ISTITUTO ITALIANO DI TECNOLOGIA
PATTERN ANALYSIS AND COMPUTER VISION

GANzzle: Reframing jigsaw puzzle solving as a retrieval task using a generative mental image

Davide Talon [1, 2] Alessio Del Bue [1] Stuart James [1]

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

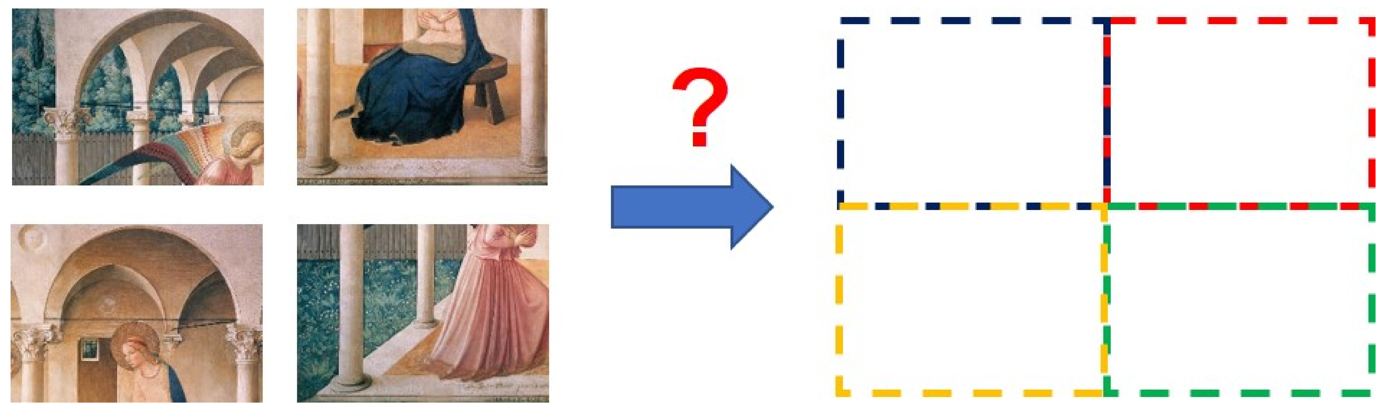
UniGe

SCAN ME

1. The Problem

Given a set of pieces (non-overlapping squares patches of an image) in a random configuration to infer the correct permutation to recover the original image.

Challenge: To overcome the combinatorially complex of matching adjacent pieces

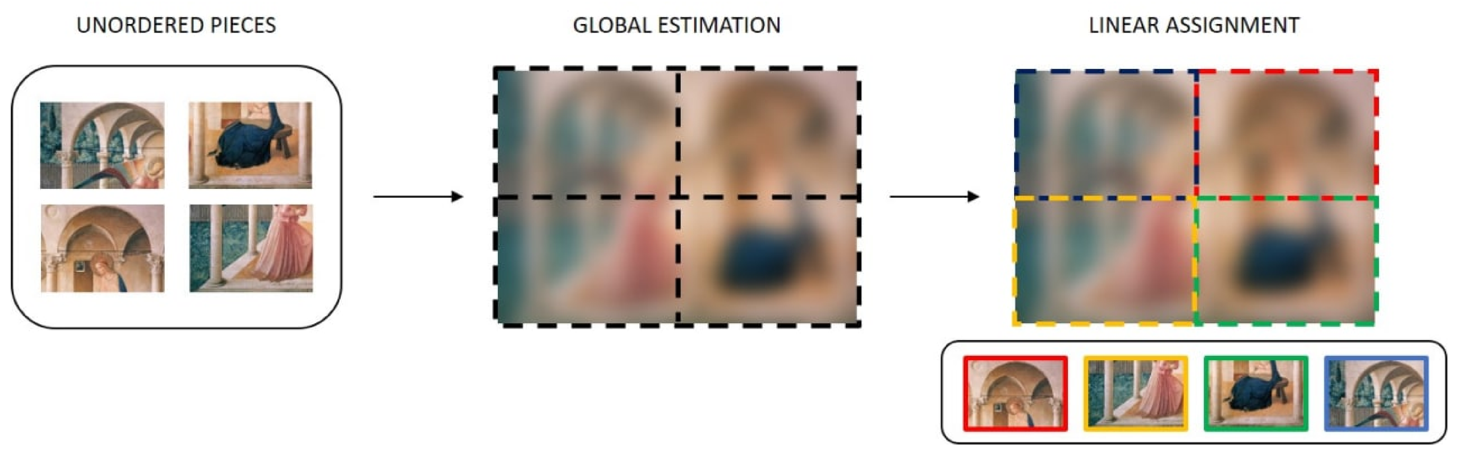


Prior work: Solving for adjacent pieces, *optimization-based* approaches [1] are time demanding and sensitive to initialization seed and erosion. *Deep Learning* approaches [7] are faster but do not generalize to multiple sizes.

2. Contribution

Exploiting advancements in Generative Adversarial Network (GAN) methods, we learn to estimate a global solution (mental image) to the problem from unordered pieces. Therefore, we frame the problem as a $R@1$ retrieval task, and then solve the linear assignment using differentiable Hungarian Attention [7].

TLDR; Estimate the global solution (mental image). Match pieces against it.



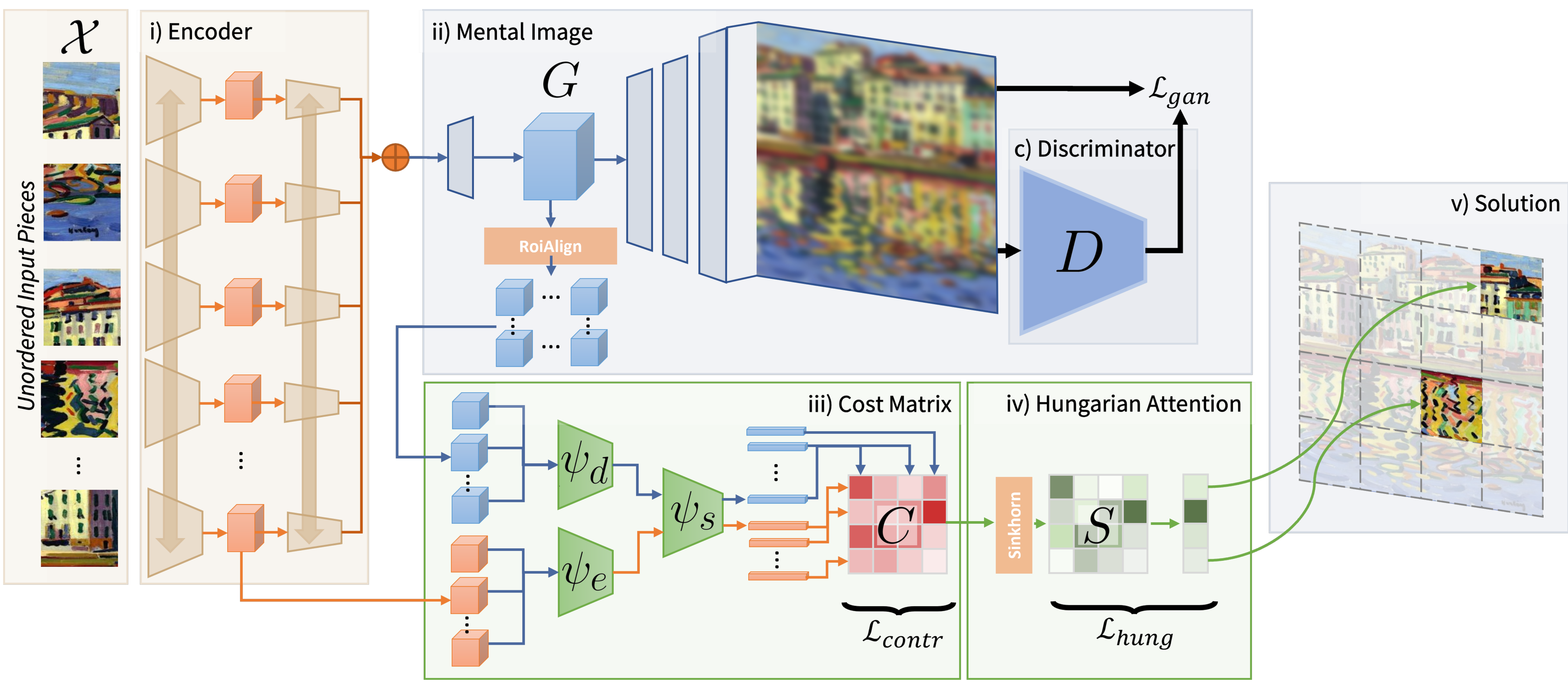
- A **many-to-one GAN Architecture** for recovering a global image from its pieces.
- **Dynamic size puzzle solver** using Hungarian attention and contrastive loss.
- **Two new large-scale puzzle solving datasets**, named *PuzzleCelebA* [3] and *PuzzleWikiArts* [6], permutations are available for direct comparison.

3. Dataset

The proposed benchmark builds on (*left*) PuzzleCelebA (30k images) and (*right*) PuzzleWikiArts (63k images), providing permutation of pieces at different puzzle sizes.

4. Architecture

Unordered pieces ($\mathcal{X}$) are independently encoded and then pooled to produce a latent vector, from which the global solution (mental image) is estimated. After RoiAlign, generator intermediate features are used as global representation slots to be matched against. A cost matrix evaluates the piece to global slot similarity. Hungarian attention solves for the final permutation.



5. Key blocks

Many-to-one GAN: unordered pieces are encoded independently. The set of encodings is merged using average pooling to generate a single encoding vector. Then, a decoder projects back to the image space trained using adversarial and reconstruction loss for Multi-Scale Gradients [2].

Cost matrix and contrastive loss: the similarity matrix is computed as dot product of all possible piece-slot pairs. A contrastive loss enforces the feature space to have similar embeddings for piece-slot correct pairs while pushing apart non-corresponding pairs:

$$\mathcal{L}_{contr} = -\mathbb{E}_i \left[\log \frac{exp\left(\psi_s^i \cdot \psi_s^j / \tau\right)}{exp(\psi_s^i \cdot \psi_s^j/\tau) + \sum_{k \neq j} exp(\psi_s^i \cdot \psi_s^k/\tau)} \right] \quad (1)$$

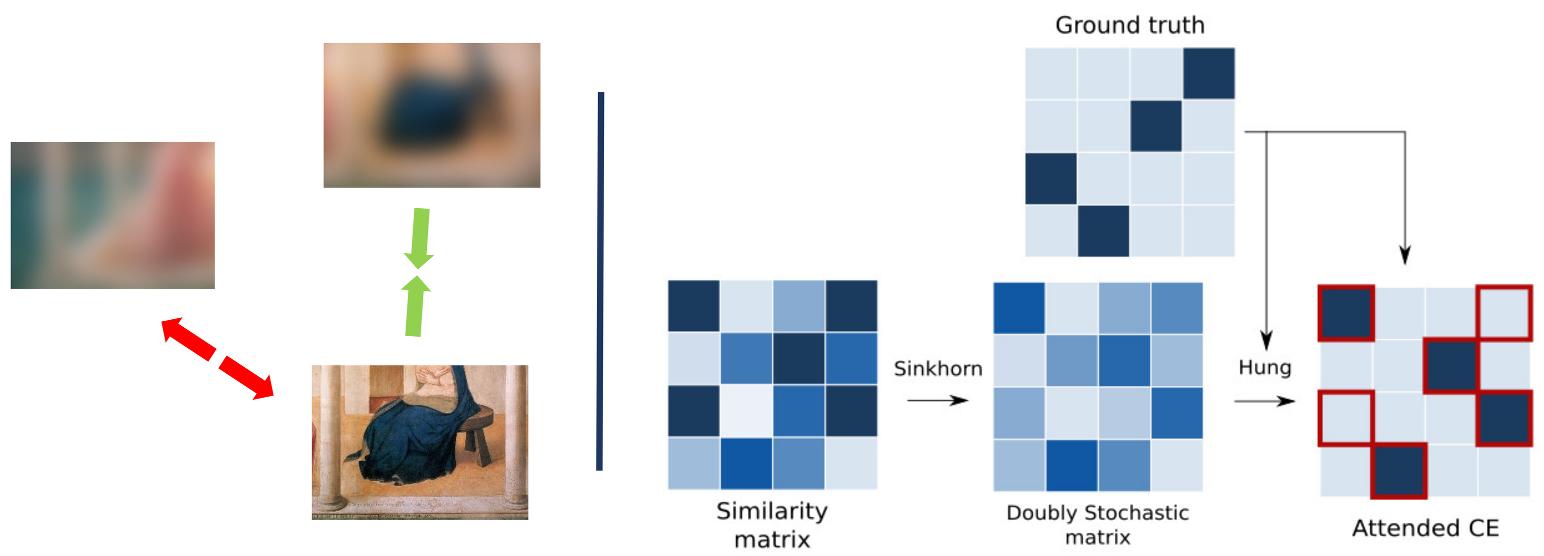
with ψ_s^i and ψ_s^j embeddings of considered piece i and j its corresponding slot.

Hungarian attention (HA): enables supervised learning of optimum assignments. The Sinkhorn normalization relaxes the cost matrix C to a doubly stochastic matrix $\mathbf{S}$. Here, both correct and misplaced pieces are attended:

$$\mathbf{Z} = OR\left(\mathrm{Hung}\left(\mathbf{S}\right), \mathbf{S}^G\right). \quad (2)$$

Binary cross-entropy loss with respect to the ground-truth $\mathbf{S}^G$ assignment matrix is attended through the mask $\mathbf{Z}$:

$$\mathcal{L}_{hung} = \sum_{i,j \in [n]} \mathbf{Z}_{ij} \left(\mathbf{S}_{ij}^G \log \mathbf{S}_{ij} + \left(1 - \mathbf{S}_{ij}^G\right) \log\left(1 - \mathbf{S}_{ij}\right)\right), \quad (3)$$



(a) Contrastive Loss (b) Hungarian Attention

6. Results

Our method handles different sizes while performing on par with deep learning single-size approaches in terms of direct comparison accuracy.

Dataset	PuzzleCelebA				PuzzleWikiArts			
	6x6	8x8	10x10	12x12	6x6	8x8	10x10	12x12
Paikin and Tal [4]	99.12	98.67	98.39	96.51	98.03	97.35	95.31	90.52
Pomeranz et al. [5]	84.59	79.43	74.80	66.43	79.23	72.64	67.70	62.13
Gallagher [1]	98.55	97.04	95.49	93.13	88.77	82.28	77.17	73.40
PO-LA [8]	71.96	50.12	38.05	-	12.19	5.77	3.28	-
Hung-perm	33.11	12.89	4.14	2.18	8.42	3.22	1.90	1.25
GANzzle-Single (Ours)	71.00	51.81	**43.74**	-	11.78	6.23	**8.97**	-
GANzzle (Ours)	**72.18**	**53.26**	32.84	**12.94**	**13.48**	**6.93**	4.10	**2.58**

Qualitative results of GANzzle for 10×10 on (*top*) PuzzleCelebA and (*bottom*) PuzzleWikiPaintings

Limitations emerge with challenging pieces, i.e., pieces with similar content, as they can be interchangeable, however GANzzle is able to resolve for the structure of the image.

Take home message

- It is possible to achieve state-of-the-art performances for puzzle problems of different sizes with a single trained model.
- Two benchmark datasets suitable for recent deep learning approaches are available for ease of comparison.
- Deep learning approaches are still far from optimization-based algorithms.

References

[1] A. C. Gallagher. Jigsaw puzzles with pieces of unknown orientation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[2] Animesh Karnewar and Oliver Wang. Msg-gan: Multi-scale gradients for generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[3] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[4] Genady Paikin and Ayellet Tal. Solving multiple square jigsaw puzzles with missing pieces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[5] D. Pomeranz, M. Shemesh, and O. Ben-Shahar. A fully automated greedy square jigsaw puzzle solver. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[6] W. R. Tan, C. S. Chan, H. E. Aguirre, and K. Tanaka. Improved artgan for conditional synthesis of natural image and artwork. *IEEE Transactions on Image Processing*, 2019.

[7] Tianshu Yu, Runzhong Wang, Junchi Yan, and Baoxin Li. Learning deep graph matching with channel-independent embedding and hungarian attention. In *International conference on learning representations (ICLR)*, 2019.

[8] Yan Zhang, Jonathon Hare, and Adam Prügel-Bennett. Learning representations of sets through optimized permutations. In *International conference on learning representations (ICLR)*, 2019.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964854.

RePAIR