Image Feature Consensus with Deep Functional Maps

Xinle Cheng[1], Congyue Deng[2], Adam W. Harley[2], Yixin Zhu[1,3], Leonidas Guibas[2]

[1]Institute for AI, Peking University, China [2]Department of Computer Science, Stanford University, USA [3]PKU-WUHAN Institute for Artificial Intelligence, China

Preliminary: Functional Map

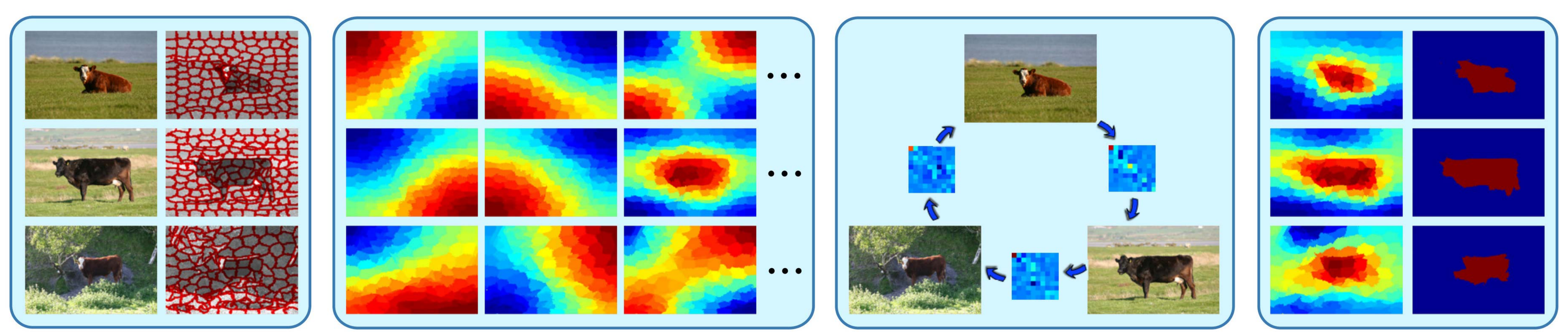
Represent dense correspondences in the function space
Basis: Eigenfunctions of the Laplacian-Beltrami operator
A function can be represented as a linear combination of the basis functions: $f = \sum_i a_i \varphi_i^{\mathcal{M}}$

$$T_F(f) = T_F\left(\sum_i a_i \varphi_i^{\mathcal{M}}\right) = \sum_i a_i T_F(\varphi_i^{\mathcal{M}})$$

A bijective mapping between the two spaces becomes a linear mapping in the function space:

$$T_F(f) = \sum_i a_i \sum_j c_{ij} \varphi_j^{\mathcal{N}} = \sum_h \sum_i a_i c_{ij} \varphi_j^{\mathcal{N}}$$

Preliminary: Functional Map for Images?



Images don't naturally have manifold structure 😕
How to construct the manifold/graph structure on images?

Method Overview

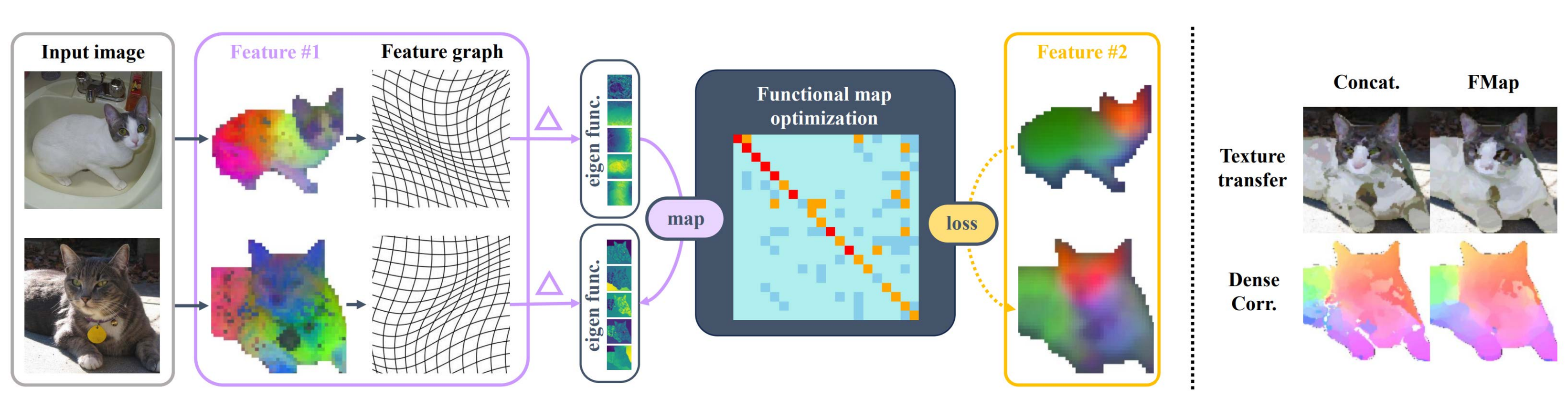


Fig. 1: Overview. Left: Given two sets of features, E^M, E^N, and F^M, F^N, we compute the Laplacian eigenfunction basis with E^M, E^N, and apply regularizations to the functional map optimization using F^M, F^N. This method optimizes a mapping in the spectral domain derived from one feature set to achieve a *consensus* with the other set. **Right:** With a better understanding of the global image structure, our method produces smoother and more accurate correspondences in a zero-shot manner.

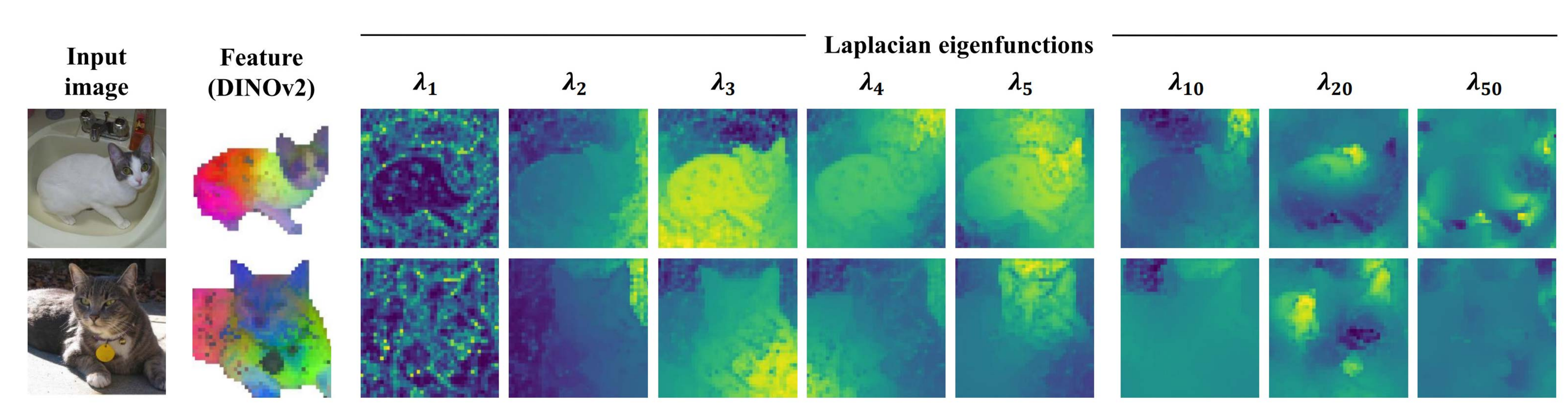
Method

Image Laplacian from Feature #1
Construct weighted graph among image pixels

$$\|e_{\mathbf{xy}}\| = \exp\left(-\frac{\|E_{\mathbf{x}}^M - E_{\mathbf{y}}^M\|}{\sigma}\right)$$

Feature #2 as function regularizer $\tilde{F}^M = \varphi^M g_{\mathcal{R}}(F^M) \quad \tilde{F}^N = \varphi^N g_{\mathcal{R}}(F^N)$

$$\mathcal{L}_{\text{feat}} = \|\mathbf{C}\tilde{F}^M - \tilde{F}^N\|_2$$



Method: Additional Constraints

Diagonality regularization: $\mathcal{L}_{\text{diag}} = \left(\left|\lambda_i^M - \lambda_j^N\right| c_{ij}\right)^2$

the magnitudes of eigenvalues reflect the frequencies of the corresponding eigenfunctions
eigenfunctions of similar frequencies are more likely to be related

Bijectivity (consistency) regularization: $\mathcal{L}_{\text{cons}} = \left\|\mathbf{C}\mathbf{Z}^M - \mathbf{Z}^N\right\|_2$

Overall objective

$$\mathcal{L} = \mathcal{L}_{\text{feat}} + \lambda_{\text{diag}}\mathcal{L}_{\text{diag}} + \lambda_{\text{cons}}\mathcal{L}_{\text{cons}}$$
$$+ \lambda_Z\left(\text{tr}\left((\mathbf{Z}^M)^t\mathbf{W}\mathbf{Z}^M\right) + \text{tr}\left((\mathbf{Z}^N)^t\mathbf{W}\mathbf{Z}^N\right)\right)$$
$$+ \lambda_{\text{reg}}\left(\left\|(\mathbf{Z}^M)^t\mathbf{Z}^M - \mathbf{I}\right\|_2 + \left\|(\mathbf{Z}^N)^t\mathbf{Z}^N - \mathbf{I}\right\|_2\right)$$

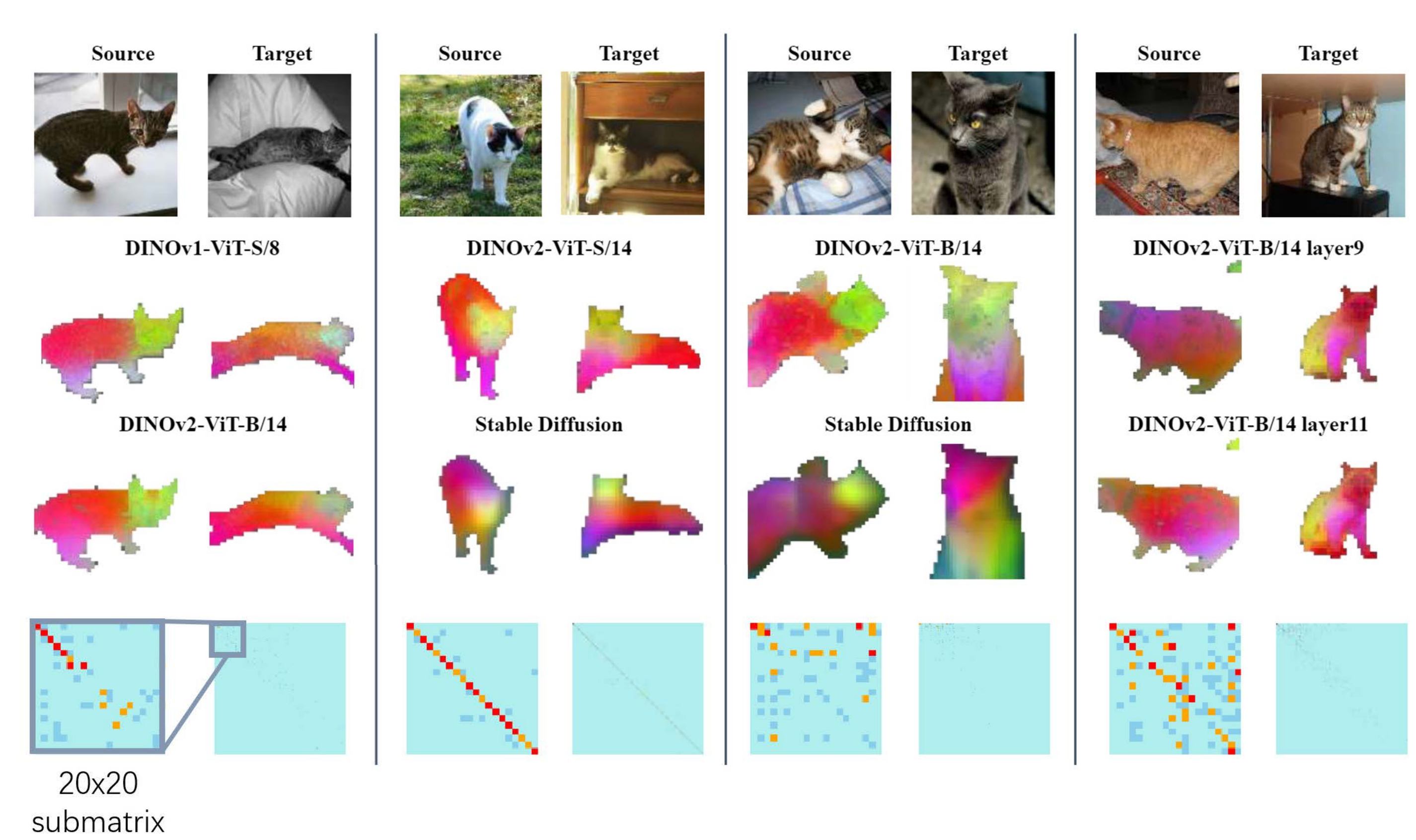
Resultant Functional Maps

Input images
(840x840 each)

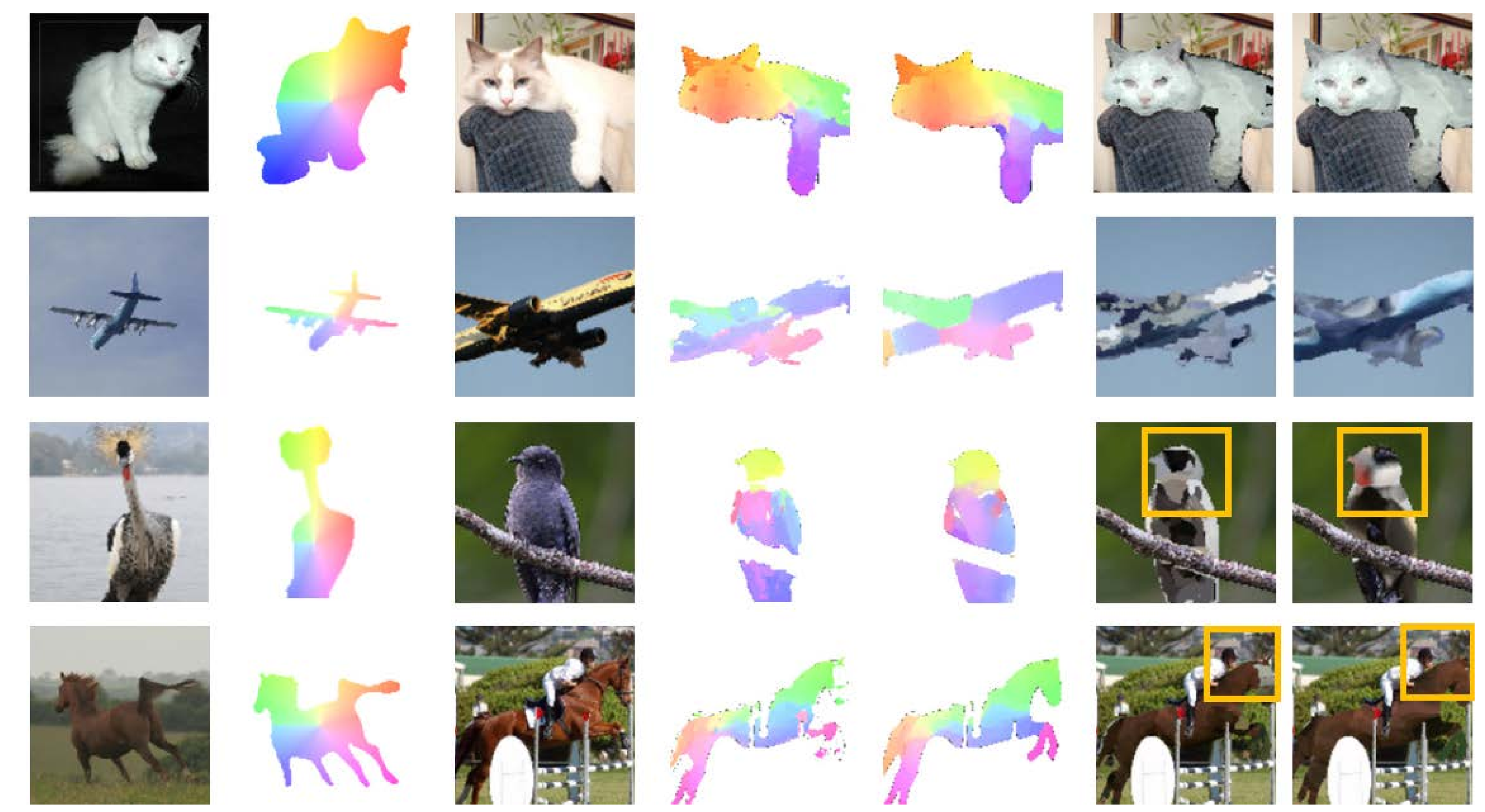
Feature #1
(patch feat. interp. to pixels)

Feature #2
(patch feat. interp. to pixels)

Functional map
(200x200)



Dense Correspondence



Keypoint Correspondences (Spair-71k)

Main Results on TSS

Setting	Method	FG3DCar	JODS	Pascal	Avg.
Supervised	SCOT [23]	95.3	81.3	57.7	78.1
	CATs* [7]	92.1	78.9	64.2	78.4
	PWarpC-CATs* [49]	95.5	85.0	85.5	88.7
Unsupervised task-specific	CNNGeo [33]	90.1	76.4	56.3	74.4
	PARN [15]	89.5	75.9	71.2	78.8
	GLU-Net [46]	93.2	73.3	71.1	79.2
	Semantic-GLU-Net [48]	95.3	82.2	78.2	85.2
Unsupervised zero-shot	DINOv1-ViT-S/8 [1]	68.7	44.7	36.7	52.7
	DINOv2-ViT-B	81.2	68.4	51.5	69.4
	Stable Diffusion (SD)	92.1	62.6	48.4	72.5
	Concat. DINOv2 + SD [55]	92.9	73.8	59.6	78.7
	FMap DINOv2(basis) + DINOv2(loss)	83.5	69.2	52.7	71.0
	FMap SD(basis) + SD(loss)	80.0	63.4	51.5	67.8
	FMap DINOv2(basis) + SD(loss) **(ours)**	84.8	70.4	53.5	72.2
	FMap DINOv2(loss) + SD(basis) **(ours)**	**93.1**	**74.0**	**59.9**	**78.9**

References

Functional maps: a flexible representation of maps between shapes. In: ACM TOG 31(4), 1–11(2012)

A tale of two features: Stable diffusion complements dino for zero-shot semantic correspondence. In: arXiv preprint arXiv:2305.15347 (2023)

Twitter

山大王
中国大陆 上海