UNIVERSITÀ DEGLI STUDI DI GENOVA

ISTITUTO ITALIANO DI TECNOLOGIA
PATTERN ANALYSIS AND COMPUTER VISION

RePAIR

FAIR FUTURE AI RESEARCH

DiffAssemble: A Unified Graph-Diffusion Model for 2D and 3D Reassembly

Gianluca Scarpellini*[1,2], Stefano Fiorini*[1], Francesco Giuliari*[1,2]
Pietro Morerio[1], Alessio Del Bue[1]

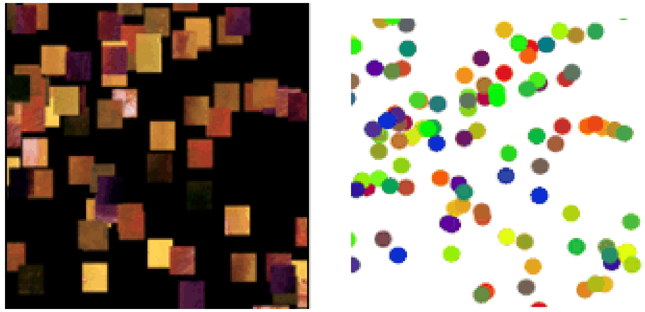


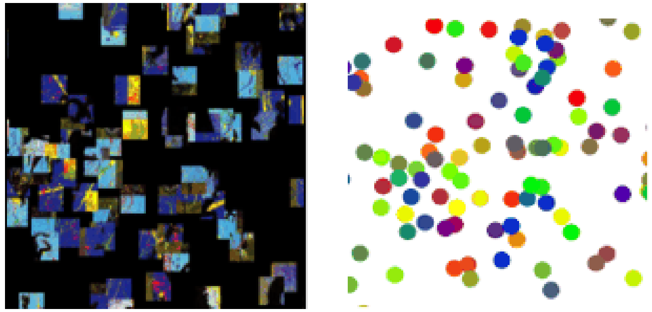


*Equal Contribution
[1]PAVIS, Istituto Italiano di Tecnologia (IIT)
[2]Università degli Studi di Genova

2D & 3D Reassembly Tasks

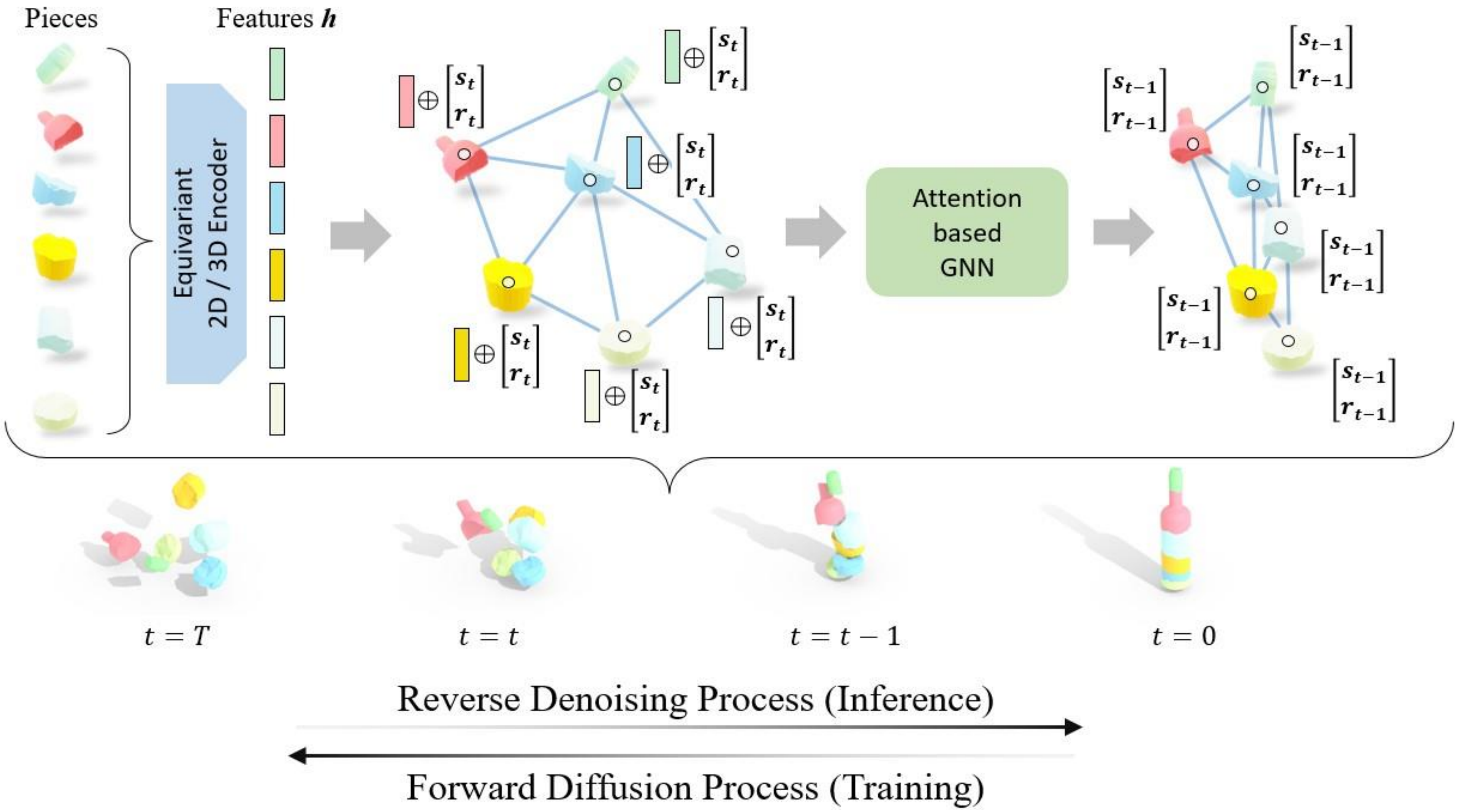
Problem. Placing each individual component in its correct position and orientation to form a coherent structure





DiffAssemble

General framework for solving reassembly tasks using graph representations and a diffusion model formulation

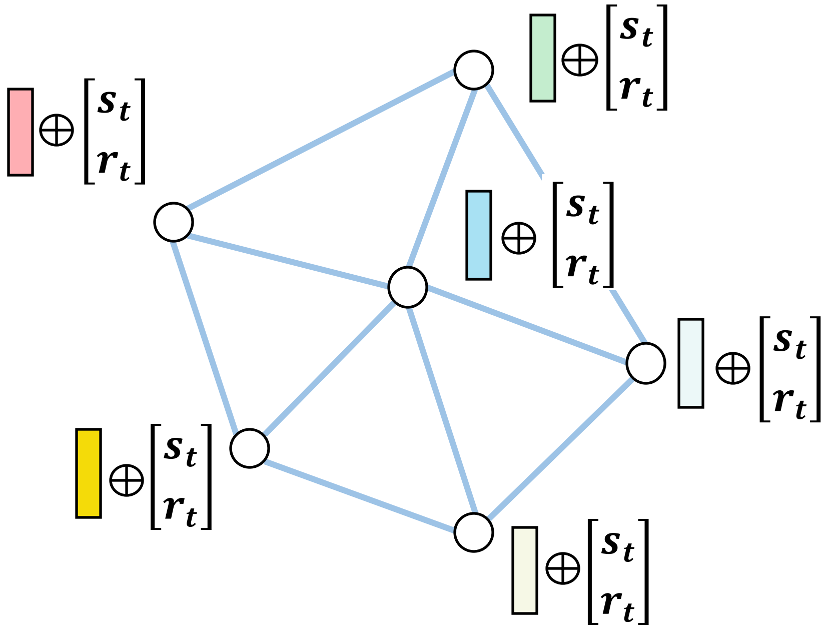


Graph Formulation

Features $\mathbf{h} \in \mathbb{R}^d$ → Features generated by a backbone

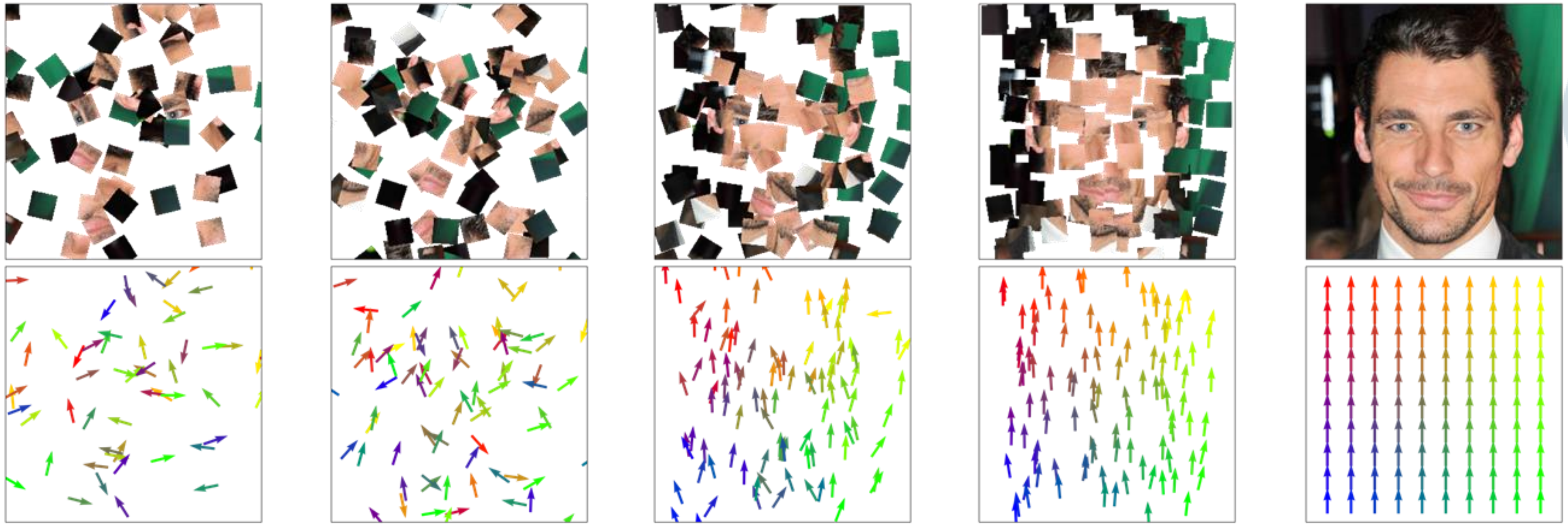
Position $\mathbf{s} \in \mathbb{R}^n$ → Represent the dimensionality of the continuous Euclidean space

Rotation $R \in SO(n)$ → Represent the matrix belonging to the Special Orthogonal Group in *n* dimensions. We also define **r**, where $f_r(\mathbf{r}) = R$.



The Key Point of Using Diffusion

Create **random starting scenarios** and learn how to reverse this process **step by steps**



t = T | t = 100 | t = 40 | t = 10 | t = 0

Reverse Denoising Process (Inference)

Forward Diffusion Process (Training)

3D Reassembly Task: Breaking Bad[1]

Number of Objects. Contains around 10k meshes from PartNet and Thingi10k.

Pieces. Number of re-compute 20 fracture modes and then simulate 80 fractures from them, resulting in a total of 1,047,400 breakdown patterns.

Subsets. *Everyday*, *Artifact* and *Other* to facilitate different applications.

[1] Sellán, Silvia, et al. "Breaking bad: A dataset for geometric fracture and reassembly." *Advances in Neural Information Processing Systems* 35 (2022).

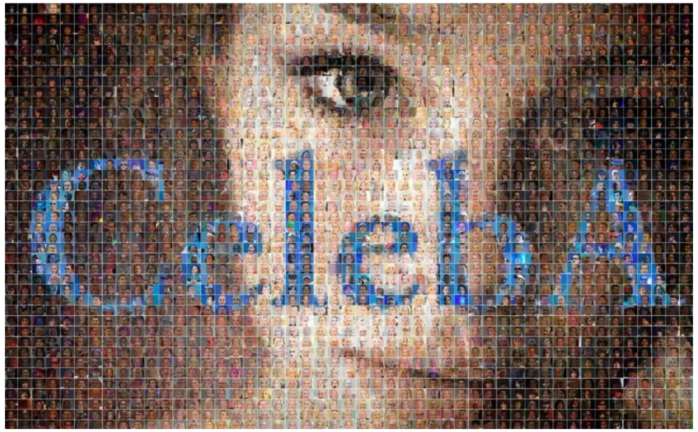
3D Reassembly Task: Results

METHOD	RMSE (R) ↓ degree	RMSE (T) ↓ $\times 10^{-2}$	PA ↑ %
Global [34]	81.6	15.2	17.5
DGL [34]	81.4	<u>14.9</u>	<u>25.4</u>
LSTM [34]	87.4	15.8	11.3
SE(3)-Equiv [46]	<u>77.9</u>	16.7	8.1
DiffAssemble - No Diffusion Process	83.6	17.1	3.1
DiffAssemble - No Equivariant Enc.	81.7	17.0	18.3
DiffAssemble	**73.3**	**14.8**	**27.5**

Insights

- No trade accuracy between rotation and translation
- Benefit in deploying the Diffusion Process and the Equivariant Backbone

2D Reassembly Task: Dataset



CelebA-HQ. Contains 30K images of celebrities in High Definition (HD). The images are cropped and positioned to show only centered faces.[1]

WikiArt. Contains 63K images of paintings in HD. This dataset contains paintings with very different content and artistic styles.[2]

[1] Lee, Cheng-Han, et al. "Maskgan: Towards diverse and interactive facial image manipulation." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
[2] Tan, Wei Ren, et al. "Improved ArtGAN for conditional synthesis of natural image and artwork." *IEEE Transactions on Image Processing* 28.1 (2018).

2D Reassembly Task: Results

Method		PuzzleCelebA 6x6	PuzzleCelebA 8x8	PuzzleCelebA 10x10	PuzzleCelebA 12x12	PuzzleWikiArts 6x6	PuzzleWikiArts 8x8	PuzzleWikiArts 10x10	PuzzleWikiArts 12x12
Optimization Based	Gallagher [15]	80.21	55.18	71.19	69.81	71.88	61.63	54.15	44.68
	Yu *et al.* [48]	98.63	<u>94.65</u>	98.33	93.33	**94.62**	**92.95**	**90.99**	**89.88**
	Huroyan *et al.* [21]	98.47	**97.45**	98.65	<u>97.08</u>	<u>92.69</u>	<u>91.37</u>	<u>89.74</u>	<u>88.28</u>
Learning Based	**DiffAssemble** - No Diff.	<u>99.43</u>	79.84	<u>99.05</u>	91.28	73.07	54.70	22.68	18.27
	DiffAssemble - No Equiv.	96.12	71.62	91.98	64.15	25.31	14.63	8.19	4.96
	DiffAssemble	**99.51**	87.66	**99.30**	**97.76**	90.65	72.79	63.33	53.08

Method	CelebA 6x6	CelebA 12x12	WikiArts 6x6	WikiArts 12x12
Gallagher [15]	33.28 *(-46.93)*	19.18 *(-50.63)*	32.19 *(-39.69)*	24.12 *(-20.56)*
Yu [21]	<u>33.45</u> *(-66.85)*	<u>21.78</u> *(-72.84)*	<u>32.53</u> *(-62.09)*	<u>24.65</u> *(-65.23)*
Huroyan [48]	18.18 *(-80.29)*	0.09 *(-88.45)*	17.14 *(-75.55)*	0.08 *(-80.28)*
DiffAssemble	**96.92** *(-2.59)*	**76.49** *(-32.81)*	**51.21** *(-39.44)*	**27.09** *(-25.99)*

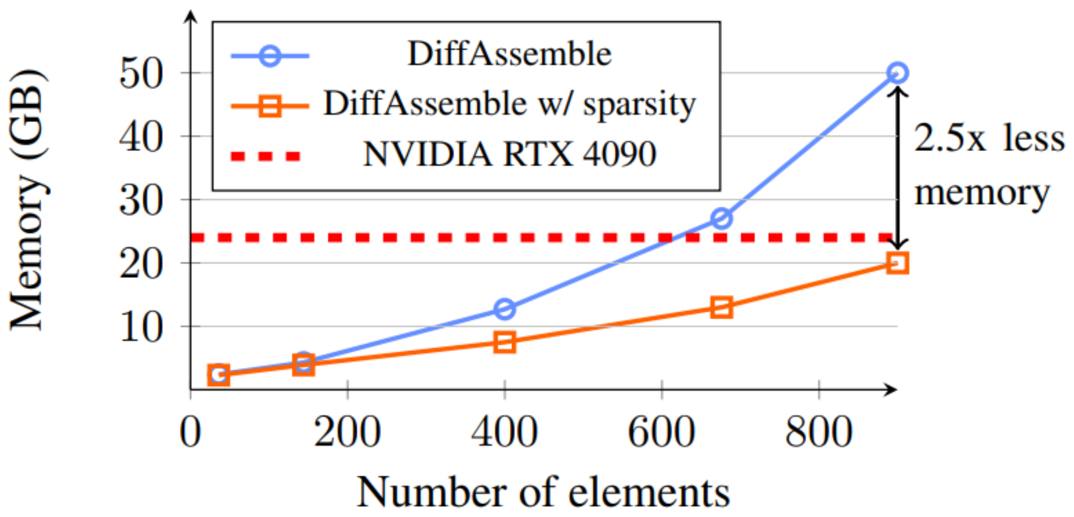
Insights

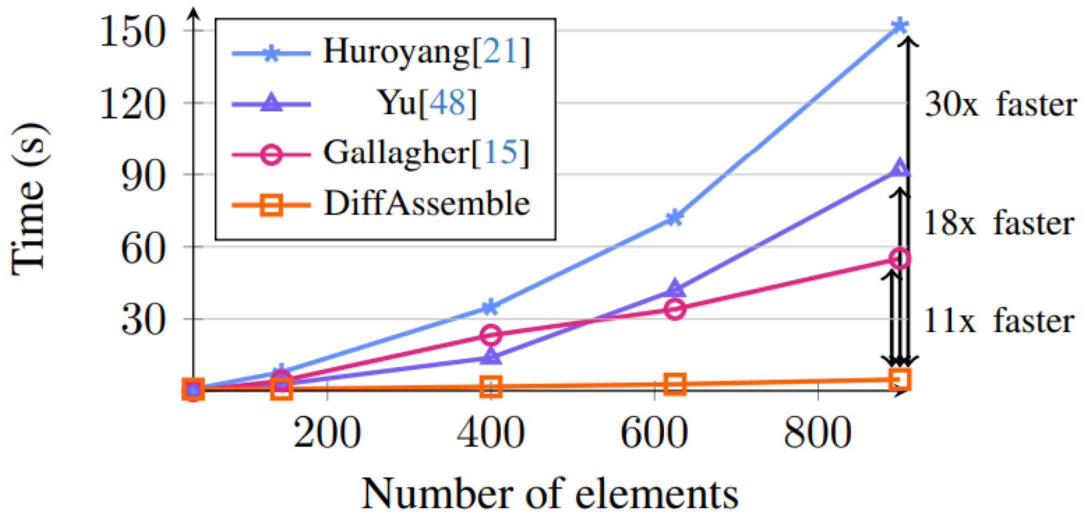
- We do not rely only on visual appearances but also on the semantic content
- We are robust to real-scenarios like missing pieces

Scaling to Larger Graphs: Results

Dataset. PuzzleCelebA

Implementation Details. Puzzles of 900 pieces (30 x 30 puzzles)





Insights

- The sparsity attention mechanism reduces 2.5x the GPU memory

Insights

- Faster than optimization-based model
- No reduction in Accuracy

Conclusion & Future Work

- Introduction of **DiffAssemble,** a general framework for tackling sorting tasks via graph representations and a diffusion model formulation
- Demonstration of the effectiveness of our method spanning 3D object reassembly and 2D puzzles with translated and rotated pieces:
  - Robustness compared to optimization-based methods
  - State-of-the-art results in 3D domain
  - Possibility to scale on larger graphs

This work is part of the RePAIR project, funded by the European Union's Horizon 2020 program, and also supported by the project Future Artificial Intelligence Research (FAIR)