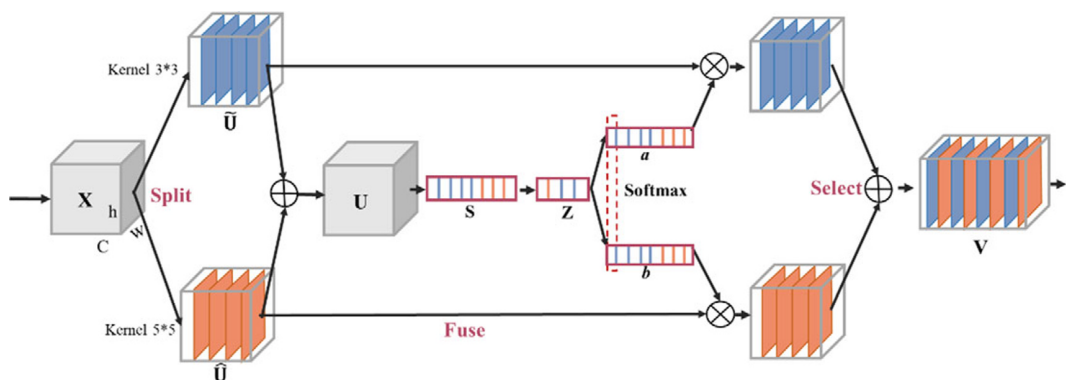


Figure 4. SKAttention module structure diagram.

overhead saved by the preamble part and improve accuracy by adding one-dimensional convolutions.

As shown in Figure 4, the primary processing process of this module is divided into the following three parts:

- (1) Split: complete convolution operation (group convolution) of input vector X with different kernel sizes. In particular, to further improve the efficiency, the traditional convolution of 5×5 is replaced by the cavity convolution with a division = 2 and a convolution core of 3×3 .
- (2) Fusion: After adding the two feature maps, the global average pooling operation is performed. The fully connected layer that reduces the dimension and then increases the dimension is a two-layer fully connected layer: the output of two attention coefficient vectors, a and b , where $a + b = 1$.
- (3) Select: Select uses two weight matrices, a and b , to weigh the previous two feature maps. There is an operation similar to feature selection between them.

3.2.4. Improved pedestrian target detection network: YOLO-TP

After the traditional YOLO v5s network is integrated with the above three improved modules, a new network, which is referred to as the YOLO-TP network, and especially suitable for high-precision target recognition of workshop staff. The network is optimized by inserting the Transformer V2 structure in the Backbone part, using the SKAttention attention mechanism in the neck part, and introducing the Asff structure in the head part. The modified YOLO-TP network achieves accurate recognition of pedestrian targets in different scenes; effectively overcomes the problem of inconsistent data characteristics caused by scene changes, changes in lighting conditions and changes in personnel actions and has high robustness. The improved network structure is shown in Figure 5.

3.3. Comparison of human target recognition test results under a complex background environment

For the training model, we selected the video stream data collected from (Tongji University) Design Institute and Shanghai NICE 2035 Creative Work Community.

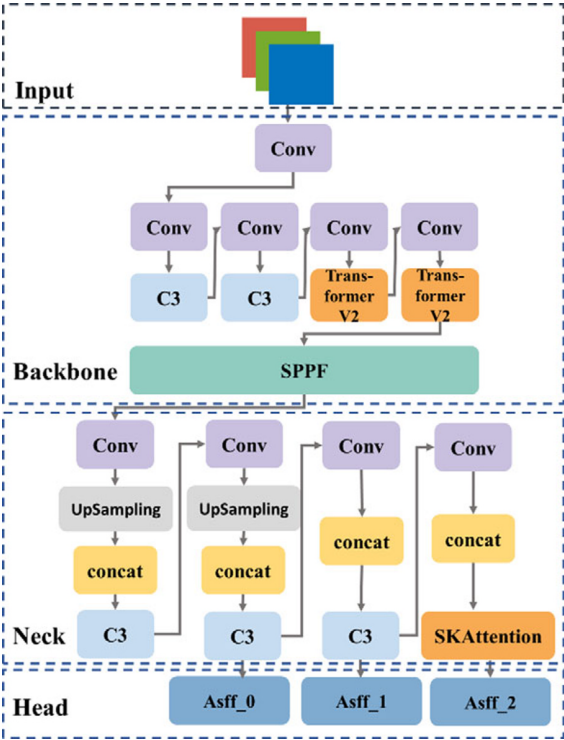


Figure 5. Schematic of the improved YOLO-TP.

After the video frame was drawn, we marked the personnel target samples to build a dataset. A total of 7,000 sample datasets were built in multiple scenes. The training and verification sets were divided according to a ratio of 9:1. The GPU used for training is an Nvidia GeoForce 2080 Ti, with Epoch = 10 training rounds. Figure 6 compares the effects of the traditional and improved networks on various evaluation indicators.

As shown in Figure 6(a), the detection accuracy of the traditional algorithm reaches 97.39%, and the detection accuracy of the improved algorithm proposed in this paper reaches 98.47%, an increase of 1.08%. Although our method can achieve a recognition accuracy of over 98%, in the process of processing a large amount of data, there are still 1% to 2% errors in discrimination when facing large changes in environmental lighting and fast movement of actions. This is a very normal phenomenon in the automatic processing of large amounts of data, so we believe that this does not affect subsequent analysis. Figure 6(b) shows that the improved YOLO-TP model in this paper is superior to the traditional algorithm in terms of mAP50 and mAP50:90, with increases of 0.11% and 1.07%, respectively. According to Figure 6(c), for the F1-Score index, the optimal index of the traditional algorithm reaches 97.59%. However, the highest calculation in this paper is 97.86%, an increase of 0.27%. As shown in Figure 6(d), compared with the traditional YOLO v5 model, the YOLO-TP model proposed in this paper reduces the target frame loss and target detection loss. The smaller the former is, the more accurate the mark position is when the frame mark is marked, and the smaller the latter is, the more

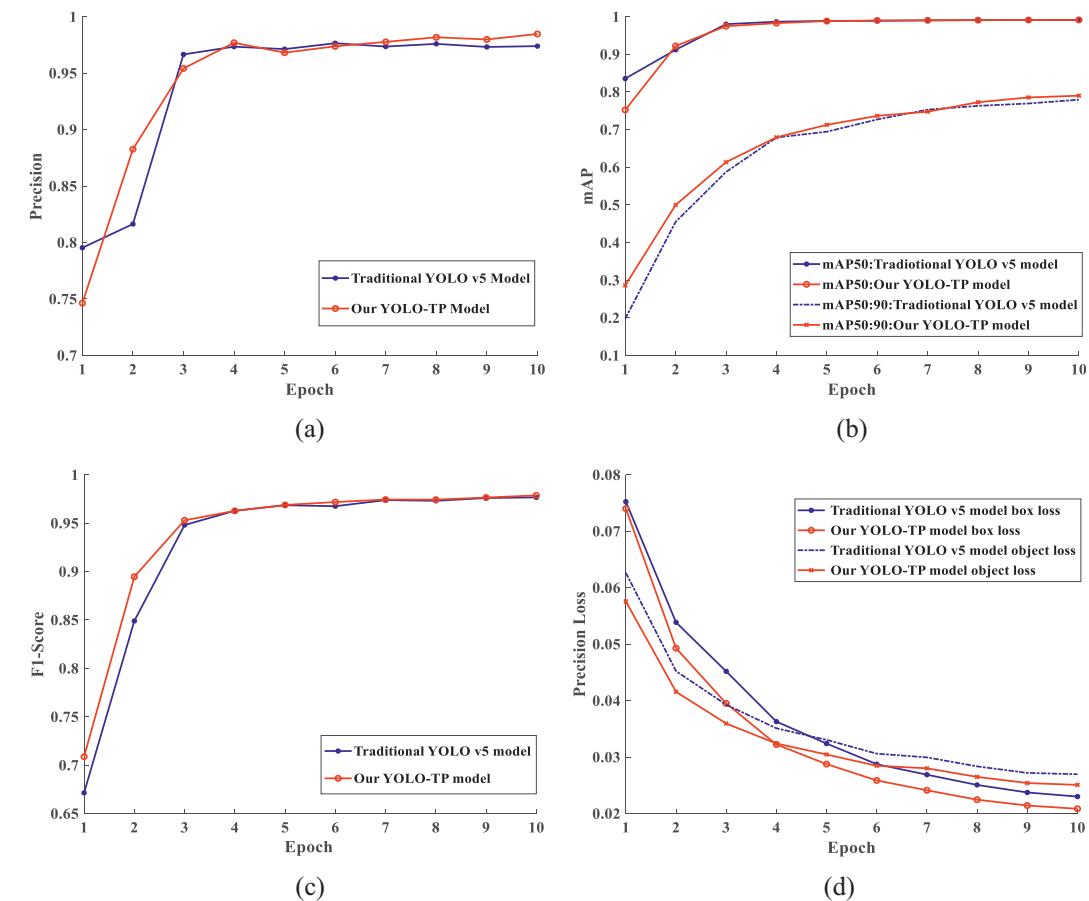


Figure 6. Comparison of the effect between the improved YOLO-TP network structure and the traditional network: (a) comparison of personnel target detection accuracy, (b) comparison of mAP50 and mAP50:90 accuracy indicators, (c) comparison of F1-Score model evaluation indicators and (d) comparison of training accuracy loss.

accurate the detection of personnel targets. The improved algorithm in this paper has improved in terms of both detection accuracy and detection effect.

To assess the performance of the proposed model in a design workshop setting, we utilized real-world data collected from design workshops at (Tongji University). The results of the test are depicted in Figure 7.

Figure 7 illustrates that the conventional YOLO v5 model exhibits some errors in the actual personnel target detection within the workshop, leading to a degree of omission. The red marks in the figure highlight the targets that were missed during detection. In contrast, the YOLO-TP model proposed in this paper accurately identifies personnel targets across various workshop scenarios, establishing a solid foundation for high-precision target recognition in subsequent personnel statistical analyses.

In summary, the YOLO-TP network proposed in this paper demonstrates a strong capability for accurately identifying participants (detecting participant targets) within various complex design workshop scenarios, significantly improving



Figure 7. Test results of real design workshop environment.

upon baseline models as shown in our validation (Section 3.3). Building upon this reliable foundation of participant detection and localization, we proceed in this study to analyze specific behavioral indicators related to joint attention exhibited during these workshop activities. The eight specific indicators utilized for this quantitative assessment, which are derived from the YOLO-TP output, are formally introduced and detailed in Section 4.2 and the Appendix. As an innovative endeavor, we apply this non-contact visual measurement technique to the study of design activities, leveraging deep learning to precisely capture participant presence and interaction patterns in diverse contexts. This provides a robust methodological basis for future analyses of design activities. To further demonstrate the application of these methods, the subsequent sections elaborate on experiments conducted using data from real-world design settings.

3.4. The measurement significance of the passive measurement method for joint attention

Joint attention constitutes a significant area of research within the field of interactive scene design. The precise measurement of joint attention in the milieu of design creativity presents an enduring challenge. This paper proposes the employment of the YOLO-TP network as a primary method for the quantification of joint attention, achieved through the efficient identification of individuals within a scene. Moreover, this network forms the foundation for the eight quantifiable indicators detailed in the Appendix of this manuscript, thereby underscoring its considerable import.

In the present study, the examination of human activities within design scenes and the quantification of joint attention are executed through video and image data gleaned from visual sensors. This paper contrasts conventional methodologies

dependent on scenario analysis and questionnaire statistics, adopting instead a non-contact measurement approach via video and image data. This method thereby addresses the inherent limitations of survey-based approaches in accurately quantifying the involvement of individuals.

Moreover, the YOLO personnel analysis network is trained on specialized datasets, particularly tailored for individuals involved in design scenarios. This network is especially adapted for the automated detection and statistical analysis of the number of participants during design workshops and exchanges. It has been optimized to quantify joint attention accurately in creative design scenarios.

In summation, the network and methodologies proposed within this paper facilitate the capturing of data from diverse perspectives, thereby enabling their application across a range of scenarios. This approach offers quantitative processing support for data collated from different contexts. It is recognized that the observation and quantification method presented herein is not the only approach to analyze joint attention in the design field. Nonetheless, it provides a potent means for quantification, thus enabling the transition from qualitative to quantitative analysis in design research.

4. Data and experiments

4.1. Data

The data presented in this study were collected from design studios within the NICE 2035 living labs at the College of Design and Innovation, Tongji University over a period of 24 months, between 2021 and 2023. A total of 296 students were sampled from 24 design workshops, with an average workshop size of 12.33 members. Of these participants, 46% were male and 54% were female, with an average age of 27.7 years. The sampled students had backgrounds in Design, Technology, and Business, primarily from (Tongji University). All participants had normal communication and activity abilities, no language barriers or psychological issues and ensured normal interaction among workshop participants. We installed video cameras in each workshop scenario to capture and record the participants' design activities and processes. [Figure 8](#) depicts an example of the camera fields of view after YOLO-TP processing.

[Table 3](#) shows the basic data of each design studio, among which we have screened a total of 40 pieces of video data to better measure joint attention in design workshops. The duration of each piece of data is controlled at 10 minutes. This selection was guided primarily by the need to balance analytical depth with the significant computational cost associated with processing long video sequences using our deep learning pipeline. Furthermore, clips were prioritized that exhibited high levels of participant interaction relevant to co-creation and joint attention. The algorithm designed in this paper is used to count eight key indicators and then to conduct a subsequent principal component analysis.

4.2. Experimental processing

4.2.1. Image measurement

In the camera imaging system used in this article, there are four coordinate systems: the world coordinate system, the camera coordinate system, the image

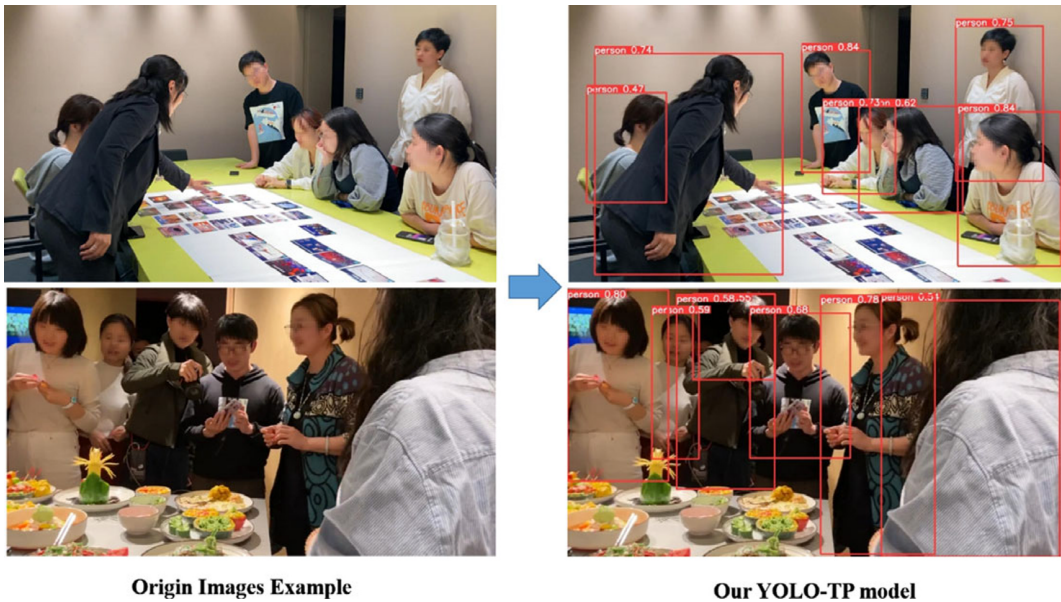


Figure 8. Example of tagged information picture after YOLO-TP processing.

coordinate system and the pixel coordinate system. There is a rigorous mathematical correlation between these four coordinate systems, as shown in Figure 9. The image measurement method in this study strictly proves that the pixel distance in the image is positively correlated with the spatial distance.

The transformation relationship between the above four coordinate systems is:

$$Z \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{dX} & -\frac{\cot\theta}{dX} & u_0 \\ 0 & \frac{1}{dY\sin\theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U \\ V \\ W \\ 1 \end{pmatrix} \quad (5)$$

Where, (U, V, W) is the physical coordinate of a point in the world coordinate system, (u, v) is the pixel coordinate corresponding to the point in the pixel

coordinate system and Z is the scale factor. The $\begin{pmatrix} \frac{1}{dX} & -\frac{\cot\theta}{dX} & u_0 \\ 0 & \frac{1}{dY\sin\theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix}$ to the

right of the equal sign represents an affine transformation matrix, and the $\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$ represents a projection transformation matrix. These two matrices

constitute the camera's internal parameter matrix. Where, f represents the image distance, dX, dY represents the physical length of a pixel in the X and Y directions on the camera's photosensitive plate (i.e., how many millimeters a pixel is on the photosensitive plate), and u_0, v_0 represents the coordinates of the center of the camera's photosensitive plate in the pixel coordinate system, respectively, θ

Table 3. List of video data of the design workshop					
Serial number	Brief description of the scenario	Number of conversations	Total duration of video recording	Number of included scene clips	Number of selected clips
1	Design workshop activities for insect ecology	20+	3000s	28	2
2	Research on visual design and visual communication design	30+	7000s	40	3
3	Carry out food creativity workshop	35+	4000s	65	4
4	Design workshop to study body and dance movements	40+	2500s	41	3
5	Design workshop for children’s extracurricular activities	30+	9000s	107	7
6	Design workshop for children’s singing activities	20+	3500s	16	1
7	Design workshop for children’s performance activities	25+	3700s	19	1
8	Design workshop for human behavior and food tasting research	15+	2700s	18	1
9	Workshop on Children’s Bird Cognition and Creative Research and Design	11+	1500s	6	1
10	Environmental protection awareness design workshop	15+	1600s	19	1
11	Design workshop for garden party activities	18+	2300s	17	1
12	Design workshop on cat adoption	15+	2500s	20	1
13	Mainly carries out the design workshop of agricultural-related courses	25+	3200s	16	1
14	Design workshop on climate change	30+	3500s	20	2
15	Garden Design Workshop	25+	3600s	19	1
16	Creative Workshop of Fashion Design	27+	4000s	21	1
17	Fashion Decorative Design Creative Workshop	28+	4500s	22	2
18	Creative Workshop on Children’s Activity Design	30+	3500s	20	1
19	Game Design Workshop	25+	3200s	17	1

Continued

Table 3. Continued

Serial number	Brief description of the scenario	Number of conversations	Total duration of video recording	Number of included scene clips	Number of selected clips
20	Workshop on Farm Garden Design	30+	5000s	17	1
21	Incense Brick Workshop	15+	1500s	7	1
22	Creative design workshop for making fallen leaves	18+	1800s	9	1
23	Workshop on design creativity related to plant environmental protection	25+	3000s	18	1
24	Creative workshops related to green food	30+	3600s	18	1

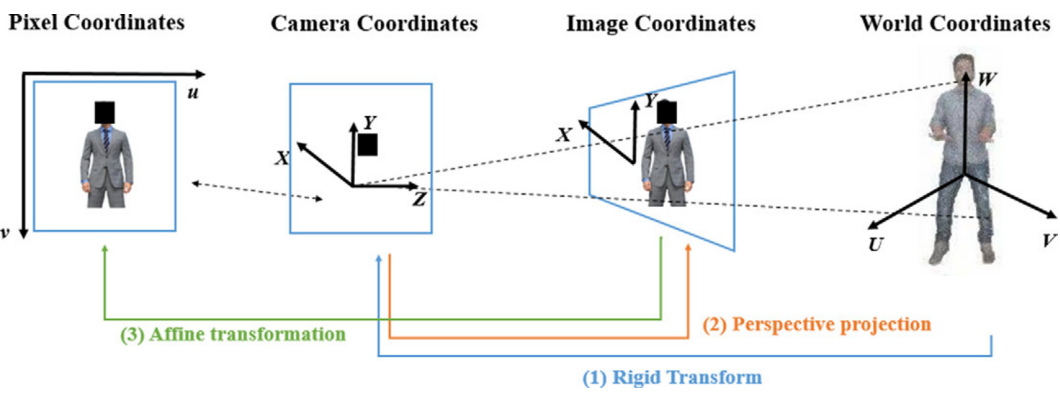


Figure 9. Relationship between image coordinates and real coordinates.

Indicates the angle between the horizontal and vertical edges of the photosensitive plate (90 ° indicates no error). $\begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U \\ V \\ W \\ 1 \end{pmatrix}$ represents a rigid body transformation matrix that constitutes the camera’s external parameter matrix. Where R represents the rotation matrix and T represents the translation vector.

When we get an image and perform recognition, the distance between the two parts obtained is a certain pixel value, but how many meters do these pixels correspond to in the real world? This requires using the camera calibration results to convert pixel coordinates to physical coordinates to calculate the distance. For camera calibration, we directly use the Zhang Zhengyou calibration method to calibrate the camera (Lu, Liu & Guo 2016). The purpose of calibration is to obtain the camera’s internal and external parameter matrices, as shown in Formula (5), which is used to convert the distance on the map to the actual distance. Through the strict proof mentioned above, we found that the distance on the image and the

actual distance present a strict positive correlation. Therefore, in our subsequent analysis, we directly use the relative distance of pixels on the image to measure the distance. These values can strictly reflect the distance situation in the real world.

4.2.2. Workshop scenario analysis

The first step in analyzing the videotapes consisted of separating peer interactions from the continuous co-creation process on the recording. The inclusion of episodes was contingent on two criteria: length and participation. During the creation of the measures, it was revealed that interactions lasting less than 60 s could not be consistently recorded. In addition, the theoretical concept of joint attention as focusing on shared activity necessitated that contacts be properly sustained to identify any shared activity. A study of the tapes confirmed that a minimum of 1 minute of contact was necessary to achieve these criteria. The second criterion for analysis inclusion was that episodes must be defined by participant continuity. The interacting people must remain the only participants throughout the episode. The previously established episode would be considered concluded if a new person joined. This approach represents the notion that joint attention among interacting partners emerges due to their particular social dynamic.

To better analyze the pedestrian activities in the design workshop, we quantitatively extracted the following eight quantifiable indicators from the video stream data collected by the visual sensor: number of people in the scene, number of activity tracks, number of people in key areas, time of appearance of people in key areas, frequency of eye contact, frequency of common facial expressions, mutual social distance and frequency of common attention. For different indicators, we designed and specified relevant algorithms. In summary, for the above eight indicators, this paper provides a statistical method using computer vision and strives to obtain the relationship among the joint attention in the design workshop from the data. The overall statistical flow chart of the above indicators is shown in Figure 10.

For each indicator, we have designed a special calculation method flow, which is included in the Appendix of this article. Additionally, during the process of data analysis, our two authors, as data analysis engineers, examined 40 analysis video clips, reviewed the correctness of the algorithm’s output on 8 indicators and evaluated the reliability of the algorithm based on this, providing a highly credible data source for subsequent principal component analysis (PCA) data analysis, as shown in Table 4. We paid special attention to the indicators of “eye contact frequency,” “common facial expression frequency,” and “common attention frequency.” This involves comparing the key event time points identified by the algorithm with their direct visual interpretation of the interactions in the corresponding video segments. These qualitative evaluations confirm that the automatically generated indicators are consistent with their observations and common patterns in these co creation activities, providing confidence in the effectiveness of the indicators. Finally, we also believe that introducing high-level experts for interpretation is a necessary operation and one of the improvement points for our future research.

As shown in Table 4, each column represents different observation indicators, and each row represents different scenarios. The values in the table represent the

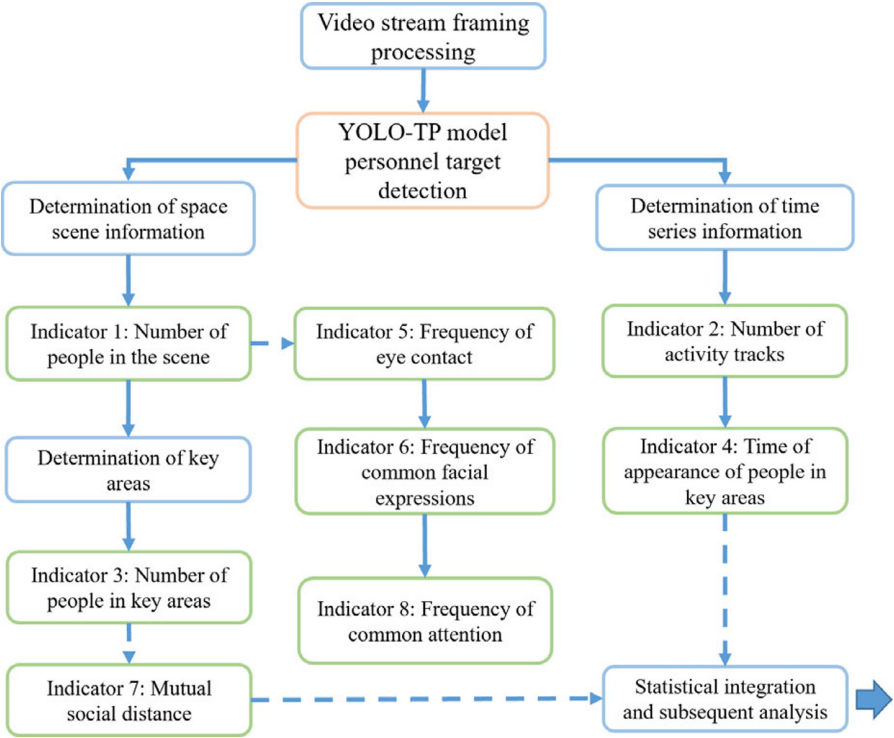


Figure 10. Key indicator statistics flow chart.

accuracy of the program’s calculated results and manual interpretation results. The closer the value is to 1, the closer the accuracy of the program’s inferred indicator value is to manual interpretation. The specific results can be found in the [Appendix](#) of this article.

5. Results and analysis

5.1. Results related to the three dimensions of joint attention

To reduce the dimensionality of the behavioral measures derived from the video analysis and identify underlying latent constructs of joint attention, we performed principal component analysis (PCA). The analysis was conducted on the dataset comprising the calculated values for the eight quantitative indicators (number of people in the scene, number of activity tracks, number of people in key areas, time of appearance of people in key areas, frequency of eye contact, frequency of common facial expressions, mutual social distance and frequency of common attention) obtained from the 40 video clips, as detailed in [Section 4.2](#) and the [Appendix](#). PCA is a multivariate statistical method used to transform multiple variables linearly, reducing the number of variables. In this study, we used PCA to identify comprehensive indicators for joint attention measurement. The KMO test value of 0.646 indicates that the information contained in each indicator has more common factors. The significance of Bartlett’s spherical test is less than 0.01 ($p = 0.000$), indicating that the indicators are independent and that the data are

Table 4. Reliability evaluation table for 8 indicators (compared with manual interpretation)								
Test data number	Indicator 1	Indicator 2	Indicator 3	Indicator 4	Indicator 5	Indicator 6	Indicator 7	Indicator 8
1	1.00	1.00	1.00	0.94	0.96	0.99	0.92	0.99
2	1.00	0.99	1.00	0.91	0.99	0.99	0.99	1.00
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	0.99	1.00	1.00	0.96	1.00	1.00	0.95	1.00
6	1.00	1.00	1.00	0.96	0.98	0.99	0.95	0.99
7	0.97	1.00	1.00	0.99	0.97	0.97	0.94	0.99
8	1.00	1.00	1.00	0.96	1.00	0.99	0.92	0.99
9	1.00	1.00	1.00	0.97	1.00	1.00	0.99	1.00
10	1.00	1.00	1.00	0.92	0.99	1.00	0.97	1.00
11	1.00	1.00	1.00	0.99	1.00	0.98	0.95	0.99
12	1.00	1.00	1.00	0.97	0.99	0.99	0.98	0.98
13	0.99	1.00	1.00	1.00	0.99	0.99	0.95	0.99
14	1.00	1.00	1.00	0.99	1.00	0.99	0.99	0.99
15	1.00	1.00	1.00	0.94	0.99	0.99	0.98	0.99
16	1.00	1.00	1.00	0.94	1.00	0.99	0.99	1.00
17	1.00	1.00	1.00	0.97	0.99	1.00	0.97	0.98
18	1.00	1.00	1.00	0.96	0.98	1.00	0.99	0.99
19	0.99	1.00	1.00	0.99	1.00	0.99	0.99	0.99
20	0.99	1.00	1.00	0.99	0.97	0.99	0.96	0.99
21	0.99	1.00	1.00	0.99	0.99	0.98	0.96	0.99
22	1.00	0.99	1.00	0.92	0.99	0.99	0.99	0.99
23	1.00	1.00	1.00	0.98	0.99	1.00	0.96	1.00
24	1.00	1.00	1.00	0.99	1.00	1.00	0.99	0.99
25	0.99	0.99	1.00	0.93	0.99	1.00	0.97	1.00
26	0.99	0.99	1.00	0.93	0.99	0.99	0.98	1.00
27	0.99	0.99	1.00	0.98	0.99	0.99	0.96	0.99
28	1.00	0.99	1.00	0.99	0.99	0.99	0.98	1.00
29	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.99
30	0.99	1.00	0.97	0.95	0.96	0.96	0.99	0.97
31	1.00	1.00	1.00	0.97	0.99	0.99	0.96	0.99
32	1.00	1.00	1.00	0.99	1.00	0.99	1.00	0.99
33	1.00	0.99	1.00	0.94	1.00	0.99	0.96	0.96
34	0.99	0.99	1.00	0.99	0.99	0.99	0.99	1.00
35	0.99	1.00	1.00	0.97	0.98	0.99	1.00	0.99
36	0.99	1.00	1.00	0.99	0.99	0.99	0.96	0.99

Continued

Table 4. Continued

Test data number	Indicator 1	Indicator 2	Indicator 3	Indicator 4	Indicator 5	Indicator 6	Indicator 7	Indicator 8
37	0.99	1.00	1.00	0.96	0.99	0.99	0.97	1.00
38	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.99
39	0.99	0.99	1.00	0.95	0.99	0.97	0.99	1.00
40	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00

Table 5. Factor loadings of the generalized principal component

Items	Loadings		
	PCA 1	PCA 2	PCA 3
Explained Variance	31.72%	27.18%	15.19%
Number of people in the scene	0.442	−0.619	−0.146
Number of activity tracks	0.138	0.645	0.550
Number of people in key areas	0.468	−0.487	0.566
Time of appearance of people in key areas	0.315	−0.519	0.573
Frequency of eye contact	0.775	0.385	−0.209
Frequency of common facial expressions	0.824	0.313	0.006
Mutual social distance	−0.521	0.610	0.431
Frequency of common attention	0.674	0.500	−0.111

suitable for principal component analysis. Three principal components were extracted from the PCA, each with eigenvalues greater than 1. The factor loadings of the generalized principal component (shown in Table 5) indicate that the frequency of eye contact, common facial expressions and common attention have the highest loadings on the first principal component, which we define as the empathic sharing dimension (PCA 1). Mutual social distance, the number of people in the scene, and the number of activity tracks have the highest loadings on the second principal component, which we define as the social context dimension (PCA 2). The number of people in key areas and the time of appearance of people in key areas have the highest loadings on the third principal component, which we define as the key area dimension (PCA 3).

Empathic sharing dimension is the key dimension in Joint Attention, which includes the frequency of eye contact, common facial expressions and common attention. Previous research has shown that empathic sharing is associated with higher levels of social connection and cooperation between participants in a design process (Swan & Riley 2015). Designers can use this dimension to understand participants’ emotional states and needs and facilitate interactions between them, leading to better design outcomes. Social context dimension is another important

Table 6. Linear combination coefficients and weights

Items	PCA 1	PCA 2	PCA 3	Comprehensive Score	Weights
Characteristic root	2.538	2.174	1.215		
Number of people in the scene	0.2776	0.4195	0.1326	0.2999	11.47%
Number of activity tracks	0.0868	0.4374	0.4988	0.2999	11.47%
Number of people in key areas	0.2936	0.3306	0.5138	0.3523	13.48%
Time of appearance of people in key areas	0.1975	0.3523	0.5200	0.3204	12.26%
Frequency of eye contact	0.4866	0.2610	0.1899	0.3430	13.12%
Frequency of common facial expressions	0.5175	0.2121	0.0054	0.3005	11.49%
Mutual social distance	0.3268	0.4136	0.3911	0.3718	14.22%
Frequency of common attention	0.4229	0.3394	0.1006	0.3262	12.48%

dimension of Joint Attention, including Mutual social distance, the number of people in the scene and the number of activity tracks. These factors influence the level of social interaction and attention among participants, which can affect design outcomes. For example, a higher number of people in the scene may increase the complexity of the design process and require more careful coordination among participants. Thus, designers should consider the social context dimension when designing collaborative environments and activities. Key area dimension, which includes the number of people in key areas and the time of appearance of people in key areas, is also crucial in Joint Attention. Key areas are places that participants focus on during the design process. Designers can use this dimension to identify the most important areas of focus for participants and facilitate collaboration and interaction in these areas.

5.2. Results related to the items of joint attention

The weightings assigned to each of the eight indicators in the PCA (shown in Table 6) reflect their respective contributions to the variability within the dataset. Specifically, the weights represent the proportion of the total variance in the original eight indicators that can be accounted for by each resulting principal component. In this case, the weights for the eight indicators range from 11.47% to 14.22%, indicating that each indicator contributes to the variability within the dataset, albeit to varying degrees. For example, the indicator “Mutual social distance” has the highest weighting at 14.22%, indicating that it is the most influential factor in shaping the underlying patterns of Joint Attention in the dataset. Conversely, the indicators “Number of people in the scene” and “Number of activity tracks” have the lowest weightings at 11.47%, suggesting that they contribute the least to the overall variability of the dataset. These weightings provide a useful understanding of the relative importance of each indicator in shaping Joint Attention, which can inform the design of collaborative environments and activities.

5.3. The Relevance for design researchers

In design research, the complex interactions between joint attention dimensions and their respective weightings present significant implications for co-creation processes. This discussion explores the consequences of each dimension and the weightings of indicators, grounded in academic literature and analytical reflection, while clarifying how our findings contribute to and extend the broader discourse on co-creation rather than merely collaboration.

5.3.1. *Empathic sharing dimension: a nexus of emotional resonance in co-creation*

The empathic sharing dimension, encapsulating the frequency of eye contact, common facial expressions and common attention, serves as a linchpin in the co-creation process. This dimension directly addresses the intersubjective aspects of co-creation, which identified as critical for meaningful collaborative design. Co-creation, as conceptualized by Sanders & Stappers (2008), goes beyond mere collaboration to encompass “collective creativity applied across the whole span of a design process,” with joint attention being a fundamental mechanism through which this collective creativity manifests. Our findings extend the current understanding of co-creation by demonstrating that joint attention, specifically through empathic sharing, quantifiably contributes to more effective co-creation outcomes. Cash & Maier (2016) demonstrated in their Design Science research that gestural communication plays a crucial role in establishing shared understanding during collaborative design activities. Our work builds upon theirs by providing a computational framework for measuring these previously qualitative aspects of co-design interaction.

The computational measurement of empathic sharing provides design researchers with unprecedented insights into the quality of co-creation processes. Rather than relying on subjective assessments or post-hoc analyses, our framework allows for real-time evaluation and enhancement of co-creation activities. As Calvo, Sclater & Smith (2021) note, “achieving collaboration through co-design is challenging as people need to understand each other, and develop trust and rapport.” Our measurement framework specifically addresses this challenge by providing objective metrics for these hard-to-quantify aspects of intersubjective engagement. Furthermore, our research contributes to resolving what Trischler *et al.* (2018) identify as a key challenge in co-design: balancing diverse participant contributions while maintaining cohesive progress toward shared goals. The empathic sharing dimension provides a metric for assessing this balance, offering design researchers a tool for orchestrating more effective co-creation sessions where collaborative empathy can be fostered and measured.

5.3.2. *Social context and key area dimensions: the infrastructure of co-creation*

The social context dimension, encompassing mutual social distance, the number of people present, and activity tracks, reveals the spatial and social infrastructure of co-creation environments. This dimension builds upon Menold, Jablowski & Simpson’s (2017) “Prototype for X (PFX)” framework by extending its principles to the social dynamics of co-creation, demonstrating how physical arrangements and movement patterns directly influence the quality of collaborative design. Our research contributes to the co-creation literature by establishing that physical

proximity and movement patterns significantly impact the quality of co-creative design. Cash, Dekoninck & Ahmed-Kristensen (2017) previously identified in their Design Science research that the spatial arrangement of design teams influences communication patterns, but our study extends this understanding by providing a measurable framework for optimizing spatial configurations in co-creation settings. The implications of our findings on the Social Context Dimension align with what Protzen & Harris (2010) term the “ecology of design spaces,” wherein the physical environment serves not merely as a backdrop but as an active agent in shaping design discourse. By quantifying the impact of spatial arrangements on joint attention, our research provides design facilitators with evidence-based guidelines for configuring co-creation environments to maximize collaborative potential. Furthermore, our work contributes to what Carlile (2002) identifies as the challenge of “knowledge boundaries” in cross-functional team interactions. By measuring how spatial proximity influences joint attention across participants from diverse backgrounds, our framework offers insights into how physical space can be leveraged to overcome disciplinary boundaries that often hinder effective co-creation.

The key area dimension, highlighting the number of individuals in pivotal areas and their temporal presence, offers critical insights into attention allocation during co-creation processes. This dimension builds upon and extends Cash *et al.*'s (2021) research on the role of prototypes in facilitating shared understanding, suggesting that key areas—whether physical spaces or conceptual domains—serve as “boundary objects” that facilitate cross-disciplinary exchange in co-design activities. Our research advances the co-creation discourse by quantifying how attention to specific physical or conceptual spaces correlates with collaborative outcomes. Erichsen *et al.* (2021) in their Design Science research explored how physical prototypes capture design knowledge, but our findings provide a measurable framework for identifying and leveraging these focal points in co-creation settings. The key area dimension intersects with what Kleinsmann *et al.* (2017) term “collaborative design loops,” wherein participants iterate between individual exploration and collective synthesis. Our measurement framework provides a means to assess the effectiveness of these loops by tracking how participants converge around and engage with key areas during the co-creation process. The practical implications of this dimension extend to what Sanders & Stappers (2014) describe as the “front end” of co-design, where the problem space is still being explored and defined. By identifying which key areas attract joint attention during early co-creation phases, facilitators can more effectively structure subsequent activities to build upon emergent shared understanding.

5.3.3. Weightings of indicators: the quantifiable metrics of co-creation

The weightings, ranging from 11.47% to 14.22%, offer a nuanced, data-driven understanding of each indicator's contribution to co-creation effectiveness. “Mutual social distance” emerges as a dominant force (14.22%), which extends research on proximity in design collaboration by quantifying its precise contribution to co-creation outcomes. Our research significantly contributes to the co-creation literature by providing a weighted framework that allows design researchers to prioritize specific aspects of co-creation environments based on their measurable impact. Previously, as noted by Dorst & Cross (2001), such

prioritization was largely intuitive or based on qualitative assessments. Our findings transform this approach by offering a quantitative basis for decision-making in co-creation facilitation. The weighted indicator framework connects directly to what Björgvinsson, Ehn & Hillgren (2012) describe as “infrastructuring” in co-design—the process of establishing conditions that enable productive participation. Our research provides empirical evidence for which aspects of this infrastructuring most significantly impact the quality of joint attention and, by extension, co-creation outcomes. In synthesizing these reflections, we advance the understanding of co-creation by moving beyond treating it as merely a collaborative activity to recognizing it as a complex, measurable phenomenon with specific dimensions that can be optimized. Co-creation is not just a designerly collaboration to involve people but a formal research practice with a general model that produces new academic knowledge. Our weighted framework provides the empirical foundation for such a formal model of co-creation effectiveness.

6. Conclusion

6.1. Practical applications for co-creation process optimization

The joint attention measurement framework developed in this study offers several concrete pathways for optimizing co-creation processes in real-world design settings. First, real-time monitoring capabilities enable facilitators to identify when joint attention is declining during co-creation sessions and implement targeted interventions to re-engage participants (Trischler *et al.* 2018). For example, when the Empathic Sharing Dimension indicators show decreased eye contact frequency and common facial expressions, facilitators can introduce structured interaction activities or modify the physical arrangement to enhance face-to-face engagement (Cash & Maier 2016).

Second, the weighted indicator framework provides evidence-based guidance for designing optimal co-creation environments. Given that mutual social distance emerged as the most influential factor (14.22% weighting), design practitioners can prioritize spatial configurations that promote appropriate proximity levels (Cash *et al.* 2017). This might involve adjusting table arrangements, seating configurations, or workspace layouts to facilitate the social distances that correlate with enhanced joint attention. Similarly, the importance of the key area dimension suggests that co-creation spaces should include clearly defined focal points that naturally draw participant attention and provide shared reference points for collaborative work (Menold *et al.* 2017).

Third, the three-dimensional framework enables diagnostic assessment of co-creation sessions, allowing researchers and practitioners to identify specific areas for improvement (Sanders & Stappers 2014). Sessions scoring low on the social context dimension might benefit from interventions targeting group size optimization or movement pattern enhancement, while sessions with poor key area dimension scores might require better definition of focal work areas or improved tool accessibility (Christensen & Ball 2016). This diagnostic capability transforms co-creation facilitation from an intuitive practice to an evidence-based discipline (Calvo *et al.* 2021).

Fourth, the framework supports the development of adaptive co-creation protocols that respond to real-time joint attention measurements. Advanced

implementations could incorporate automated feedback systems that adjust lighting, spatial arrangements, or activity structures based on ongoing joint attention assessments, creating responsive environments that continuously optimize collaborative conditions (Oertzen *et al.* 2018). This approach aligns with emerging trends in design research toward more responsive and data-driven facilitation methods (Bjögvinsson *et al.* 2012).

6.2. Key findings and theoretical implications

In this study, we endeavored to elucidate a machine learning-oriented paradigm for ascertaining joint attention within the co-creation milieu. Our research directly addresses the growing need within design research for objective, quantifiable measures of co-creation effectiveness, moving beyond the traditional reliance on subjective assessments that has limited the field's advancement (Kleinsmann *et al.* 2017). Leveraging the finesse of computer vision, we have developed a novel methodological approach that not only captures joint attention data from multifaceted co-creation scenarios but also circumvents potential pitfalls inherent in traditional sampling surveys. This methodological contribution responds directly to calls within the co-creation literature for more robust measurement frameworks (Sanders & Stappers 2014).

The empirical revelations from our study proffer three significant contributions to the co-creation discourse. First, our Principal Component Analysis has identified and quantified three cardinal dimensions of joint attention in co-creation – empathic sharing, social context and key area – providing a structured framework for understanding the previously amorphous concept of intersubjectivity in collaborative design. This framework extends beyond mere collaboration to address the distinctive characteristics of co-creation as a specific form of collaborative activity where shared attention and intersubjectivity are paramount. Second, we have demonstrated that joint attention, as a specific manifestation of intersubjectivity, can be objectively measured and correlated with co-creation effectiveness. This finding bridges the gap between the experiential aspects of design collaboration and measurable outcomes, offering a quantitative basis for evaluating and enhancing co-creation processes. As Nguyen & Mougenot (2022) note in their systematic review of empirical studies on multidisciplinary design collaboration, shared understanding is consistently identified as crucial across diverse collaborative contexts, yet methodological approaches for measuring it remain inconsistent. Our computational framework for measuring joint attention provides a standardized approach to evaluating this critical aspect of co-creation. Third, our weighted indicator framework provides design researchers and practitioners with a practical tool for optimizing co-creation environments and activities. As Cash & Maier (2016) demonstrated in their Design Science research on gestural communication in design, non-verbal interactions significantly impact shared understanding development. Our framework extends this line of inquiry by providing quantitative metrics for measuring these interactions and their effects on joint attention in co-creation settings.

Central to our discourse is the pivotal role of refined deep learning frameworks in objectively measuring design processes. Our approach aligns with recent advances in applying computational methods to design research, such as Kent *et al.*'s (2022) network analysis approach to prototyping and Erichsen *et al.*'s (2021)

digital capture of physical prototypes. These approaches collectively represent a paradigm shift toward more objective, data-driven assessment of design activities. Our research clarifies the distinction between general collaboration and co-creation by focusing specifically on the intersubjective dimensions that make co-creation a unique form of collaborative activity. While collaboration broadly encompasses coordinated effort toward shared goals, co-creation, as defined by Sanders & Stappers (2008) and expanded by Oertzen *et al.* (2018), distinctively involves “joint value creation among multiple actors through resource integration.” Our quantitative framework for measuring joint attention provides a means to distinguish co-creation from other collaborative activities by quantifying the degree to which participants achieve shared focus, understanding, and engagement – the hallmarks of true co-creation. This distinction is particularly important in light of Castañer & Oliveira’s (2020) systematic review clarifying the differences between collaboration, coordination and cooperation in organizational contexts. While these terms are often used interchangeably, co-creation represents a specific form of collaborative activity characterized by mutual focus, shared understanding and joint value creation. Our measurement framework provides empirical support for this distinction by quantifying the degree to which participants achieve joint attention during co-creation activities.

However, the current study focused on developing and validating the methodology for measuring joint attention components. Consequently, we did not investigate the direct correlation between these automatically extracted measures (either the eight indicators or the three PCA dimensions) and specific outcomes of the co-design process, such as participant experience ratings or the quality/quantity of design outputs. Exploring these relationships is a critical avenue for future research to determine the practical utility and predictive validity of our joint attention metrics for assessing and potentially enhancing co-creation effectiveness. Future studies should aim to collect both the automated behavioral data and corresponding process/outcome measures.

6.3. Limitations and future research directions

Looking beyond the current horizon, we are poised to integrate speech emotion recognition with computer vision, aligning with calls for multimodal approaches to design research. This future direction aims to unravel the nuanced behavioral attributes of designers in co-creation settings via a comprehensive analytical lens. The integration of multiple data streams will enable a more holistic assessment of co-creation dynamics, potentially revealing interaction patterns that remain invisible when examined through a single modality. However, a critical introspection warrants the acknowledgment of certain limitations. At the outset, our focus on human-centric parameters to gauge joint attention in intersubjectivity provides but a glimpse into the vast expanse of co-creation analytics. As Christensen & Ball (2016) noted in their Design Studies research on creative analogy use in heterogeneous design teams, the cognitive aspects of design collaboration extend beyond observable behaviors. Future research should explore how joint attention correlates with cognitive processes underlying co-creation, potentially through mixed-methods approaches combining our computational framework with qualitative assessment of participants’ thought processes.

Furthermore, the inherent reliance on visual sensor deployment, especially in expansive co-creation environments, amplifies both fiscal and computational burdens. This limitation echoes concerns about the scalability of technological approaches to design research. As Hansen & Özkil (2020) demonstrated in their longitudinal case study of prototyping strategies, design processes unfold across multiple spaces and timeframes, presenting challenges for comprehensive data capture. Future research should explore more efficient sensor systems and sampling strategies to reduce the resource intensiveness of our approach while maintaining measurement validity. A pivotal caveat lies in the modest sample size, underscoring the preliminary nature of our findings and beckoning corroborative studies with augmented sample sizes. While our study demonstrates the feasibility and potential value of computational approaches to measuring joint attention, broader deployment across diverse co-creation contexts is needed to establish the generalizability of our findings. As Erichsen *et al.* (2021) note in their work on prototyping data capture, design research methods must balance specificity with generalizability to maximize their value to the field.

In conclusion, our research advances the co-creation discourse by providing a quantitative framework for measuring and enhancing the intersubjective dimensions that distinguish co-creation from general collaboration. By objectively measuring joint attention, we offer design researchers a powerful tool for understanding and optimizing co-creation processes, addressing a significant gap in the current literature. As co-creation continues to gain prominence across diverse domains – from product development to service design to policy formulation – our framework provides a foundation for evidence-based enhancement of co-creation processes across these contexts. This contribution represents not merely an incremental advancement in co-creation methodology but a fundamental shift toward more objective, data-driven approaches to understanding and facilitating this complex form of collaborative design activity.

Acknowledgments

We would like to thank NICE2035 for providing the experimental facilities essential for this study. We also appreciate the support and feedback from colleagues and participants, which greatly contributed to the success of this work. The data used in the paper can be obtained by contacting the corresponding author or the following link (<https://www.wjx.cn/vm/PQfR5gm.aspx#>). Due to the involvement of personal image data of experimental participants, the researchers in this article have the right to review the qualifications of the data requesters during the data sharing process.

Financial support

This research was funded by the Chinese Ministry of Education Humanities and Social Sciences Research Youth Fund Project [23YJC760101] and MOE (China) Research Innovation Team on “Design-Driven High-Quality Urban Development [20242717]”.

Competing interest

The authors declare none.

References

- Andersen, P. V. K. & Mosleh, W. S. 2021 Conflicts in co-design: Engaging with tangible artefacts in multi-stakeholder collaboration. *CoDesign* 17 (4), 473–492; doi:[10.1080/15710882.2020.1740279](https://doi.org/10.1080/15710882.2020.1740279).
- Aziz, L., Salam, M. S. B. H., Sheikh, U. U. & Ayub, S. 2020 Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review. *IEEE Access* 8; doi:[10.1109/ACCESS.2020.3021508](https://doi.org/10.1109/ACCESS.2020.3021508).
- Bateman, A., Campbell, C. & Fonagy, P. 2021 Rupture and repair in mentalization-based group psychotherapy. *International Journal of Group Psychotherapy* 71 (2), 371–392; doi:[10.1080/00207284.2020.1847655](https://doi.org/10.1080/00207284.2020.1847655).
- Beebe, B. (2005). Faces-in-relation: Forms of intersubjectivity in an adult treatment of early trauma. In *Forms of Intersubjectivity in Infant Research and Adult Treatment*. Other Press.
- Beebe, B. & Lachmann, F. M. 2002 *Infant Research and Adult Treatment: Co-Constructing Interactions*. The Analytic Press/Taylor & Francis Group, pp. xv, 272–272.
- Behoora, I. & Tucker, C. S. 2015 Machine learning classification of design team members' body language patterns for real time emotional state detection. *Design Studies* 39, 100–127; doi:[10.1016/j.destud.2015.04.003](https://doi.org/10.1016/j.destud.2015.04.003).
- Bjögvinsson, E., Ehn, P. & Hillgren, P. A. 2012 Design things and design thinking: Contemporary participatory design challenges. *Design Issues* 28 (3), 101–116; doi:[10.1162/DESI_a_00165](https://doi.org/10.1162/DESI_a_00165).
- Button, G. & Sharrock, W. 1996 Project work: The organisation of collaborative design and development in software engineering. *Computer Supported Cooperative Work (CSCW)*, 5 (4), 369–386; doi:[10.1007/BF00136711](https://doi.org/10.1007/BF00136711).
- Calvo, M., Sclater, M. & Smith, P. 2021 Creating spaces for collaboration in community co-design. *International Journal of Art & Design Education* 40 (1), 232–250; doi:[10.1111/jade.12349](https://doi.org/10.1111/jade.12349).
- Carlile, P. R. 2002 A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science* 13 (4), 442–455; doi:[10.1287/orsc.13.4.442.2953](https://doi.org/10.1287/orsc.13.4.442.2953).
- Carpenter, M., Nagell, K. & Tomasello, M. 1998 Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development* 63 (4), i–174. doi:[10.2307/1166214](https://doi.org/10.2307/1166214)
- Cash, P., Dekoninck, E. A. & Ahmed-Kristensen, S. 2017 Supporting the development of shared understanding in distributed design teams. *Journal of Engineering Design* 28 (3), 147–170; doi:[10.1080/09544828.2016.1274719](https://doi.org/10.1080/09544828.2016.1274719).
- Cash, P., Dekoninck, E. & Ahmed-Kristensen, S. 2020 Work with the beat: How dynamic patterns in team processes affect shared understanding. *Design Studies* 69, 100943; doi:[10.1016/j.destud.2020.04.003](https://doi.org/10.1016/j.destud.2020.04.003).
- Cash, P., Hicks, B., Culley, S. & Salustri, F. 2021 Designer behaviour and activity: An industrial observation method. In *Proceedings of the 18th International Conference on Engineering Design (ICED 11)*, Impacting Society through Engineering Design, Vol. 2: Design Theory and Research Methodology, pp. 151–162. The Design Society. doi:[10.32920/ryerson.14639703](https://doi.org/10.32920/ryerson.14639703).

- Cash, P. & Maier, A. 2016 Prototyping with your hands: The many roles of gesture in the communication of design concepts. *Journal of Engineering Design* 27 (1–3), 118–145; doi:[10.1080/09544828.2015.1126702](https://doi.org/10.1080/09544828.2015.1126702).
- Castañer, X. & Oliveira, N. 2020 Collaboration, coordination, and cooperation among organizations: Establishing the distinctive meanings of these terms through a systematic literature review. *Journal of Management* 46 (6), 965–1001; doi:[10.1177/0149206320901565](https://doi.org/10.1177/0149206320901565).
- Christensen, B. T. & Abildgaard, S. J. J. 2021 Kinds of ‘moving’ in designing with sticky notes. *Design Studies* 76, 101036; doi:[10.1016/j.destud.2021.101036](https://doi.org/10.1016/j.destud.2021.101036).
- Christensen, B. T. & Ball, L. J. 2016 Creative analogy use in a heterogeneous design team: The pervasive role of background domain knowledge. *Design Studies* 46, 38–58; doi:[10.1016/j.destud.2016.07.004](https://doi.org/10.1016/j.destud.2016.07.004).
- Cooper, L. 2023 Constructing accounts of decision-making in sustainable design: A discursive psychology analysis. *Design Studies* 84, 101158; doi:[10.1016/j.destud.2022.101158](https://doi.org/10.1016/j.destud.2022.101158).
- Damen, S., Janssen, M. J., Ruijsenaars, W. A. J. J. M. & Schuengel, C. 2015 Communication between children with deafness, blindness and Deafblindness and their social partners: An Intersubjective developmental perspective. *International Journal of Disability, Development and Education* 62 (2), 215–243; doi:[10.1080/1034912X.2014.998177](https://doi.org/10.1080/1034912X.2014.998177).
- De Jaegher, H., Peräkylä, A. & Stevanovic, M. 2016 The co-creation of meaningful action: Bridging enaction and interactional sociology. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1693), 20150378; doi:[10.1098/rstb.2015.0378](https://doi.org/10.1098/rstb.2015.0378).
- Devos, T. & Loopmans, M. 2022 Stimulating embodied intersubjectivities: Two participatory experiments in Antwerp north, Belgium. *CoDesign* 18 (3), 322–339; doi:[10.1080/15710882.2021.1894179](https://doi.org/10.1080/15710882.2021.1894179).
- Djenar, D., Ewing, M. & Howard, M. 2017 *Style and Intersubjectivity in Youth Interaction*. De Gruyter Mouton; doi:[10.1515/9781614516439](https://doi.org/10.1515/9781614516439).
- Dorst, K. & Cross, N. 2001 Creativity in the design process: Co-evolution of problem–solution. *Design Studies* 22 (5), 425–437; doi:[10.1016/S0142-694X\(01\)00009-6](https://doi.org/10.1016/S0142-694X(01)00009-6).
- Edvardsson, B., Tronvoll, B. & Gruber, T. 2011 Expanding understanding of service exchange and value co-creation: A social construction approach. *Journal of the Academy of Marketing Science* 39 (2), 327–339; doi:[10.1007/s11747-010-0200-y](https://doi.org/10.1007/s11747-010-0200-y).
- Erichsen, J. A., Sjöman, H., Steinert, M. & Welo, T. 2021 Protobooth: Gathering and analyzing data on prototyping in early-stage engineering design projects by digitally capturing physical prototypes. *AI EDAM* 35 (1), 65–80; doi:[10.1017/S0890060420000414](https://doi.org/10.1017/S0890060420000414).
- Falck-Ytter, T., Kleberg, J. L., Portugal, A. M. & Thorup, E. 2023 Social attention: Developmental foundations and relevance for autism Spectrum disorder. *Biological Psychiatry* 94 (1), 8–17; doi:[10.1016/j.biopsych.2022.09.035](https://doi.org/10.1016/j.biopsych.2022.09.035).
- Finsterwalder, J. & Kuppelwieser, V. G. 2011 Co-creation by engaging beyond oneself: The influence of task contribution on perceived customer-to-customer social interaction during a group service encounter. *Journal of Strategic Marketing* 19 (7), 607–618; doi:[10.1080/0965254X.2011.599494](https://doi.org/10.1080/0965254X.2011.599494).
- Frow, P., Nenonen, S., Payne, A. & Storbacka, K. 2015 Managing co-creation design: A strategic approach to innovation. *British Journal of Management* 26 (3), 463–483; doi:[10.1111/1467-8551.12087](https://doi.org/10.1111/1467-8551.12087).
- Fuchs, T. & De Jaegher, H. 2009 Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences* 8, 465–486; doi:[10.1007/s11097-009-9136-4](https://doi.org/10.1007/s11097-009-9136-4).

- Gallagher, S. 2001 The practice of mind. Theory, simulation or primary interaction? *Journal of Consciousness Studies* 8 (5–6), 83–108.
- Gallagher, S. 2009 Two problems of Intersubjectivity. *Journal of Consciousness Studies* 16 (6–8), 6–8.
- Garte, R. R. 2015 Intersubjectivity as a measure of social competence among children attending head start: Assessing the measure's validity and relation to context. *International Journal of Early Childhood* 47 (1), 189–207; doi:[10.1007/s13158-014-0129-2](https://doi.org/10.1007/s13158-014-0129-2).
- Göncü, A., Patt, M. B. & Kouba, E. 2002 Understanding young children's pretend play in context. In P. K. Smith & C. H. Hart (Eds.), *Blackwell Handbook of Childhood Social Development*, pp. 418–437. Blackwell Publishers.
- Grönroos, C. & Voima, P. 2013 Critical service logic: Making sense of value creation and co-creation. *Journal of the Academy of Marketing Science* 41 (2), 133–150; doi:[10.1007/s11747-012-0308-3](https://doi.org/10.1007/s11747-012-0308-3).
- Guo, Y., Liu, Z., Luo, H., Pu, H. & Tan, J. 2022 Multi-person multi-camera tracking for live stream videos based on improved motion model and matching cascade. *Neurocomputing* 492; doi:[10.1016/j.neucom.2021.12.047](https://doi.org/10.1016/j.neucom.2021.12.047).
- Hansen, C. A. & Özkil, A. G. 2020 From idea to production: A retrospective and longitudinal case study of prototypes and prototyping strategies. *Journal of Mechanical Design* 142 (3), 031115; doi:[10.1115/1.4045385](https://doi.org/10.1115/1.4045385).
- Heiss, L. & Kokshagina, O. 2021 Tactile co-design tools for complex interdisciplinary problem exploration in healthcare settings. *Design Studies* 75, 101030; doi:[10.1016/j.destud.2021.101030](https://doi.org/10.1016/j.destud.2021.101030).
- Ho, D. K. L. & Lee, Y. C. 2012 The quality of design participation: Intersubjectivity in design practice. *International Journal of Design* 6 (1), 71–83.
- Howard, C. D. & Bevins, L. 2022 'The blue dot thing': A discourse analysis of learner interlanguage in instructional design. *CoDesign* 18 (2), 186–207; doi:[10.1080/15710882.2020.1789173](https://doi.org/10.1080/15710882.2020.1789173).
- Ind, N. & Coates, N. 2013 The meanings of co-creation. *European Business Review*, 25 (1), Article 1; doi:[10.1108/09555341311287754](https://doi.org/10.1108/09555341311287754)
- Jaegher, H. D., Paolo, E. D. & Gallagher, S. 2010 Can social interaction constitute social cognition? *Trends in Cognitive Sciences* 14 (10), 441–447; doi:[10.1016/j.tics.2010.06.009](https://doi.org/10.1016/j.tics.2010.06.009).
- Kassner, M., Patera, W. & Bulling, A. 2014 Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1151–1160. ACM. doi:[10.1145/2638728.2641695](https://doi.org/10.1145/2638728.2641695)
- Kent, L., Gopsill, J., Giunta, L., Goudswaard, M., Snider, C. & Hicks, B. 2022 Prototyping through the lens of network analysis and visualisation. *Proceedings of the Design Society* 2, 743–752; doi:[10.1017/pds.2022.76](https://doi.org/10.1017/pds.2022.76).
- Kleinsmann, M., Valkenburg, R. & Sluijs, J. 2017 Capturing the value of design thinking in different innovation practices. *International Journal of Design* 11 (2), 25–40.
- Le Bail, C., Baker, M. & Détienne, F. 2022 Values and argumentation in collaborative design. *CoDesign* 18 (2), 165–185; doi:[10.1080/15710882.2020.1782437](https://doi.org/10.1080/15710882.2020.1782437).
- Liang, H., Wu, T., Zhang, Q. & Zhou, H. 2022 Non-maximum suppression performs later in multi-object tracking. *Applied Sciences (Switzerland)* 12 (7); doi:[10.3390/app12073334](https://doi.org/10.3390/app12073334).
- Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y. ... & Guo, B. 2022 Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12009–12019. IEEE. doi:[10.48550/arXiv.2111.09883](https://doi.org/10.48550/arXiv.2111.09883)

- Lloyd, P. & Oak, A. 2018 Cracking open co-creation: Categories, stories, and value tension in a collaborative design process. *Design Studies* 57, 93–111; doi:[10.1016/j.destud.2018.02.003](https://doi.org/10.1016/j.destud.2018.02.003).
- Loots, G. & Devisé, I. 2003 An Intersubjective developmental perspective on interactions between deaf and hearing mothers and their deaf infants. *American Annals of the Deaf* 148 (4), 295–307.
- Loots, G., Devisé, I. & Jacquet, W. 2005 The impact of visual communication on the Intersubjective development of early parent–child interaction with 18- to 24-month-old deaf toddlers. *The Journal of Deaf Studies and Deaf Education* 10 (4), 357–375; doi: [10.1093/deafed/eni036](https://doi.org/10.1093/deafed/eni036).
- Lu, P., Liu, Q. & Guo, J. 2016 Camera calibration implementation based on zhang zhengyou plane method. *Lecture Notes in Electrical Engineering* 359; doi:[10.1007/978-3-662-48386-2_4](https://doi.org/10.1007/978-3-662-48386-2_4).
- Ma, J. (2013). *A phenomenological inquiry into the experience of “having a design concept.”* <https://theses.lib.polyu.edu.hk/handle/200/7012>
- Markus, J., Mundy, P., Morales, M., Delgado, C. E. F. & Yale, M. 2000 Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development* 9 (3), 302–315; doi:[10.1111/1467-9507.00127](https://doi.org/10.1111/1467-9507.00127).
- Matsumae, A. & Nagai, Y. 2018 The function of co-creation in dynamic mechanism of intersubjectivity formation among individuals: 15th international DESIGN conference, DESIGN 2018. In *15th International Design Conference, DESIGN 2018*, pp. 1925–1936. [10.21278/idc.2018.0141](https://doi.org/10.21278/idc.2018.0141)
- Menold, J., Jablokow, K. & Simpson, T. 2017 Prototype for X (PFX): A holistic framework for structuring prototyping methods to support engineering design. *Design Studies* 50, 70–112; doi:[10.1016/j.destud.2017.03.001](https://doi.org/10.1016/j.destud.2017.03.001).
- Moore, C., Dunham, P. J. & Dunham, P. 2014 *Joint Attention: Its Origins and Role in Development*. Psychology Press; doi:[10.4324/9781315806617](https://doi.org/10.4324/9781315806617).
- Mosleh, W. S. & Larsen, H. 2021 Exploring the complexity of participation. *CoDesign* 17 (4), 454–472; doi:[10.1080/15710882.2020.1789172](https://doi.org/10.1080/15710882.2020.1789172).
- Mundy, P., Sullivan, L. & Mastergeorge, A. M. 2007 A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research* 1 (1), 2–21; doi:[10.1002/aur.4](https://doi.org/10.1002/aur.4).
- Nguyen, M. & Mougenot, C. 2022 A systematic review of empirical studies on multidisciplinary design collaboration: Findings, methods, and challenges. *Design Studies* 81, 101120; doi:[10.1016/j.destud.2022.101120](https://doi.org/10.1016/j.destud.2022.101120).
- Oertzen, A. S., Odekerken-Schröder, G., Brax, S. A. & Mager, B. 2018 Co-creating services —Conceptual clarification, forms and outcomes. *Journal of Service Management* 29 (4), 641–679; doi:[10.1108/JOSM-03-2017-0067](https://doi.org/10.1108/JOSM-03-2017-0067).
- Ozdemir, S., Akin-Bulbul, I. & Yildiz, E. 2024 Visual attention in joint attention bids: A comparison between toddlers with autism Spectrum disorder and typically developing toddlers. *Journal of Autism and Developmental Disorders*; doi:[10.1007/s10803-023-06224-y](https://doi.org/10.1007/s10803-023-06224-y).
- Park, J. S., O’Brien, J., Cai, C. J., Morris, M. R., Liang, P. & Bernstein, M. S. 2023 Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–22. ACM. doi:[10.1145/3586183.3606763](https://doi.org/10.1145/3586183.3606763).
- Prahalad, C. K. & Ramaswamy, V. 2004 Co-creation experiences: The next practice in value creation. *Journal of Interactive Marketing* 18 (3), Article 3; doi:[10.1002/dir.20015](https://doi.org/10.1002/dir.20015).
- Protzen, J. P. & Harris, D. J. 2010 *The Universe of Design: Horst Rittel’s Theories of Design and Planning*. Routledge. [10.4324/9780203851586](https://doi.org/10.4324/9780203851586).

- Racine, T. P. & Carpendale, J. I. 2007 The role of shared practice in joint attention. *British Journal of Developmental Psychology* 25 (1), 3–25; doi:[10.1348/026151006X119756](https://doi.org/10.1348/026151006X119756).
- Ramaswamy, V. & Ozcan, K. 2018 What is co-creation? An interactional creation framework and its implications for value creation. *Journal of Business Research* 84, 196–205; doi:[10.1016/j.jbusres.2017.11.027](https://doi.org/10.1016/j.jbusres.2017.11.027).
- Ranjan, K. R. & Read, S. 2016 Value co-creation: Concept and measurement. *Journal of the Academy of Marketing Science* 44 (3), 290–315; doi:[10.1007/s11747-014-0397-2](https://doi.org/10.1007/s11747-014-0397-2).
- Rauschnabel, P. A., Felix, R., Heller, J. & Hinsch, C. 2024 The 4C framework: Towards a holistic understanding of consumer engagement with augmented reality. *Computers in Human Behavior* 154, 108105; doi:[10.1016/j.chb.2023.108105](https://doi.org/10.1016/j.chb.2023.108105).
- Richardson, D. C., Dale, R. & Kirkham, N. Z. 2007 The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science* 18 (5), 407–413; doi:[10.1111/j.1467-9280.2007.01914.x](https://doi.org/10.1111/j.1467-9280.2007.01914.x).
- Rochat, P., Passos-Ferreira, C. & Salem, P. 2009 Three levels of intersubjectivity in early development. In A. Carassa, F. Morganti, & G. Riva (Eds.), *Enacting Intersubjectivity: Paving the Way for a Dialogue between Cognitive Science, Social Cognition and Neuroscience*, pp. 173–190. Larioprint.
- Sanders, E. B.-N. & Stappers, P. J. 2008 Co-creation and the new landscapes of design. *CoDesign* 4 (1), 5–18; doi:[10.1080/15710880701875068](https://doi.org/10.1080/15710880701875068).
- Sanders, E. B. N. & Stappers, P. J. 2014 Probes, toolkits and prototypes: Three approaches to making in codesigning. *CoDesign* 10 (1), 5–14; doi:[10.1080/15710882.2014.888183](https://doi.org/10.1080/15710882.2014.888183).
- Sani-Bozkurt, S. & Bozkus-Genc, G. 2023 Social robots for joint attention development in autism Spectrum disorder: A systematic review. *International Journal of Disability, Development and Education* 70 (5), 625–643; doi:[10.1080/1034912X.2021.1905153](https://doi.org/10.1080/1034912X.2021.1905153).
- Shalley, C. E., Zhou, J. & Oldham, G. R. 2004 The effects of personal and contextual characteristics on creativity: Where should we go from here? *Journal of Management* 30 (6), 933–958; doi:[10.1016/j.jm.2004.06.007](https://doi.org/10.1016/j.jm.2004.06.007).
- Sharma, T., Debaque, B., Duclos, N., Chehri, A., Kinder, B. & Fortier, P. 2022 Deep learning-based object detection and scene perception under bad weather conditions. *Electronics (Switzerland)* 11 (4); doi:[10.3390/electronics11040563](https://doi.org/10.3390/electronics11040563).
- Shteynberg, G. & Galinsky, A. D. 2011 Implicit coordination: Sharing goals with similar others intensifies goal pursuit. *Journal of Experimental Social Psychology* 47 (6), 1291–1294; doi:[10.1016/j.jesp.2011.04.012](https://doi.org/10.1016/j.jesp.2011.04.012).
- So, W.-C., Cheng, C.-H., Law, W.-W., Wong, T., Lee, C., Kwok, F.-Y., Lee, S.-H. & Lam, K.-Y. 2023 Robot dramas may improve joint attention of Chinese-speaking low-functioning children with autism: Stepped wedge trials. *Disability and Rehabilitation: Assistive Technology* 18 (2), 195–204; doi:[10.1080/17483107.2020.1841836](https://doi.org/10.1080/17483107.2020.1841836).
- Spagnoli, A., Guardigli, E., Orso, V., Varotto, A. & Gamberini, L. 2014 Measuring user acceptance of wearable symbiotic devices: Validation study across application scenarios. In *Symbiotic Interaction* (ed. G. Jacucci, L. Gamberini, J. Freeman & A. Spagnoli), pp. 87–98. Springer International Publishing. [10.1007/978-3-319-13500-7_7](https://doi.org/10.1007/978-3-319-13500-7_7).
- Swan, P. & Riley, P. 2015 Social connection: Empathy and mentalization for teachers. *Pastoral Care in Education* 33 (4), 220–233; doi:[10.1080/02643944.2015.1094120](https://doi.org/10.1080/02643944.2015.1094120).
- Tomasello, M. 1995 Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development*, pp. 103–130. Lawrence Erlbaum Associates Inc.
- Trevarthen, C. 1998 The concept and foundations of infant intersubjectivity. *Intersubjective Communication and Emotion in Early Ontogeny* 15, 46.

- Trevarthen, C.** 2012 Embodied human Intersubjectivity: Imaginative agency, to share meaning. *Cognitive Semiotics* 4 (1), 6–56; doi:[10.1515/cogsem.2012.4.1.6](https://doi.org/10.1515/cogsem.2012.4.1.6).
- Trevarthen, C. & Aitken, K. J.** 2001 Infant intersubjectivity: Research, theory, and clinical applications. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 42 (1), 3–48; doi:[10.1017/S0021963001006552](https://doi.org/10.1017/S0021963001006552).
- Trischler, J., Pervan, S. J., Kelly, S. J. & Scott, D. R.** 2018 The value of codesign: The effect of customer involvement in service design teams. *Journal of Service Research* 21 (1), 75–100; doi:[10.1177/1094670517714060](https://doi.org/10.1177/1094670517714060).
- Wang, Y., Kim, Y. & Lin, H.** 2024 Social viewing of news and political participation: The mediating roles of information acquisition, self-expression, and partisan identity. *Computers in Human Behavior* 154, 108158; doi:[10.1016/j.chb.2024.108158](https://doi.org/10.1016/j.chb.2024.108158).
- Wang, Z., Wu, Y., Yang, L., Thirunavukarasu, A., Evison, C. & Zhao, Y.** 2021 Fast personal protective equipment detection for real construction sites using deep learning approaches. *Sensors* 21 (10); doi:[10.3390/s21103478](https://doi.org/10.3390/s21103478).
- Weick, K. E. & Roberts, K. H.** 1993 Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly* 38 (3), 357–381; doi:[10.2307/2393372](https://doi.org/10.2307/2393372).
- Wu, W., Liu, H., Li, L., Long, Y., Wang, X., Wang, Z., Li, J. & Chang, Y.** 2021 Application of local fully convolutional neural network combined with YOLO v5 algorithm in small target detection of remote sensing image. *PLoS One* 16 (10); doi:[10.1371/journal.pone.0259283](https://doi.org/10.1371/journal.pone.0259283).
- Yi, Y. & Gong, T.** 2013 Customer value co-creation behavior: Scale development and validation. *Journal of Business Research* 66 (9), 1279–1284; doi:[10.1016/j.jbusres.2012.02.026](https://doi.org/10.1016/j.jbusres.2012.02.026).
- Zhang, X., Wan, T., Wu, Z. & Du, B.** 2022 Real-time detector design for small targets based on a bi-channel feature fusion mechanism. *Applied Intelligence* 52 (3); doi:[10.1007/s10489-021-02545-6](https://doi.org/10.1007/s10489-021-02545-6).
- Zhao, D., Zhang, J. & Fu, J.** 2020 Research on real-time statistics of people flow based on deep Learning. *Chinese Journal of Sensors and Actuators* 33 (8); doi:[10.3969/j.issn.1004-1699.2020.08.013](https://doi.org/10.3969/j.issn.1004-1699.2020.08.013).

A. Appendix

The specific measurement methods for the eight indicators mentioned in this article are shown in the Appendix. For each indicator, we provide a detailed calculation process to facilitate the digital statistics of the information collected from visual sensors and achieve the transformation of design methods from qualitative analysis to quantitative analysis.

(1) Number of people in the scene: This paper uses the target statistics algorithm based on the YOLO-TP deep learning network. After accurately recognizing pedestrian targets using the YOLO-TP deep learning network proposed in this paper, the number of pedestrian targets is counted according to time. The algorithm flow is shown in [Table A1](#).

(2) Number of activity tracks: This indicator is closely related to the circulation in the design workshop space and the activities of participants. For the quantitative analysis of this indicator, the number of trace lines in the specified time series is mainly counted. The key point of this part is how to accurately draw the dynamic trajectory data of the same pedestrian target in the video stream data. Here, we use the DeepSORT algorithm and a dynamic Kalman filter for processing

Table A1. Statistical algorithm flow of the number of people in the scene

Process	Processing flow
Step 1	Frames from the video scene data stream are extracted, the time interval $n = 5$ s is determined and key frames are extracted according to the fixed time interval.
Step 2	The key frame images are sorted according to the time sequence, and the time represented by each image is recorded.
Step 3	The model trained by the YOLO-TP depth learning network is employed to recognize pedestrian targets in key frame images, and the statistical confidence of the identified targets and the frame ID are drawn.
Step 4	The following image is detected. If the target is the same, step 3 is performed, and the ID is not changed. If there is a new target, the ID is updated by increasing the current time series as the new ID. If the above conditions are not met, Step 4 performed.
Step 5	According to the time series, the number of IDs in each key frame image, that is, the number of people, is counted.
Step 6	The time length to be counted is determined, the ID quantity within the time length range is accumulated and the final number of people within the time range is output.

Table A2. Statistical algorithm flow of the number of activity tracks

Process	Processing flow
Step 1	ID information is assigned to all pedestrian targets detected in the first video sequence, which are initialized as independent new tracks.
Step 2	Simultaneously, the Kalman filter is initialized, and the trajectory state is evaluated. When the track is successfully matched for three consecutive frames, it is in a confirmed state; otherwise, it is regarded as unconfirmed.
Step 3	The Kalman filter is used to determine whether the trajectory is confirmed.
Step 4	When the track is unconfirmed, IOU matching is performed between the pedestrian track and the detection execution (Liang <i>et al.</i> 2022), and three matching results are output: (1) When the pedestrian track and the detection are matched, Kalman prediction is performed to predict the position of the new track, and then step 3 is performed. (2) When the pedestrian target detection does not match the track, it is initialized as a new track, a new pedestrian ID is assigned, and step 3 is performed. (3) When the pedestrian track does not match the detection completion, the track's state is evaluated. When a track is in an indeterminate state, the track is deleted. When the track is determined, whether the number of mismatched frames exceeds the set threshold is determined. When the number of mismatched frames exceeds the set threshold, the track is deleted. When the number of mismatches is less than the set threshold, step 3 is performed.
Step 5	When the track is in a confirmed state, the pedestrian track is cascaded with the new detection (Guo <i>et al.</i> 2022), and the cascade matching also outputs three results: (1) When the pedestrian track matches the detection, step 2 is performed. (2) When the track is not matched, it is input into the IOU matching model and associated with the unmatched detection. (3) When the detection is not matched, it is input into the IOU matching model and associated with the unmatched track.

Continued

Table A2. Continued	
Process	Processing flow
Step 6	Steps 2, 3, 4, and 5 are repeated until the video sequence is completed.
Step 7	The number of continuous tracks in the monitoring period is counted according to the time to be detected.

Table A3. Statistical algorithm flow of the number of people in key areas	
Process	Processing flow
Step 1	The location and number of key areas according to the scene and the marking box with a fixed size in the key areas are determined.
Step 2	The pedestrian target is detected frame by frame, and the detected target is marked with a marker box.
Step 3	When the personnel target box collides with the frame mark of the key area, the overlap ratio of the target range in the key area to the target frame mark is calculated.
Step 4	If the overlap ratio exceeds 30%, it is counted once; otherwise, it will not be counted.
Step 5	Steps 2, 3, and 4 are repeated until the end of the video stream detection data.
Step 6	The above steps are performed once for one area until all are completed.
Step 7	The sum of the final count within the specified time, that is, the number of people in key areas, is counted.

(Zhao, Zhang & Fu 2020). The statistical algorithm for this trajectory is shown in Table A2.

(3) Number of people in key areas: the principle of algorithmic statistics here is to count the effective collisions of the frame labels in key areas. The key areas described here are determined according to different situations. For example, in the kitchen scene, we usually set the key area in the kitchen stove area to effectively count the number of people cooking. The steps and principles of the specific statistical algorithm are shown in Table A3.

(4) Time of appearance of people in key areas: here, we refer to the previous indicator 3, count the time of effective collisions in the key areas, and accumulate them. This indicator reflects the continued attractiveness of key areas to personnel in the design workshop. The statistical algorithm flow for this indicator is shown in Table A4.

(5) Frequency of eye contact: This indicator mainly counts the frequency of eye contact between two people in communication. Here, two people's eye contact is evaluated mainly by the same duration of their eye gaze. When the duration is greater than 3 seconds, we believe that these two people are making eye contact. The statistical algorithm flow for this indicator is shown in Table A5.

(6) Frequency of common facial expressions: Common facial expressions are used to describe the interaction between two people in the design workshop space. In general, if people have the same facial expressions in the design workshop, it can

Table A4. Statistical algorithm flow of time of appearance of people in key areas	
Process	Processing flow
Step 1	The location and number of key areas according to the scene and the marking box with a fixed size in the key areas are determined.
Step 2	The pedestrian target is detected frame by frame, and the detected target is marked with a marker box.
Step 3	When the personnel target box collides with the frame mark of the key area, a timer is started to count the time of this event until the personnel target box is out of contact with the frame mark of the key area.
Step 4	When a new person target collides with the frame mark of the key area, step is repeated3, a new timer is created, and the time until the video stream detection within the count time is completed is counted.
Step 5	The timer with a screening time of less than 3 s is deleted to eliminate the false detection caused by accidental collision.
Step 6	The time counted by all timers in the detection period and the time when the number of people in the key area appears is counted.
Step 7	The above steps for each key area are repeated until the detection is completed.

Table A5. Statistical algorithm flow of frequency of eye contact	
Process	Processing flow
Step 1	The video stream data are split frame by frame.
Step 2	Using the YOLO-TP network model to identify the person's target, the head orientation of the person is further determined and the head orientation direction line is drawn and used as the focus direction of the person's eyes.
Step 3	The gaze direction lines of the two nearest pedestrian targets are checked. When the included angle between the two direction lines is between 150° and 180°, data collection will commence. If the time exceeds 3 s, it will be recorded. If the time is insufficient, it will not be recorded.
Step 4	Steps 2 and 3 are repeated for the detection time series until the detection is completed.
Step 5	All the recorded times are accumulated and summed.

be considered that there is a typical emotional interaction between these people. We choose to count the number of changes in the common facial expressions, and we can observe the activity of the workshop through these data. The specific statistical algorithm is shown in [Table A6](#).

(7) Mutual social distance: this indicator mainly reflects the interaction distance of personnel in the design workshop. In the actual statistical process, we recorded the average value of the relative pixel distance. We connect all the adjacent nearest pedestrian targets in each frame detection scene according to the center of gravity position of the frame marker, sum the pixel distance obtained

Table A6. Statistical algorithm flow of the frequency of common facial expressions	
Process	Processing flow
Step 1	The video stream data is split and detected frame by frame.
Step 2	Based on using the YOLO-TP network model to recognize the human target, the position of the human head further determined, and the facial expressions are distinguished.
Step 3	Initially, if the facial expressions of the two nearest pedestrian targets are consistent, the facial expression information is retained, and the facial expression of the person target with the retained information is continuously monitored.
Step 4	A common expression change is recorded if one or more pairs of the monitored people's expression types are inconsistent.
Step 5	People with the same ID attribute have one and only one expression type in each frame. Only when the expressions of one or more adjacent pairs of people are consistent can the expression information be retained for subsequent discrimination.
Step 6	Steps 2, 3 and 4 are repeated until the end of the time to be detected, all times are accumulated and data are recorded.

Table A7. Statistical algorithm flow of mutual social distance	
Process	Processing flow
Step 1	The video stream data is split and detected frame by frame.
Step 2	Based on identifying the personnel target using the YOLO-TP network model, the location of the center of gravity of the personnel marking frame is further determined.
Step 3	The nearest target of adjacent people is calculated, the center of gravity is connected, the pixel distance length of the connection is calculated and the sum of the distance length is divided by the total number of people in the scene.
Step 4	The average distance of each frame is recorded after the end of each frame, and the frames with no target detected are deleted from the total frames.
Step 5	The average distance of all recorded frames is divided by the total number of frames.
Step 6	The above steps are cycled until all tests are completed within the test time range.

by connecting multiple pairs of human targets, divide it by the total number of people in the scene and obtain the average distance of each frame. We sum the average distance of each frame and average it. The size of this indicator reflects the degree of intimacy between two people in the design workshop. The specific statistical algorithm is shown in Table A7.

(8) Frequency of common attention: although different design workshops have different themes and different environmental scenes, the theme of each workshop is relatively fixed. The critical things in the scene are relatively fixed, so it is important that the people in the statistical design workshop pay attention to things in the same scene. Here, we count the number of times two people pay attention to the same thing. The specific statistical algorithm is shown in Table A8.

Table A8. Statistical algorithm flow of the frequency of common attention

Process	Processing flow
Step 1	Video stream data are split and detected frame by frame.
Step 2	Based on using the YOLO-TP network model to identify the person's target, the head orientation of the person is further determined and the head orientation direction line is drawn and used as the gaze direction of the person's eyes.
Step 3	The gaze direction lines of the two nearest pedestrian targets are checked. When the common intersection of the two direction lines is similar, the time is recorded and counted.
Step 4	When the direction gaze direction line simultaneously deviates from the thing, timing is stopped, and the time whose duration is less than 2 s is deleted to suppress the accidental error.
Step 5	The above steps are repeated until all video streams within the time to be detected are completed.
Step 6	The number of times that people in the scene simultaneously focus on the same thing is counted.

Table A9. Comparison between algorithm calculation of 8 indicators and manual interpretation data

Test data number	Indicator	1	2	3	4	5	6	7	8
1	Algorithm calculation	34	42	4	230	69	14	428	14
	Manual interpretation	34	42	4	224	65	13	420	13
	Accuracy	1	1	1	0.94	0.96	0.99	0.92	0.99
2	Algorithm calculation	22	45	3	310	45	17	368	12
	Manual interpretation	22	44	3	319	44	18	369	12
	Accuracy	1	0.99	1	0.91	0.99	0.99	0.99	1
3	Algorithm calculation	28	8	2	420	10	17	279	18
	Manual interpretation	28	8	2	425	10	18	278	18
	Accuracy	1	1	1	0.95	1	0.99	0.99	1
4	Algorithm calculation	20	9	6	530	15	21	320	19
	Manual interpretation	21	9	6	525	18	20	326	19
	Accuracy	0.99	1	1	0.95	0.97	0.99	0.94	1
5	Algorithm calculation	15	4	2	150	11	10	299	10
	Manual interpretation	14	4	2	154	11	10	304	10
	Accuracy	0.99	1	1	0.96	1	1	0.95	1
6	Algorithm calculation	18	4	1	230	35	15	235	29
	Manual interpretation	18	4	1	234	33	14	230	28
	Accuracy	1	1	1	0.96	0.98	0.99	0.95	0.99
7	Algorithm calculation	25	3	2	356	51	25	215	37
	Manual interpretation	28	3	2	355	48	28	209	36
	Accuracy	0.97	1	1	0.99	0.97	0.97	0.94	0.99

Continued

Table A9. Continued									
Test data number	Indicator	1	2	3	4	5	6	7	8
8	Algorithm calculation	15	9	1	250	48	19	180	25
	Manual interpretation	15	9	1	254	48	18	172	26
	Accuracy	1	1	1	0.96	1	0.99	0.92	0.99
9	Algorithm calculation	16	5	2	314	21	18	245	21
	Manual interpretation	16	5	2	311	21	18	244	21
	Accuracy	1	1	1	0.97	1	1	0.99	1
10	Algorithm calculation	10	25	1	180	9	12	397	15
	Manual interpretation	10	25	1	188	8	12	400	15
	Accuracy	1	1	1	0.92	0.99	1	0.97	1
11	Algorithm calculation	8	18	5	510	18	24	355	18
	Manual interpretation	8	18	5	511	18	22	360	19
	Accuracy	1	1	1	0.99	1	0.98	0.95	0.99
12	Algorithm calculation	21	21	3	498	56	15	348	25
	Manual interpretation	21	21	3	495	55	16	350	23
	Accuracy	1	1	1	0.97	0.99	0.99	0.98	0.98
13	Algorithm calculation	15	5	10	466	51	17	211	29
	Manual interpretation	14	5	10	466	52	18	206	28
	Accuracy	0.99	1	1	1	0.99	0.99	0.95	0.99
14	Algorithm calculation	41	9	2	499	54	38	198	39
	Manual interpretation	41	9	2	500	54	39	199	38
	Accuracy	1	1	1	0.99	1	0.99	0.99	0.99
15	Algorithm calculation	15	3	1	450	25	19	222	29
	Manual interpretation	15	3	1	444	24	18	220	28
	Accuracy	1	1	1	0.94	0.99	0.99	0.98	0.99
16	Algorithm calculation	27	5	1	466	18	22	231	19
	Manual interpretation	27	5	1	460	18	23	230	19
	Accuracy	1	1	1	0.94	1	0.99	0.99	1
17	Algorithm calculation	35	3	3	512	26	24	255	24
	Manual interpretation	35	3	3	509	25	24	258	22
	Accuracy	1	1	1	0.97	0.99	1	0.97	0.98
18	Algorithm calculation	34	8	2	454	15	18	241	22
	Manual interpretation	34	8	2	450	13	18	240	21
	Accuracy	1	1	1	0.96	0.98	1	0.99	0.99

Continued

Table A9. Continued									
Test data number	Indicator	1	2	3	4	5	6	7	8
19	Algorithm calculation	33	4	2	411	18	26	251	21
	Manual interpretation	34	4	2	410	18	25	250	20
	Accuracy	0.99	1	1	0.99	1	0.99	0.99	0.99
20	Algorithm calculation	32	6	1	401	15	22	249	20
	Manual interpretation	33	6	1	400	18	21	245	19
	Accuracy	0.99	1	1	0.99	0.97	0.99	0.96	0.99
21	Algorithm calculation	38	6	5	519	25	29	284	29
	Manual interpretation	39	6	5	520	24	27	288	28
	Accuracy	0.99	1	1	0.99	0.99	0.98	0.96	0.99
22	Algorithm calculation	12	12	1	477	21	10	399	31
	Manual interpretation	12	11	1	469	22	11	398	30
	Accuracy	1	0.99	1	0.92	0.99	0.99	0.99	0.99
23	Algorithm calculation	10	1	2	320	26	18	251	35
	Manual interpretation	10	1	2	322	25	18	255	35
	Accuracy	1	1	1	0.98	0.99	1	0.96	1
24	Algorithm calculation	39	5	1	459	35	29	395	39
	Manual interpretation	39	5	1	458	35	29	396	38
	Accuracy	1	1	1	0.99	1	1	0.99	0.99
25	Algorithm calculation	29	39	5	429	51	35	393	38
	Manual interpretation	28	38	5	422	52	35	396	38
	Accuracy	0.99	0.99	1	0.93	0.99	1	0.97	1
26	Algorithm calculation	40	72	6	581	52	45	358	29
	Manual interpretation	41	71	6	588	51	44	356	29
	Accuracy	0.99	0.99	1	0.93	0.99	0.99	0.98	1
27	Algorithm calculation	25	36	2	457	26	28	321	29
	Manual interpretation	24	35	2	455	25	27	325	28
	Accuracy	0.99	0.99	1	0.98	0.99	0.99	0.96	0.99
28	Algorithm calculation	15	22	5	488	28	29	333	31
	Manual interpretation	15	21	5	489	27	30	335	31
	Accuracy	1	0.99	1	0.99	0.99	0.99	0.98	1
29	Algorithm calculation	11	4	1	522	16	18	328	21
	Manual interpretation	11	4	1	521	15	17	326	20
	Accuracy	1	1	1	0.99	0.99	0.99	0.98	0.99

Continued

Table A9. Continued									
Test data number	Indicator	1	2	3	4	5	6	7	8
30	Algorithm calculation	61	75	12	185	49	37	405	36
	Manual interpretation	60	75	15	180	45	33	404	33
	Accuracy	0.99	1	0.97	0.95	0.96	0.96	0.99	0.97
31	Algorithm calculation	9	15	2	325	15	14	384	24
	Manual interpretation	9	15	2	322	14	13	388	23
	Accuracy	1	1	1	0.97	0.99	0.99	0.96	0.99
32	Algorithm calculation	4	6	1	410	10	9	510	21
	Manual interpretation	4	6	1	409	10	8	510	20
	Accuracy	1	1	1	0.99	1	0.99	1	0.99
33	Algorithm calculation	8	12	1	214	14	10	489	29
	Manual interpretation	8	11	1	220	14	9	485	25
	Accuracy	1	0.99	1	0.94	1	0.99	0.96	0.96
34	Algorithm calculation	15	31	3	356	29	32	351	39
	Manual interpretation	14	30	3	355	28	33	350	39
	Accuracy	0.99	0.99	1	0.99	0.99	0.99	0.99	1
35	Algorithm calculation	12	2	2	600	62	41	211	45
	Manual interpretation	11	2	2	603	64	40	211	44
	Accuracy	0.99	1	1	0.97	0.98	0.99	1	0.99
36	Algorithm calculation	11	3	1	510	12	10	362	21
	Manual interpretation	10	3	1	511	11	9	366	20
	Accuracy	0.99	1	1	0.99	0.99	0.99	0.96	0.99
37	Algorithm calculation	18	5	2	451	32	25	341	29
	Manual interpretation	19	5	2	455	33	26	344	29
	Accuracy	0.99	1	1	0.96	0.99	0.99	0.97	1
38	Algorithm calculation	8	9	2	484	10	15	358	24
	Manual interpretation	8	9	2	485	9	14	356	23
	Accuracy	1	1	1	0.99	0.99	0.99	0.98	0.99
39	Algorithm calculation	17	25	1	410	15	10	499	29
	Manual interpretation	16	26	1	415	14	13	500	29
	Accuracy	0.99	0.99	1	0.95	0.99	0.97	0.99	1
40	Algorithm calculation	25	4	2	500	29	13	351	27
	Manual interpretation	25	4	2	500	29	13	350	27
	Accuracy	1	1	1	1	1	1	0.99	1

To verify the reliability of the 8 statistical data obtained by our method, we selected 40 scene segments for algorithm metric calculation. We compared the data obtained by our method with the manually interpreted data indicators and obtained a detailed comparison table as shown in [Table A9](#).