The 1st International Workshop and Challenge on Disentangled Representation Learning for Controllable Generation

Xin Jin[1], Qiuyu Chen[1,2], Yue Song[3], Xihui Liu[4], Shuai Yang[5], Tao Yang[6], Ba'ao Xie[1], Nicu Sebe[7], and Wenjun (Kevin) Zeng[1]

[1] *Eastern Institute of Technology, Ningbo, China*
[2] *Shanghai Jiao Tong University, Shanghai, China*
[3] *California Institute of Technology, USA*
[4] *The University of Hong Kong, Hong Kong, China*
[5] *Peking University, Beijing, China*
[6] *Xi'an Jiaotong University, Xi'an, China*
[7] *University of Trento, Italy*

1 Summary

Workshop title	The 1st International Workshop and Challenge on Disentangled Representation Learning for Real-world Applications
Acronym	DRL for Real
Edition (1st, 2nd, ...)	1st
Keywords	Disentanglement, Generative model
Primary contact name and email	Xin Jin, jinxin@eitech.edu.cn
Half or full day	Half day
Anticipated audience size	50
Requested number of poster boards	20
Papers published in proceedings?	20
Special requests	N/A

2 Topic

Disentangled representation learning (DRL) is believed to be one of the possible ways for AI to fundamentally understand the world, which potentially helps alleviate wicked problems like hallucination issue in LMMs/MLLMs [1, 2] and controllability issue in Generative Models [3–6], so as to eventually achieve Artificial General Intelligence (AGI). Over the years, disentangled representation learning (DRL) has garnered significant academic interests and research contributions, which has been recognized for its ability to improve model robustness, interpretability, and generalizability [7–9]. However, the field currently faces a notable gap: **the lack of comprehensive, realistic benchmark and unified evaluation metric that reflects real-world characteristics, which makes DRL still limited in the synthetic toy scenarios and hard to apply to more practical applications.**

To bridge the above gap, this "DRL for Real" workshop aims to strengthen the emphasis of DRL research toward practical utility and real-world applicability. *The core objective of this workshop is to make disentanglement genuinely useful and practically relevant.* Specifically, we will introduce novel, realistic datasets and comprehensive benchmarks designed to better evaluate DRL methods in practical scenarios. Additionally, the workshop will discuss concrete applications of DRL in critical domains such as controllable generation and autonomous driving.

This workshop is intended for researchers and practitioners interested in DRL methodologies and their integration into real-world systems. Attendees will gain valuable insights into state-of-the-art approaches, current practical challenges, and promising future directions. We anticipate strong participation from ICCV 2025 attendees working in representation learning, interpretability, and practical

AI applications, who share our vision of making disentangled representation learning truly practical and impactful.

3 Organizers and speakers

3.1 List of organizers

- **Prof. Xin Jin**, Eastern Institute of Technology, Ningbo, China
  Email: jinxin@eitech.edu.cn
  Bio: `https://scholar.google.com/citations?user=byaSC-kAAAAJ&hl=zh-CN`
- **Dr. Qiuyu Chen**, Eastern Institute of Technology, Ningbo, China
  Email: canghaimeng@sjtu.edu.cn
  Bio: `https://scholar.google.com/citations?user=jIhRaNQAAAAJ&hl=zh-CN`
- **Prof. Yue Song**, California Institute of Technology (Caltech), USA
  Email: yue.song@unitn.it
  Bio: `https://kingjamessong.github.io/`
- **Prof. Xihui Liu**, The University of Hong Kong, Hong Kong, China
  Email: xihuiliu@eee.hku.hk
  Bio: `https://xh-liu.github.io/`
- **Prof. Shuai Yang**, Peking University, Beijing, China
  Email: williamyang@pku.edu.cn
  Bio: `https://williamyang1991.github.io/`
- **Dr.Tao Yang**, Xi'an Jiaotong University, Xi'an, China
  Email: yt14212@stu.xjtu.edu.cn
  Bio: `https://scholar.google.fi/citations?user=qT5psCEAAAAJ&hl=fi`
- **Prof. Ba'ao Xie**, Eastern Institute of Technology, Ningbo, China
  Email: bxie@idt.eitech.edu.cn
  Bio: `https://scholar.google.com/citations?user=-VXIO1sAAAAJ&hl=zh-CN&oi=ao`
- **Prof. Nicu Sebe**, Fellow of IAPR, University of Trento, Italy
  Email: sebe@disi.unitn.it
  Bio: `https://disi.unitn.it/~sebe/`
- **Prof. Wenjun (Kevin) Zeng**, IEEE Fellow, Eastern Institute of Technology, Ningbo, China
  Email: wzeng-vp@eitech.edu.cn
  Bio: `https://scholar.google.com/citations?user=_cUfvYQAAAAJ&hl=zh-CN`

3.2 Organizers' experience and background

Our organizing committee brings together a diverse team of researchers with complementary expertise in disentangled representation learning, computer vision, and deep learning. Prof. Xin Jin (EIT) leads research in deep learning and computer vision, having organized tutorial sessions on "Disentanglement and Compositionality in Computer Vision" at CVPR 2024 and ECCV 2024. Dr. Yue Song, a post-doctoral researcher at Caltech, brings expertise in structured representation learning with focus on high-order representation learning and disentangled representation learning. Prof. Wenjun (Kevin) Zeng, an IEEE Fellow and Foreign Member of the Canadian Academy of Engineering, provides strategic guidance with decades of experience in multimedia computing and deep learning. Collectively, our team has extensive experience organizing workshops, tutorials, and special sessions at major computer vision conferences (CVPR, ECCV, ICCV, IEEE ICME, IEEE ICIP, IEEE VCIP), and our diverse backgrounds spanning academia and industry uniquely position us to bridge theoretical advances in disentangled representation learning with practical applications. The team's complementary expertise covers the full spectrum of DRL research, from theoretical foundations to real-world implementations, ensuring a comprehensive and impactful workshop experience.

3.3 List of invited speakers

- **Prof. Tao Mei** (HiDream.ai) - *Tentative confirmation*
  `https://taomei.me/`
- **Prof. Jiajun Wu** (Stanford University) - *Tentative confirmation*
  `https://jiajunwu.com/`
- **Prof. Tongliang Liu** (The University of Sydney) - *Tentative confirmation*
  `https://tongliang-liu.github.io/`
- **Prof. Deva Ramanan** (Carnegie Mellon University) - *Tentative confirmation*
  `https://www.cs.cmu.edu/~deva/`
- **Prof. Olga Russakovsky** (Princeton University) - *Tentative confirmation*

https://www.cs.princeton.edu/~olgarus/

- **Prof. Phillip Isola** (Massachusetts Institute of Technology) - *Tentative confirmation*
  https://web.mit.edu/phillipi/

Our invited speakers represent the forefront of research in disentangled representation learning and controllable generation. They cover the full spectrum from theoretical foundations to practical implementations of DRL, ensuring comprehensive coverage of our workshop's themes.

3.4 Diversity

The organizing team and speakers represent a balanced mix across genders, geographical regions, and career stages—from emerging scholars to established researchers. We have intentionally brought together members from both academia and industry, integrating diverse perspectives from universities, research institutions, and technology companies. This multifaceted commitment to diversity ensures comprehensive insights into the real-world applications of disentangled representation learning, while fostering inclusive participation from across the broader research community.

4 Format and logistics

This workshop will be in-person.

4.1 Schedule

Date Range	Task/Deadline
April 10, 2025	Competition Announcement and Dataset Release
May 10, 2025	Competition Submission Deadline
May 15, 2025	Paper Submission Deadline
June 10, 2025	Notification of Competition Results
June 15, 2025	Notification of Paper Acceptance
June 25, 2025	Final Documents Submission to ICCV
October 2025	Workshop and Competition Results Presentation at ICCV 2025

The schedule above integrates key dates for both paper submissions and the competition, ensuring participants have clear timelines for preparation and participation.

4.2 Paper submission

4.2.1 Tentative Program Committee

The tentative program committee (PC) consists of experts in the list of organizers. This committee will be responsible for evaluating the submitted papers, recommending accepted papers for the workshop, and ensuring the quality of submissions aligns with the scope and goals of the workshop.

4.2.2 Paper Review Timeline

The timeline for paper submissions and reviews is as follows:

- **Paper Submission Deadline**: May 15th, 2025
- **Notification of Acceptance**: June 15th, 2025
- **Final Documents Submission to ICCV**: June 25th, 2025

4.2.3 Publication in Proceedings

Yes, all accepted papers will be published in the ICCV 2025 workshop proceedings, which will be officially distributed through the conference's channels and made accessible to all ICCV attendees.

4.3 Competition

To further promote the practical utility and real-world applicability of Disentangled Representation Learning (DRL), we will organize a competition as part of the workshop. The primary goal of this competition is to promote the development and evaluation of DRL methods on realistic datasets, thereby accelerating the transition of disentangled models from theoretical research to practical applications. By training and evaluating models on real-world data, our aim is to foster meaningful advancements in the DRL community and facilitate its integration into practical scenarios.

(a) Datasets to be used: The proposed dataset is specifically designed for disentangled representation learning (DRL), covering multiple realistic scenarios and diverse disentanglement factors, as illustrated in Figure 1.

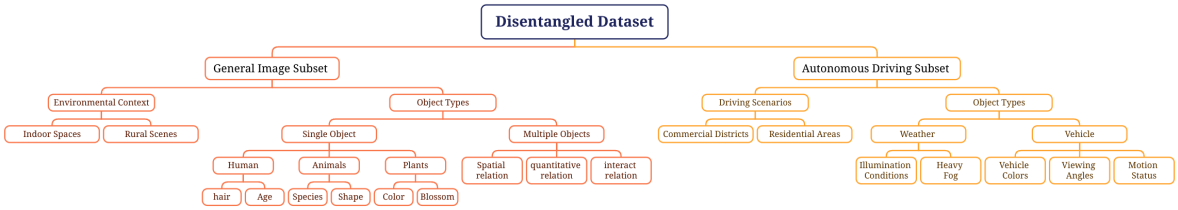

Figure 1: Hierarchical structure of the dataset, showing the organization of scenarios, object types, and disentanglement factors.

It consists of two main subsets: (i) a **general image subset** containing scenarios in daily life and (ii) an **autonomous driving subset**. The general image subset comprises about 25,000 images across several distinct categories, such as human faces, animals, plants, etc., following the conceptual framework shown in Figure 2.

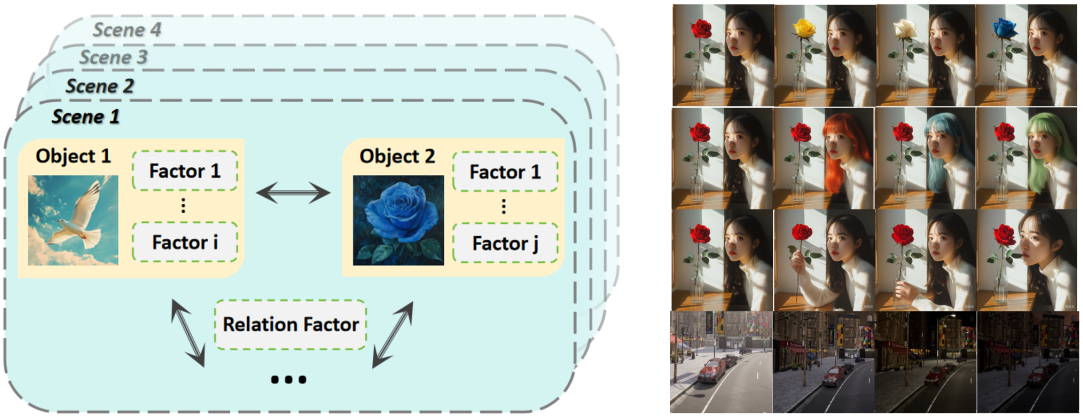

Figure 2: The image presents a comprehensive visualization of our dataset structure. On the left side, an intuitive diagram illustrates the conceptual framework, while representative examples are shown on the right. In the general image examples, the first row demonstrates variations in the color of the rose—and the second row shows changes in the person's hair color. The third row illustrates relationship factors between objects. For the autonomous driving subset shown in the bottom row, we present images that vary according to illumination conditions.

Images in this subset are partially synthesized using text-to-image generative models [10], while others are derived from real-world video sources such as time-lapse photography. Thus, the final general image subset contains a balanced mixture of both real and generated images, providing a comprehensive and realistic benchmark for evaluating disentangled representation models. Similarly, the autonomous driving subset includes about 7,000 images systematically generated using the CARLA simulator [11], covering 12 distinct vehicle types and more than 5 background scenarios (e.g., rainy, foggy, sunny, and varying illumination conditions). Each scenario serves as a base, with variations across multiple disentanglement factors such as viewing angles, distances, vehicle colors, weather conditions, and lighting changes. This subset is particularly valuable for training controllable generation models that can produce diverse autonomous driving scenarios for simulation and training purposes.

In addition, **two parallel competitions** will be organized in this workshop, corresponding to two distinct versions of dataset.

- **Scenario 1:** Each group of images in the dataset varies along only a single disentanglement factor. It's designed to evaluate the performance of disentangled representation models under idealized conditions.
- **Scenario 2:** Each group of images in the dataset simultaneously varies across multiple disen-

tanglement factors. The complex scenario closely resembles real-world conditions, and models performing well on this dataset are expected to have greater practical value.

These two types of datasets constitute the complete dataset used in this workshop, enabling both theoretical evaluation and practical assessment of existing models in the field of disentangled representation learning.

(b) Dataset availability and contingency plan: Currently, the dataset is under active development and is expected to be publicly available approximately five months prior to the ICCV 2025 workshop (tentatively by May 2025). In the event of unforeseen delays, we will release a smaller-scale preliminary version of the dataset at least six months before the workshop, ensuring participants have sufficient time to develop and test their methods.

(c) Ethical considerations for the datasets: Ethical implications in dataset construction is carefully considered. We will ensure diversity and representativeness in the dataset to avoid potential biases related to ethnicity, gender, age, or other sensitive attributes. Furthermore, clear documentation regarding data collection, annotation processes, and intended usage will be provided to all participants.

(d) Evaluation of submissions: Submissions will be evaluated using a unified evaluation framework specifically designed for disentanglement metrics, including but not limited to:

- **Factor Disentanglement Metrics:** FactorVAE score [12], Mutual Information Gap (MIG) [13], DCI disentanglement score [14].
- **Downstream Task Performance:** Accuracy and robustness on downstream tasks (e.g., classification, regression) leveraging learned representations.
- **Transferability Across Domains:** Quantitative evaluation of the transfer capability of DRL models across different domains.

Participants will submit their trained models along with a short technical report describing their methods. The evaluation committee will independently verify results and rank submissions accordingly.

4.4 Special requests

N/A

5 Broader impacts

The proposed workshop have profound impacts on both academia and industry. By providing a comprehensive benchmark dataset and unified evaluation framework, we aim to accelerate research in DRL, facilitating the development of more robust, interpretable, and generalizable AI models. Such advancements have the potential to significantly enhance AI applications in critical real-world domains, including autonomous driving, robotics, and human-computer interaction.

5.1 Social considerations

The development of DRL techniques could lead to more transparent and interpretable AI systems, thereby fostering greater public trust and understandings of AI. Enhanced interpretability empowers users and stakeholders by providing clearer insights into AI decision-making processes, enabling more responsible and informed use of AI in society. Furthermore, this workshop aims to ensure that the benefits of these advancements are equitably distributed and accessible across diverse communities, helping to prevent the exacerbation of existing social inequalities.

5.2 Ethical considerations

Ethical considerations will be explicitly addressed throughout the workshop. Participants will be encouraged to engage in thoughtful discussions and exchange best practices regarding ethical AI research and responsible deployment. Particular emphasis will be placed on promoting the ethical use of DRL technologies in sensitive or high-stakes domains, such as autonomous driving, healthcare, and surveillance, where responsible and transparent AI practices are crucial.

6 Relationship to previous workshops

This workshop builds upon and extends our recent successful tutorial sessions at CVPR 2024 and ECCV 2024 on “Visual Disentanglement and Compositionality”.

References

[1] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, *et al.*, "A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions," *ACM Transactions on Information Systems*, vol. 43, no. 2, pp. 1–55, 2025.

[2] H. Orgad, M. Toker, Z. Gekhman, R. Reichart, I. Szpektor, H. Kotek, and Y. Belinkov, "Llms know more than they show: On the intrinsic representation of llm hallucinations," *arXiv preprint arXiv:2410.02707*, 2024.

[3] A. Pati and A. Lerch, "Is disentanglement enough? on latent representations for controllable music generation," *arXiv preprint arXiv:2108.01450*, 2021.

[4] Z. Song, O. Koyejo, and J. Zhang, "Toward a controllable disentanglement network," *IEEE Transactions on Cybernetics*, vol. 52, no. 4, pp. 2491–2504, 2020.

[5] X. Shen, F. Liu, H. Dong, L. Qing, Z. Chen, and T. Zhang, "Disentangled generative causal representation learning," 2020.

[6] W. Cho, H. Ravi, M. Harikumar, V. Khuc, K. K. Singh, J. Lu, D. Inouye, and A. Kale, "Enhanced controllability of diffusion models via feature disentanglement and realism-enhanced sampling methods," in *European Conference on Computer Vision*, pp. 285–301, Springer, 2024.

[7] B. Xie, Q. Chen, Y. Wang, Z. Zhang, X. Jin, and W. Zeng, "Graph-based unsupervised disentangled representation learning via multimodal large language models," *Advances in Neural Information Processing Systems*, vol. 37, pp. 103101–103130, 2025.

[8] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[9] F. Locatello, S. Bauer, M. Lucic, G. Raetsch, S. Gelly, B. Schölkopf, and O. Bachem, "Challenging common assumptions in the unsupervised learning of disentangled representations," in *international conference on machine learning*, pp. 4114–4124, PMLR, 2019.

[10] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans, *et al.*, "Photorealistic text-to-image diffusion models with deep language understanding," *Advances in neural information processing systems*, vol. 35, pp. 36479–36494, 2022.

[11] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*, pp. 1–16, PMLR, 2017.

[12] H. Kim and A. Mnih, "Disentangling by factorising," in *International conference on machine learning*, pp. 2649–2658, PMLR, 2018.

[13] R. T. Chen, X. Li, R. B. Grosse, and D. K. Duvenaud, "Isolating sources of disentanglement in variational autoencoders," *Advances in neural information processing systems*, vol. 31, 2018.

[14] C. Eastwood and C. K. Williams, "A framework for the quantitative evaluation of disentangled representations," in *6th International Conference on Learning Representations*, 2018.