Gromov Wasserstein Optimal Transport for Semantic Correspondences

Francis Snelgar[1,2], Ming Xu[1], Stephen Gould[1], Liang Zheng[1], Akshay Asthana[2]
1) School of Computing, Australian National University, Australia
2) Seeing Machines, Australia

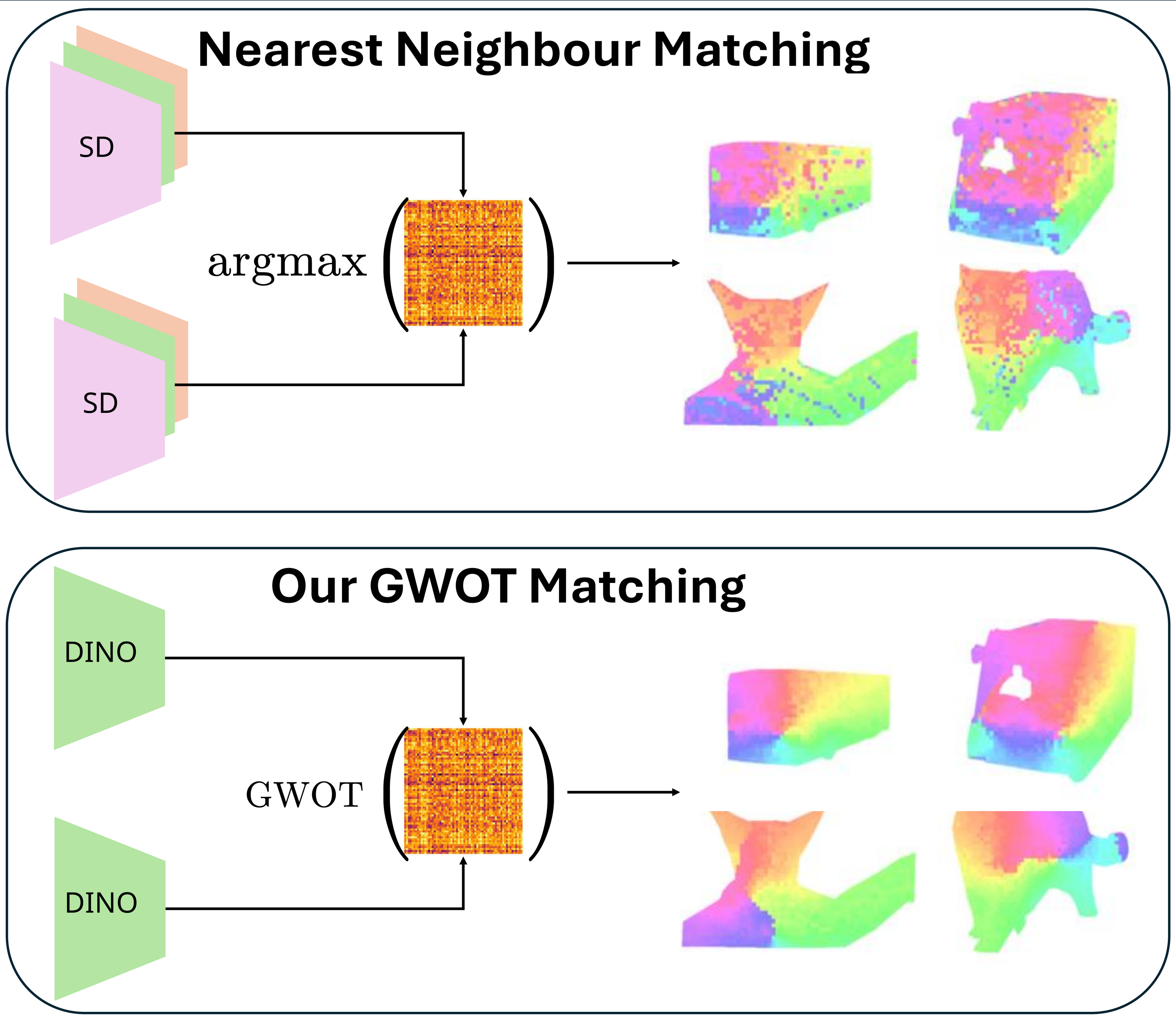
BMVC 2025

Motivation

- Foundation models provide excellent feature descriptors for semantic matching
- However matches from the standard nearest neighbour's matcher can be spatially noisy and sparse
- Recent works address this with an ensemble of foundation models
- We instead use an optimal transport based matcher with a Gromov Wasserstein spatial smoothness prior
- Our matcher boosts the performance of a DINOv2 baseline to be competitive with state-of-the-art ensemble methods
- Our matcher is 5-10x more efficient than ensemble methods

Key Takeaway: Matching algorithms with structural priors can boost performance over nearest neighbour algorithms, while being more efficient than model ensembles.

Pipeline



Matching Algorithm

$$\overbrace{T^\star}^{1} = \operatorname*{argmin}_{T} \overbrace{\langle D^p D^{q\top}, T\rangle}^{2} + \overbrace{\langle C^q T C^p, T\rangle}^{3} + \sum_{\substack{(i,k)\in\mathcal{G}\\(j,l)\in[N]}} \overbrace{S_{i,k}\hat{S}_{j,l}T_{i,j}T_{k,l}}^{4} + \overbrace{\mathrm{D}_{\mathrm{KL}}(T^\top \mathbf{1}_N || \frac{1}{N}\mathbf{1}_N)}^{5}$$

1. Transport Map
The solution or *transport map* provides a mapping between image patches

2. Feature Similarity Cost
The OT cost is between *patches of different images* - matching patches should have 'similar' DINOv2 features

3. Spatial Consistency Cost
The GWOT cost is between *pairs of patches from the same image* - a pair of patches that are 'close' in one image should not match a pair of patches that are 'far' in the second image

$$C^p_{i,k} = \begin{cases} 1, & ||x_i - x_k||_2 < \delta_{\min} \\ 0, & \text{otherwise} \end{cases}$$

$$C^q_{j,l} = \begin{cases} 1, & ||\hat{x}_j - \hat{x}_l||_2 > \delta_{\max} \\ 0, & \text{otherwise.} \end{cases}$$

C^p, C^q are sparse and implemented efficiently with 2D convolutions with kernel $k = \mathbf{1}_{(\delta\times\delta)}$

4. Symmetry Cost
Ordering of symmetric patches should be preserved e.g. the order of the 'left ear' and 'right ear' should be the same in both images

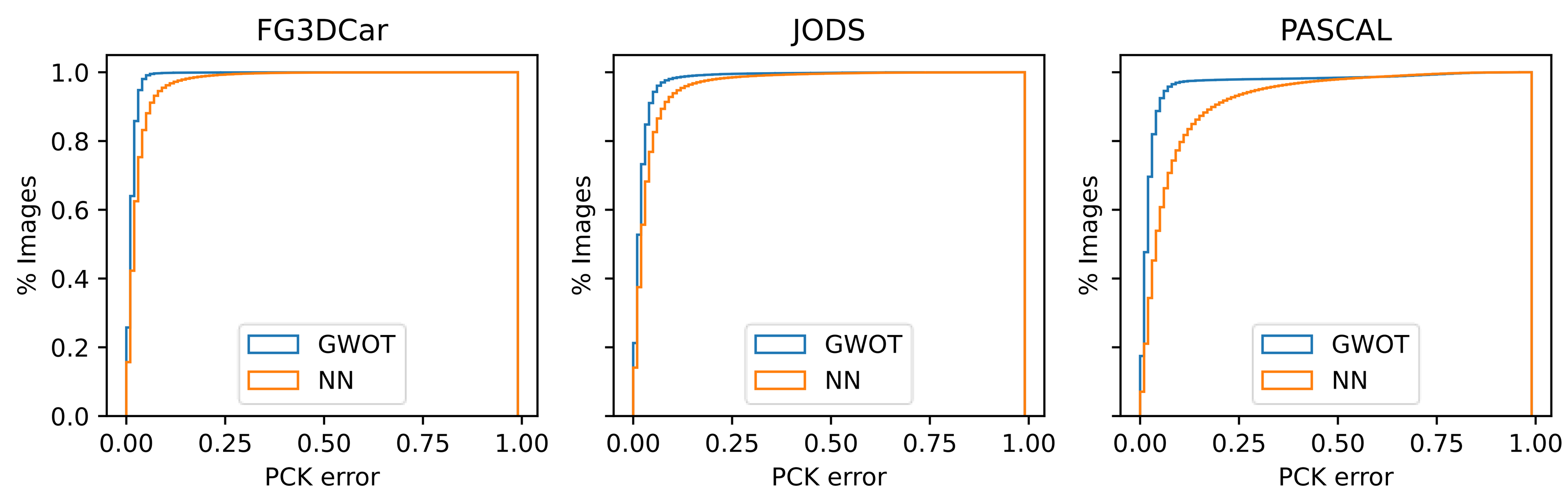
$$S_{i,k} = \operatorname{sign}\left(x_i^{(0)} - x_k^{(0)}\right)$$

5. Unbalanced Transport
one-to-one matching constraint is replaced with a KL divergence penalty – encourages one-to-one matching but does not require it e.g., different sized objects

Results

Experimental Results

- Our GWOT matcher improves over NN matcher at all PCK thresholds
- State-of-the-art performance on TSS dataset, competitive on PASCAL and Spair datasets



Failure Cases

1. Different sized objects
   - Optimal transport assumes balanced assignment
   - Incorrect smoothing due to fixed size GW kernels
2. Symmetry failure
   - Symmetry assumption is not always valid