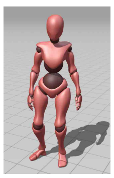
I work on **real-time** and **data-driven** algorithms for **visual systems**. My research focus is on research focus is in scene analysis and synthesis[1,4,8,11,12,15,19,20,23], the **simulation** of physically-embodied **AI agents**[18,16,13,21,22], deep learning and reinforcement learning[17,13,1,23,15,11]. My work has been recognized by the ACM SIGGRAPH conference on Motion in Games, where I received the **Best Paper Award**[21]. In addition, I was a **finalist** in both the **ACM SIGGRAPH Thesis Fast Forward** and **ACM SIGGRAPH Asia Doctoral Consortium** in 2018. Before my current **assistant professor role** at NJIT, I worked in research roles in **Amazon**, **Autodesk Research** and **Wayfair**. In the following sections, I summarize my current research and long-term goals.

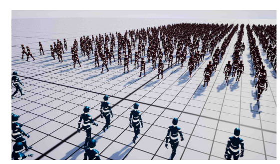
1. Multi-Agent Motion Dynamics

Multi-agent crowd simulation has direct applications in entertainment, pedestrian analysis, urban planning, robotics and autonomous systems. A **crowd** is a **collection of independent, self-actuated agents**. Each agent has individual navigational goals in this shared environment. Since agents share the same environment, they can interact and collide with each other. Agent movement is controlled by a navigation algorithm, which needs to ensure that an **agent progresses towards its goal, while avoiding collisions**. However, computing collision-free agent motion is difficult, due to the complexity of such dynamic interactions. Field experiments are one important avenue for testing a navigation algorithm. Another approach is to conduct a virtual simulated experiment, which allows quick insight into the dynamics and performance of such navigation algorithms. Furthermore, a simulation is more efficient and less costly than real-world field experiments, and allows us to easily test corner-cases, and "what-if" scenarios. Visualizing the simulated scenario allows us to capture high-level details, as well as group behaviors that might not be otherwise noticeable.



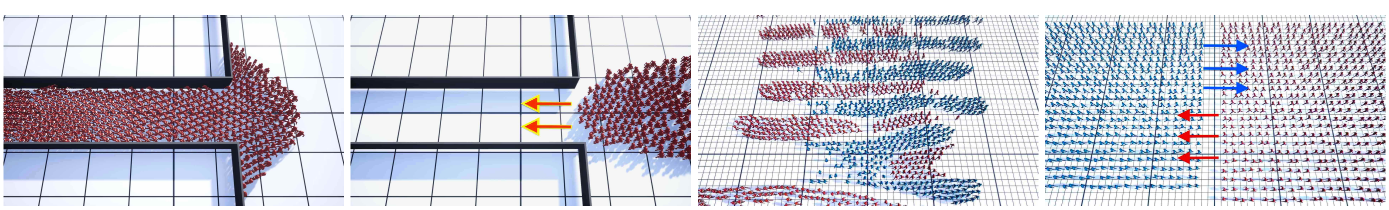
Agent

Simulations are composed of a series of discrete time steps. Agents run through a continuous cycle of sensing and acting, where each cycle correlates to a time step. At the beginning of each cycle, each agent independently computes a trajectory to its goal, while avoiding collisions with other agents or obstacles. In addition to collision-free movement, a simulation should **capture both individual and group behavior** observed in real crowds, while being computationally interactive.



Multi-Agent Crowds

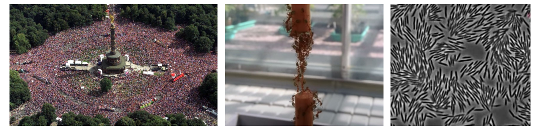
Despite 30+ years in crowd simulation research, simulation methods have subspecialized, and are **computationally effective only on a case-by-case basis**[14]. For example, methods for dense crowds smooth out individual agent motions, while methods for sparse crowds cannot computationally cope with dense situations. Since most simulation scenarios contain varying crowd densities, neither of the existing approaches are robust, scalable, or generalizable.



(a) A group of agents passing through a narrow corridor. **(b)** Two groups exchanging positions.

Figure 1: Emergent phenomena in crowd simulation. Our results[21] recreate different phenomena found in pedestrian crowds, without any scripting or other user-directed control. These phenomena include: (a) clogging and arching near bottlenecks[7], (b) groups self-organizing into lanes[6], with stable and real-time performance for 100,000+ individuals. These results are made possible by framing the crowd motion as a constraint optimization problem, which is then efficiently solved with a GPU.

Our award-winning work[21] allows real-time simulation of **both dense and sparse crowds**. These results are made possible by reframing crowd motion as a **constrained optimization problem**. Crowd motion is controlled by numerical constraints on individual agent positions. For example, a constraint for collision avoidance extrapolates each agent's trajectory to detect future collisions. In case of a collision, agents adjust their trajectory according to the direction that **minimizes constraint violation**, with the magnitude of the change depending on the amount of time until an impending collision[10]. Otherwise, agents locomote directly to their goal.



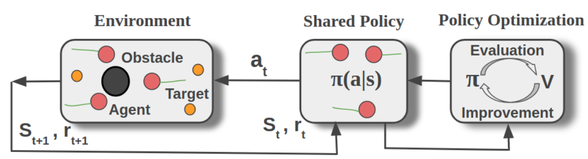
Collective behavior in crowds, ant swarms and bacteria.

All computations are processed **efficiently on a GPU**, allowing **real-time motion for 100,000+ agents**. This is achievable since agent motion constraints are solved independently, and therefore are easily parallelizable (Fig. 1). For these results, I received the **Best Paper Award** at the ACM SIGGRAPH conference on Motion in Games 2017[21].

Our group has been working on extending these results to multiple crowd benchmarks[22], including agent navigation in 3D spaces[16], formation control, and robotics applications[17]. In our latest results, we show how Deep Reinforcement Learning is valuable as a motion suggestion engine where the simulation mechanics are controlled by our method[21]. Aside from the immediate practical implications in entertainment and robotics, our lab is working on further expanding this work to collective bacteria motion, virtual ant swarms, and for the training of robotic agents to move safely in heterogeneous crowded environments.

Future Directions

Learning-based Collective Motion. Deep Learning and Reinforcement Learning (RL) have proven useful for domains where the underlying model is unknown or complex to define. Collective behavior dynamics is one such case, since comprehensive analytic models have eluded researchers. To address this challenge, we plan to: (i) Better understand collective dynamics via deep representations, where such representations allow us to predict, classify and detect anomalies collective dynamics scenarios; (ii) Simulate multi-agent dynamics driven by learning, in which we will develop RL models for emulating real-world pedestrians. In contrast to existing work that targets small or synthetic benchmarks[3], our model's goal is to learn adaptive behavior that fits a set of diverse real-world behaviors[17,9]; (iii) Better understand collective behaviors arising in nature. Understanding the dynamics of biological systems such as ant swarms and bacteria colonies is important for treating infectious diseases, reducing traffic congestions, and for enhancing robot groups with the ability to self-assemble. This research trajectory is currently funded through internal grants.



RL for multi-agent navigation[17].

2. Synthesis and Analysis of 3D Spaces

Virtual worlds have grown in terms of complexity and interactivity, becoming an integral part of domains such as engineering, education, retail, robotics and entertainment. These trends have been increasing exponentially due to the rise of gaming and VR. However, virtual worlds are **challenging to construct**, requiring professionals with training in 3D modeling software. Our lab's research accelerates 3D content creation process with contributions in:

Layout Synthesis. We developed computational methods for automatically creating spaces using optimization and data[19,11]. In the **optimization** approach, users provide a set of 3D objects, and spatial arrangement constraints from which our method computes layout proposals. This is achieved by viewing layout synthesis from the prism of simulating deformable bodies. Both layouts and deformable bodies can be described by geometric constraints, which should be satisfied for a layout to be realistic, or for the movement of a body to be plausible. Both cases can be tackled using continuous optimization procedures. A layout synthesis solution is then an arrangement that satisfies such constraints (Fig. 3). The novelty of our work[19] is in enabling **real-time, interactive synthesis of layouts**, even for **large scale** scenes, which **were previously intractable** via traditional optimization-based approaches[24]. This work was later used to create a dataset of 3D spaces, which is widely used today for multiple deep learning tasks[5]. In addition to the above, we developed a deep-learning generative transformer approach for creating spaces in scale[11]. Given an input dataset of furnished 3D spaces[5], our method creates 10K new interior layouts in a few seconds. To achieve this, we formulate the synthesis task to be similar to text sequence generation. In this manner, an interior is represented by a series of tokens, where the goal of the neural network is to generate a sequence of tokens that represent a realistic space. Our methodology is not limited to interior spaces, and can be easily be reformulated for architectural plans, city layouts, map generation and other domains[8,4].

Virtual Content Compatibility. Objects may not always fit a space, regardless of their function, due to differing styles. For example, an old-fashioned furniture piece will be out of place in a modern interior. To understand this question better, we used user feedback[8], and deep-learning for estimating room and 3D object style[1,23]. To accomplish this, we used: (i) images where each object appears in multiple "styled" backgrounds, in contrast to existing work which relies on images of objects presented in an isolated context, or solely on object geometry, (ii) comparison labels, which encode the style spectrum of each object in a relative manner. For example, a label indicates whether one table is more modern than another table. Most recently, we are working on understanding style compatibility via joint image-geometry data (Fig. 4).

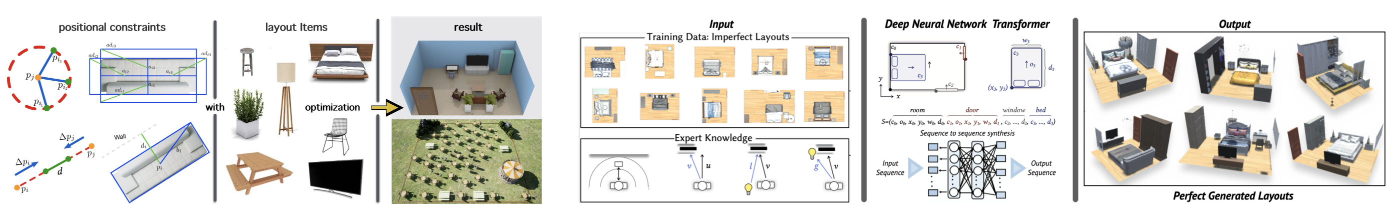


Figure 3: Left: Our results[19] in 3D scene synthesis are faster by an order of magnitude than previous probabilistic methods, and allow **real-time, interactive synthesis of scenes which were previously intractable**. Given input positional constraints and selected layout items, the algorithm rapidly outputs a variety of synthesized layouts. **Right**: We proposed a data-driven method for layout synthesis that combines interior expert knowledge with a data-driven generator based on a Deep Neural Network Transformer architecture[11]. The synthesized layouts integrate desirable properties that are not present in the original dataset.

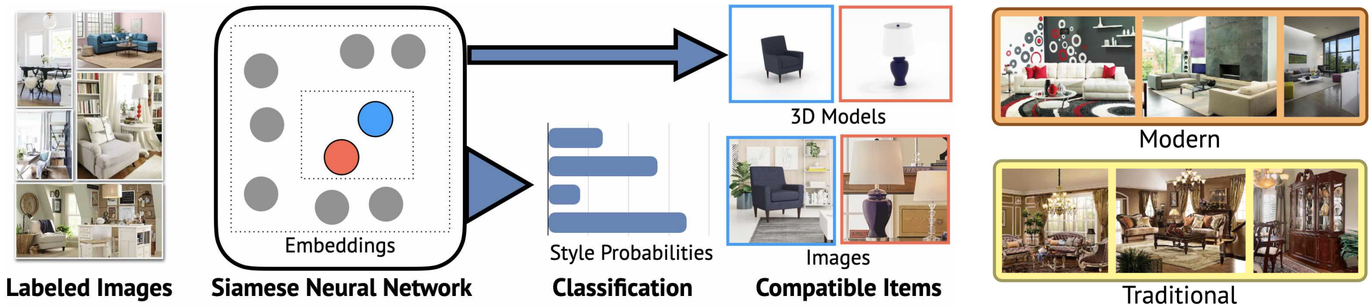
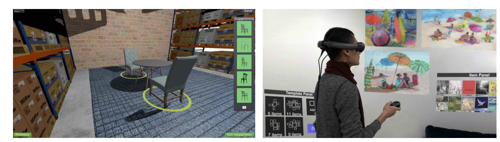


Figure 4: Left: Given style-labeled images of 3D objects and rooms, we trained a siamese network to predict image styles[23]. After the network is trained, the stylistic compatibility between 3D objects is measured by selecting the nearest neighbours within the network's embeddings for each image displaying the 3D object. **Right**: Examples of style labels. Our work can predict the predominant style of a space, which is useful for content creation, design and retail[1].

Future Directions

Multi-model Content Creation and Simulation. The majority of 3D scene synthesis is focused on purely visual elements in terms of 3D geometries[2,24], yet the **physical world** contains other **modalities** such as light, sound[12], smell, airflow, temperature, typical activities, and so on. For example, a train station, and a restaurant have particular aspects that cannot be captured only with 3D geometry. Thus, a scene synthesis model should **capture** such qualities. Our goal is twofold: first, to find such multimodal scene representations, so an end user should only have to provide a high-level description for creating scenes. Second, explore how we can manipulate scenes with an abstract instruction, such as — "clean and organize the apartment". The main challenges would be with data collection, scene understanding, inferring user intent, and in defining a coupling relationship between each of the modalities, and objects embedded in a scene.



System for designing style compatible spaces[8,23].

Virtual Environments for Workforce Training. As they continue to improve, virtual technologies will be **useful** for **training** in multiple domains, including first responders, medical staff, designers, technicians, and others. Virtual environments, as well as **simulation** of workplace activities within these environments, can be generated for each training scenario. Once the training requirements, expectations, and action space are defined, I intend to propose computational methods for creating believable, dynamic training environments. Collaboration with experts from each domain would allow enough granularity in terms of the requirements, tasks, and possible outcomes for each training scenario, which would allow to **procedurally generate and simulate** the specific training environments.

3. Funding

We received internal funding at my current institution for future work listed at Section 1. Currently, we are applying to:

NSF. Our research relates the future of work call, and to RI and HCC tracks at CISE, in respect to Sections 1 and 2.

DoD. Our crowd simulation work (Sec. 1) addresses the call for *"learning for control of autonomous dynamical systems"*.

Industry. Our research is directly applicable to the gaming, robotics, and retail industries. For example, our results in procedural content generation (Sec. 2) accelerate the development of gaming content. Our real-time simulation results (Sec. 1) can be directly incorporated into current game engine frameworks. I will also build on our industry experience and contacts (Amazon, Autodesk Research, and Wayfair) for future industry collaboration opportunities.

References

[1] Ataer-Cansizoglu, E., Liu, H., Weiss, T., Mitra, A., Dholakia, D., Choi, J.-W., and Wulin, D. Room style estimation for style-aware recommendation. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2019), IEEE, pp. 267–2673.

[2] Chaudhuri, S., Ritchie, D., Wu, J., Xu, K., and Zhang, H. Learning generative models of 3d structures. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 643–666.

[3] Fahad, M., Chen, Z., and Guo, Y. Learning how pedestrians navigate: A deep inverse reinforcement learning approach. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2018), IEEE, pp. 819–826.

[4] Farooki, H., Ataer-Cansizoglu, E., Choi, J.-W., and Weiss, T. Interactive and scalable layout synthesis with design templates. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2020), IEEE, pp. 235–238.

[5] Fu, H., Cai, B., Gao, L., Zhang, L.-X., Wang, J., Li, C., Zeng, Q., Sun, C., Jia, R., Zhao, B., et al. 3d-front: 3d furnished rooms with layouts and semantics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 10933–10942.

[6] Helbing, D., Buzna, L., Johansson, A., and Werner, T. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation science 39*, 1 (2005), 1–24.

[7] Helbing, D., Farkas, I., and Vicsek, T. Simulating dynamical features of escape panic. *Nature 407*, 6803 (2000), 487.

[8] Huang, H., Zhang, Y., Weiss, T., Perry, R. W., and Yu, L.-F. Interactive design of gallery walls via mixed reality. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2020), IEEE, pp. 17–26.

[9] Kang, B., Jie, Z., and Feng, J. Policy optimization with demonstrations. In *International conference on machine learning* (2018), PMLR, pp. 2469–2478.

[10] Karamouzas, I., Skinner, B., and Guy, S. Universal power law governing pedestrian interactions. *Phys Rev Lett 113* (Dec 2014), 238701.

[11] Leimer, K., Guerrero, P., Weiss, T., and Musialski, P. Layoutenhancer: Generating good indoor layouts from imperfect data. In *SIGGRAPH Asia Conference Proceedings* (2022).

[12] Liang, B. S., Liang, A. S., Roman, I., Weiss, T., Duinkharjav, B., Bello, J. P., and Sun, Q. Reconstructing room scales with a single sound for augmented reality displays. *Journal of Information Display* (2022), 1–12.

[13] M.Nakada, H.Chen, T.Zhao, T.Weiss, and D.Terzopoulos. Deep learning of neuromuscular control for pointing tasks in a biomechanical human model. In *ACM Transactions on Graphics (SIGGRAPH)* (2018).

[14] Pelechano, N., Allbeck, J., Kapadia, M., and Badler, N. *Simulating Heterogeneous Crowds with Interactive Behaviors*. CRC Press, 2016.

[15] Schwartz, M., Weiss, T., Ataer-Cansizoglu, E., and Choi, J.-W. Style similarity as feedback for product design. In *A New Perspective of Cultural DNA*. Springer, 2021, pp. 27–42.

[16] Sharma, R., Weiss, T., and Kallmann, M. Plane-based local behaviors for multi-agent 3d simulations with position-based dynamics. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2020), IEEE, pp. 214–217.

[17] Talukdar, B., and Weiss, T. Learning anticipatory multi-agent navigation with position-based deep reinforcement learning. *IEEE Robotics and Automation Letters. Under Review.* (2022).

[18] Weiss, T. Implementing position-based real-time simulation of large crowds. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2019), IEEE, pp. 306–3061.

[19] Weiss, T., Litteneker, A., Duncan, N., Jiang, C., Yu, L.-F., and Terzopoulos, D. Fast and scalable position-based layout synthesis. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* (2018).

[20] Weiss, T., Litteneker, A., Duncan, N., Jiang, C., Yu, L.-F., and Terzopoulos, D. Fast, scalable layout synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2018), IEEE. Extended abstract.

[21] Weiss, T., Litteneker, A., Jiang, C., and Terzopoulos, D. Position-based multi-agent dynamics for real-time crowd simulation. In *ACM SIGGRAPH Conference on Motion in Games* (2017), MIG '17, pp. 10:1–10:8. **Best Paper Award**.

[22] Weiss, T., Litteneker, A., Jiang, C., and Terzopoulos, D. Position-based real-time simulation of large crowds. *Computers & Graphics* (2018).

[23] Weiss, T., Yildiz, I., Agarwal, N., Ataer-Cansizoglu, E., and Choi, J.-W. Image-driven furniture style for interactive 3d scene modeling. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 57–68.

[24] Zhang, S.-H., Zhang, S.-K., Liang, Y., and Hall, P. A survey of 3d indoor scene synthesis. *Journal of Computer Science and Technology 34*, 3 (2019), 594–608.