

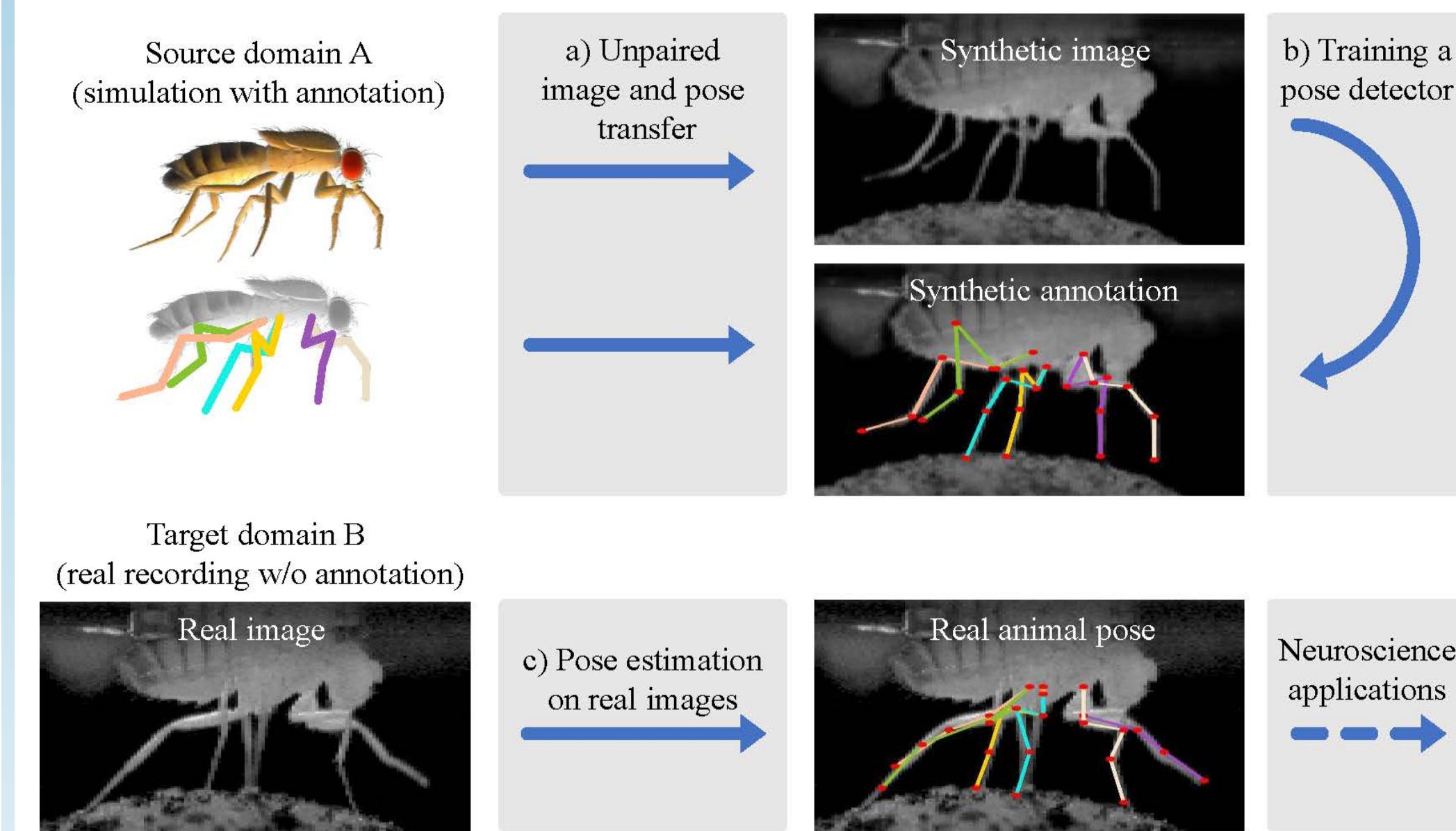
Deformation-aware Unpaired Image Translation for Pose Estimation on Laboratory Animals

Siyuan Li¹, Semih Günel¹, Mirela Ostrek¹, Pavan Ramdy¹, Pascal Fua¹, and Helge Rhodin^{1,2}

¹École polytechnique fédérale de Lausanne ² The University of British Columbia

Problem Definition

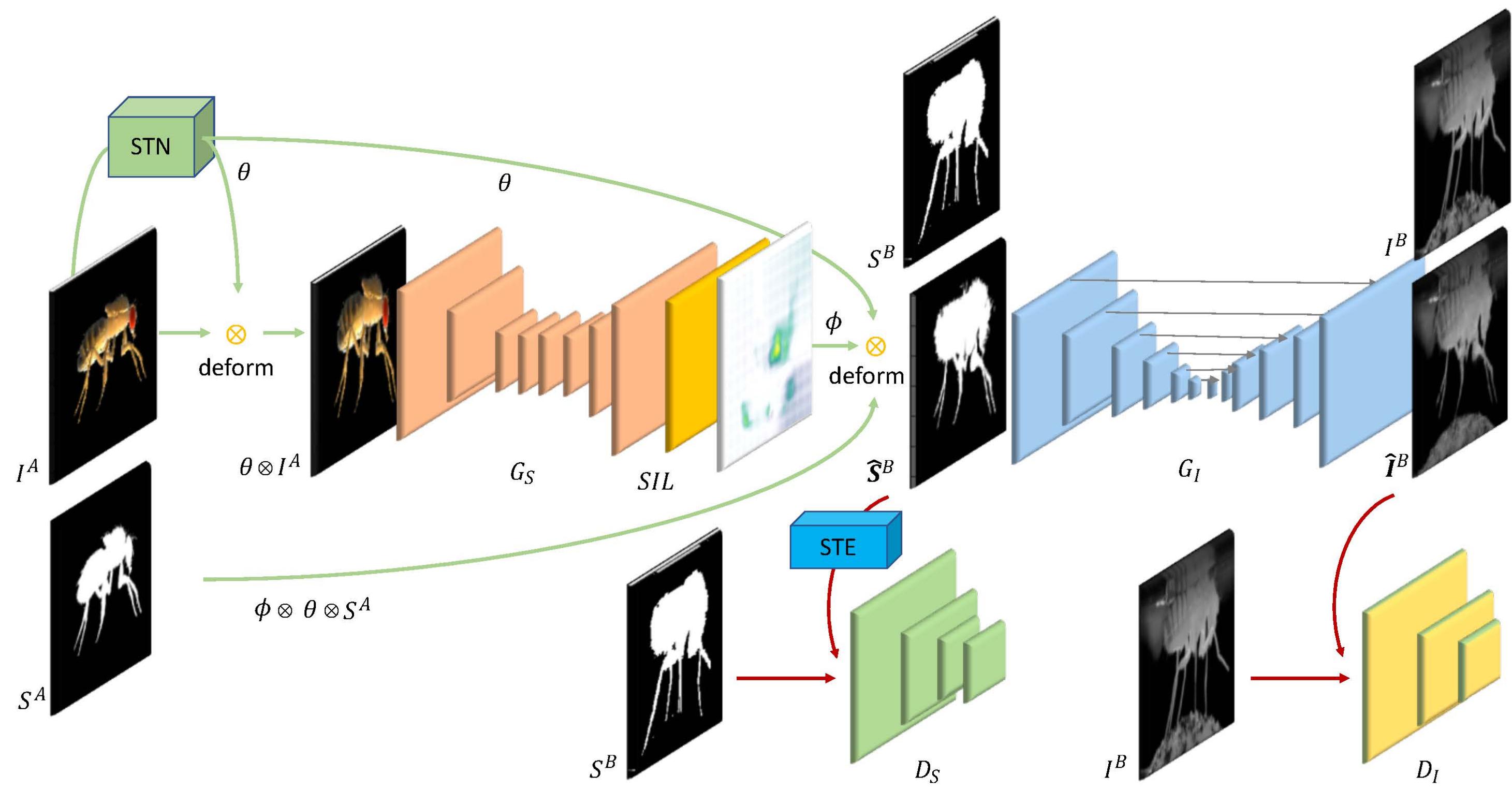
Goal: Capture the pose of real animals using synthetic training examples, **without using any manual annotations**. In order to support the study how neural circuits orchestrate behavior.



Key Contributions:

- An efficient model that generates both realistic appearances and accurate annotations from simple synthetic animal models.
- Explicit and independent modeling of appearance, shape and pose in an unpaired image translation framework.
- Introducing a pixel-wise deformation module that overcomes large structure difference across domains.

Method

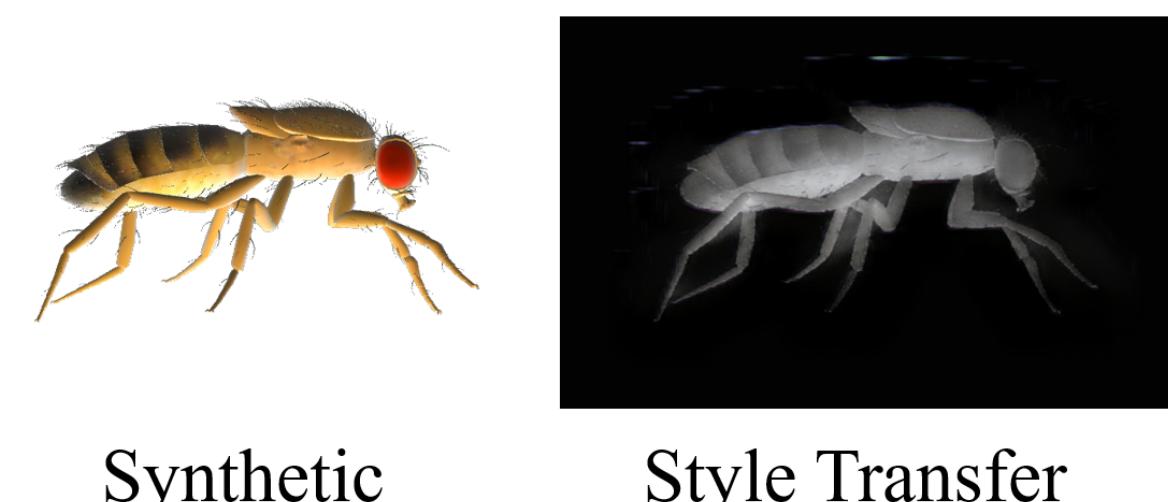


Overview of our deformation-based image translation method. Our model has two steps. In the first step, the deformation from source domain A to target domain B is estimated for input image I^A and its silhouette S^A via network G_S and a Spatial Transformer Network (STN). Their output is an explicit deformation field parameterized by the global, affine transformation θ and a local, non-linear warping ϕ , using a spatial integral layer (SIL). Then, the deformed silhouette is transformed into the full output image I^B with image generator G_I . Discriminators D_S and D_I enable unpaired training. D_S uses the Straight Through Estimator (STE) for backpropagation.

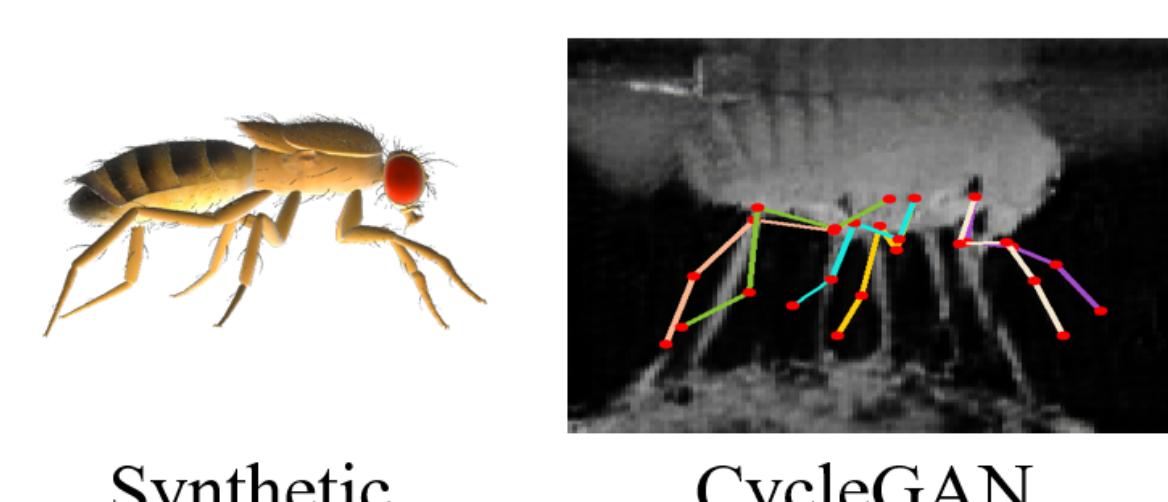
Related Works

Unpaired image translation

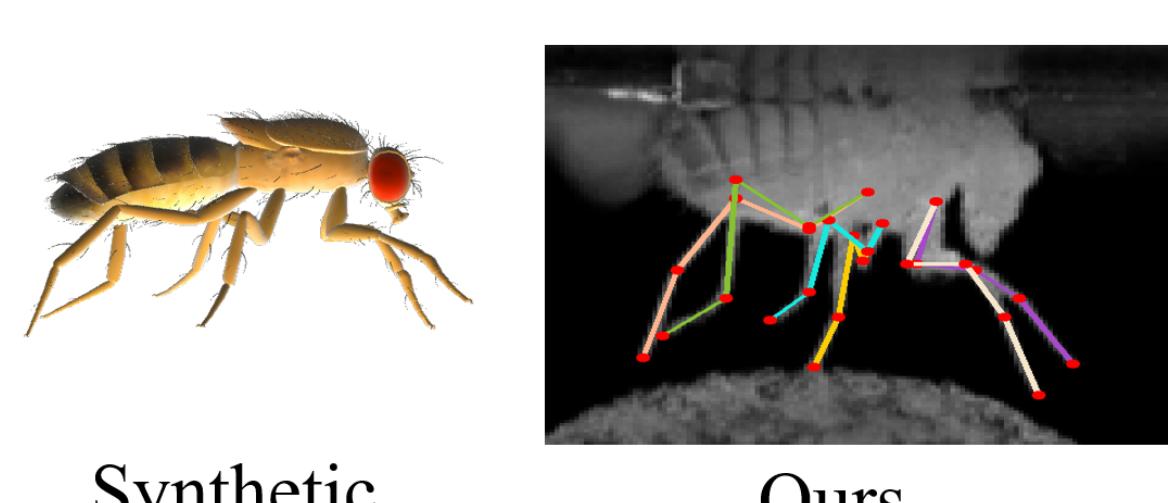
- Style Transfer based methods** keep the pose unchanged fail to generate realistic appearances.



- cycleGAN based methods** generate realistic appearances fail to keep the pose when source and target domain have large structural differences.



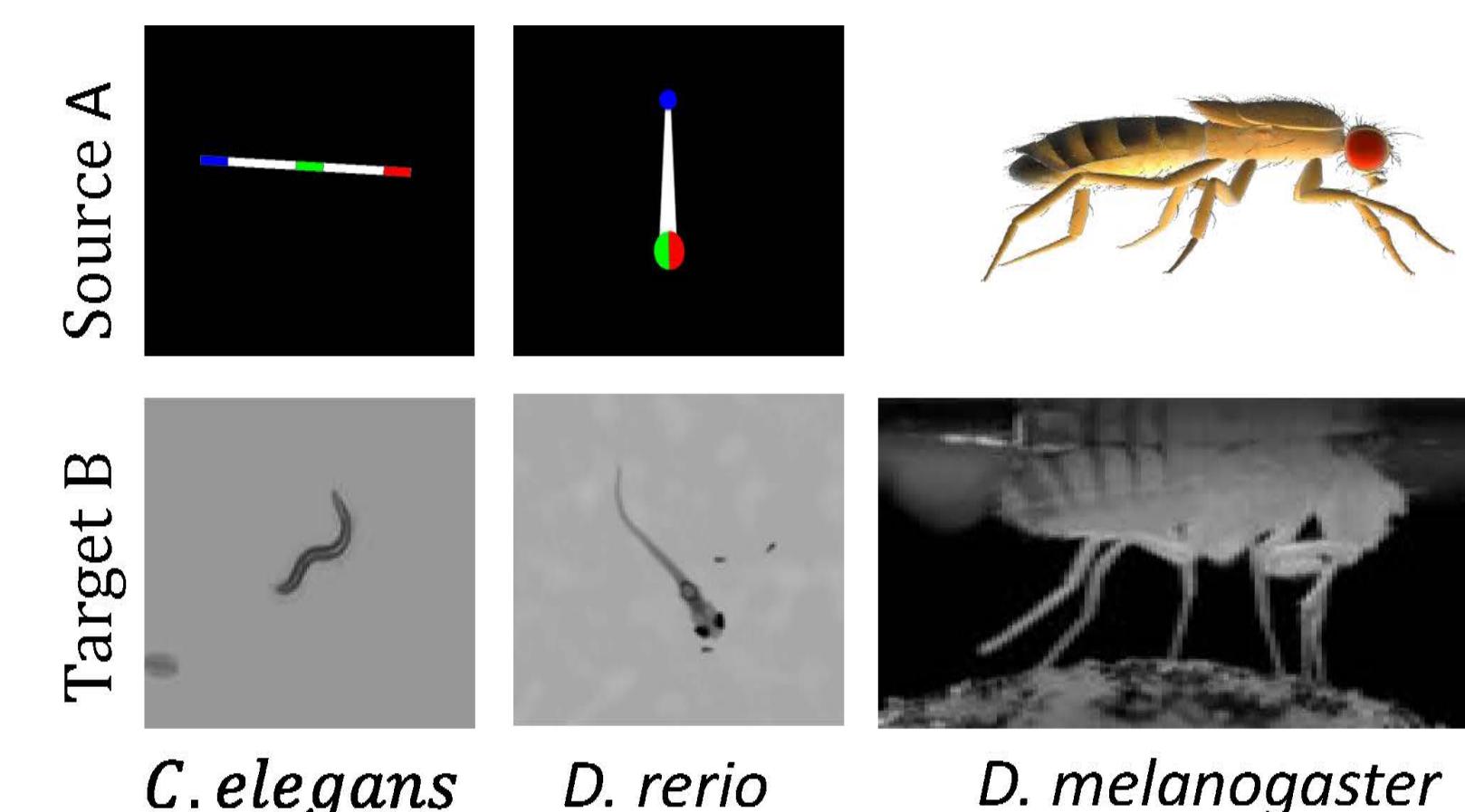
- Our proposed method** can generate realistic images together with accurate annotations.



Experiments: image translation

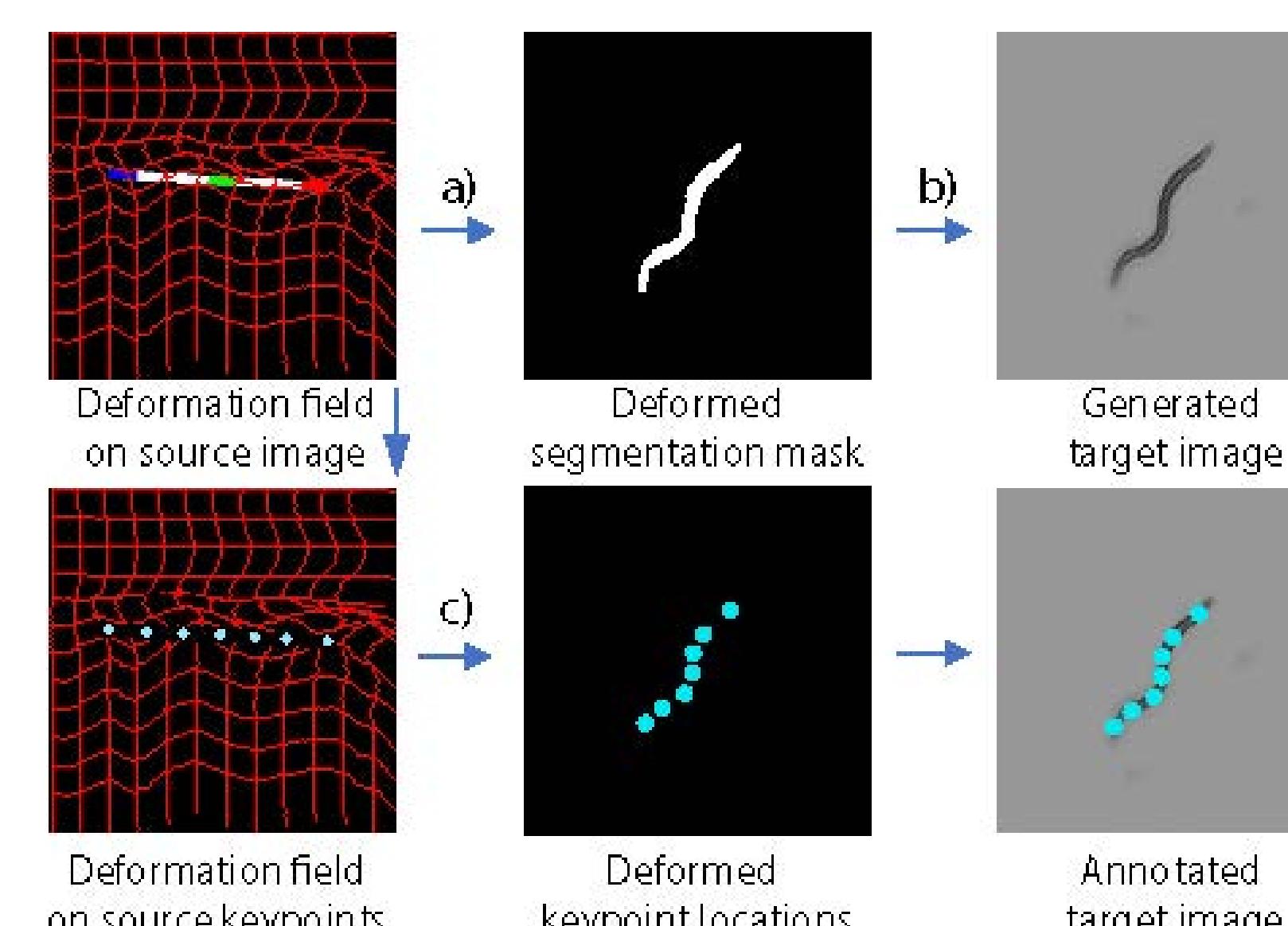
Dataset overview

- Samples of synthetic and real data.** We focus on three animals with large discrepancy in appearance, shape and pose.



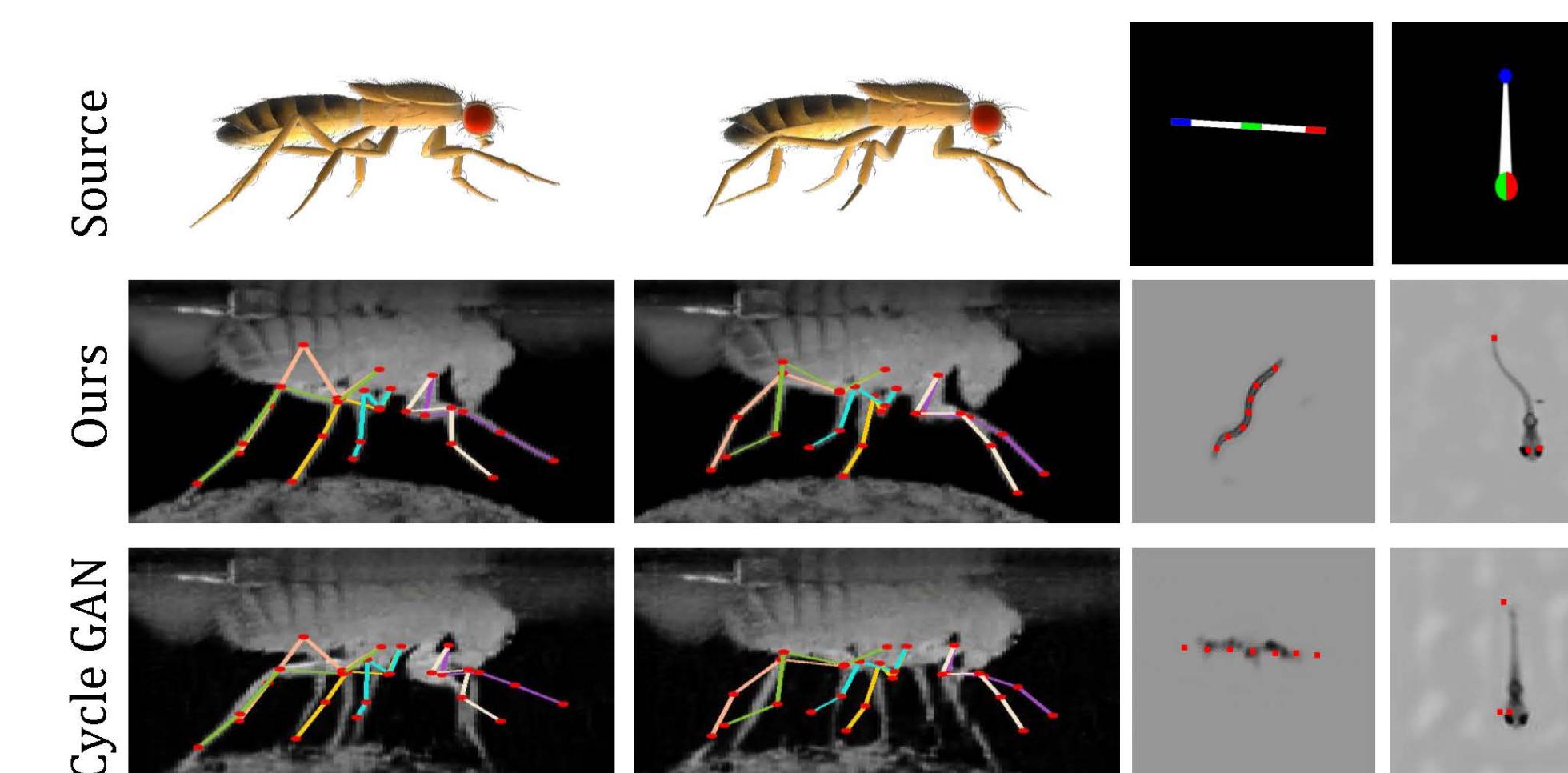
Annotation transfer

- Annotations can be transferred from source to target domain by learned deformation field.



Results

- Qualitative comparison.** Our methods can generate realistic images with accurate annotations in target domain



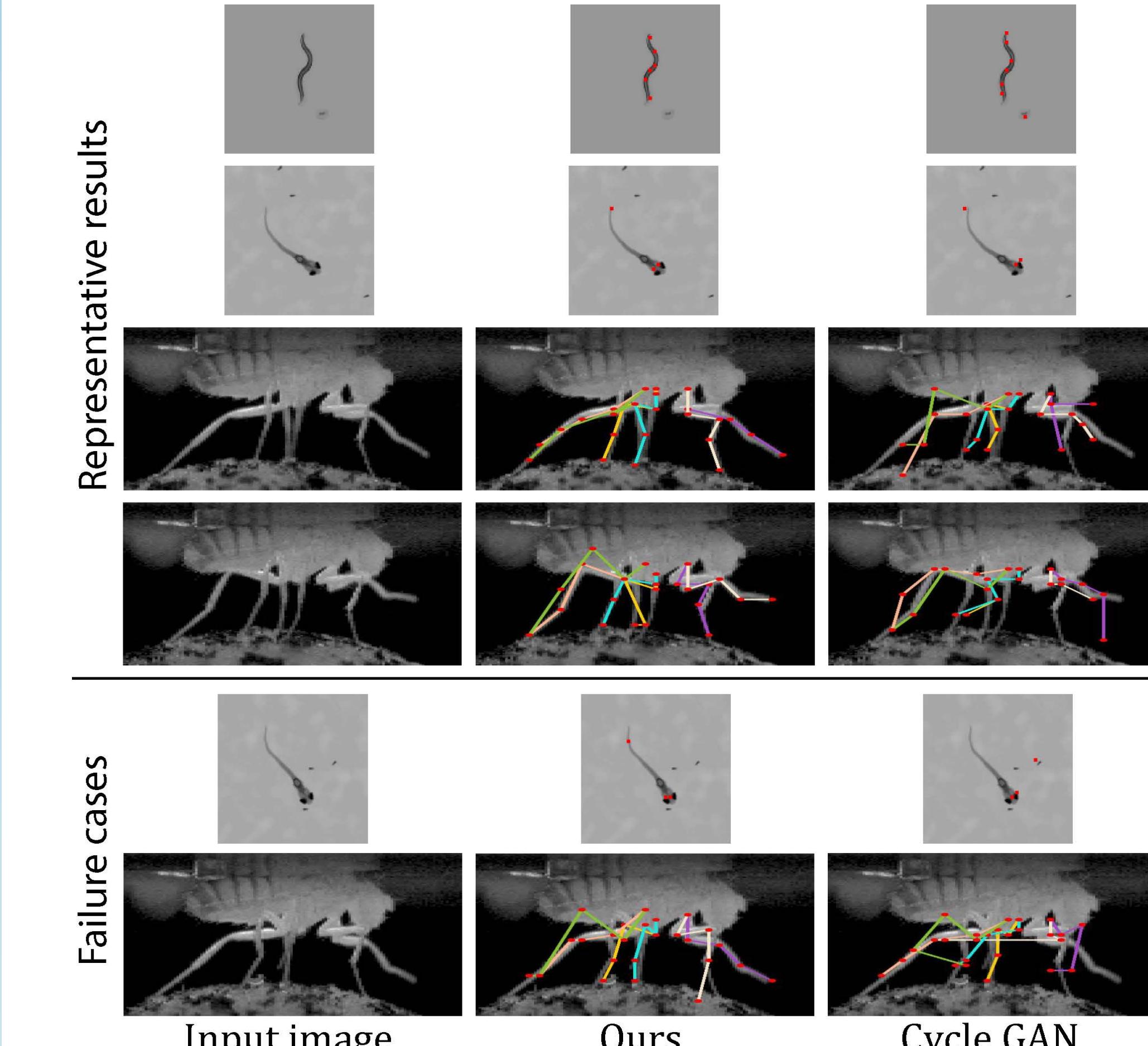
- Structured similarity (SSIM) comparison.**

Task	D.M.	C.E.	D.E.
Fast-Style-Transfer	0.3932	0.0539	0.6385
Cycle-GAN	0.6543	0.9034	0.8504
Gc-GAN	0.6392	0.8915	0.8586
Ours	0.6746	0.9076	0.8771
Supervised	72.2	88.8	90.35

Experiments: pose estimation

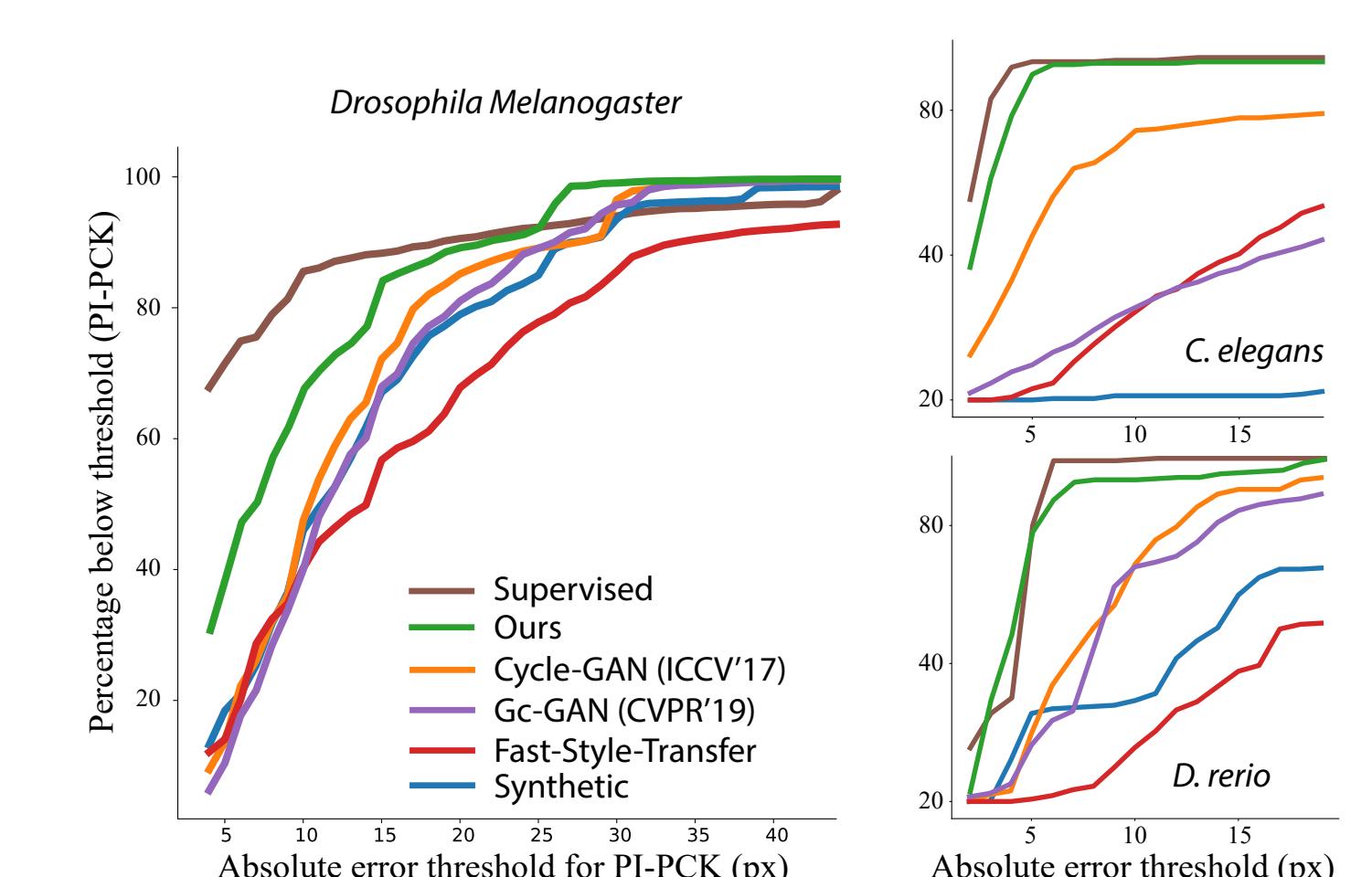
Qualitative results

- The estimator provides close to fully-supervised results across all three animals.



Quantitative results

- Pose estimation accuracy**



- Pose estimation results for Drosophila:

Metric	Drosophila Melanogaster			
	PI-PCK ↑ (5 pix)	PI-PCK ↑ (15 pix)	PI-AUC ↑ (4-45 pix)	PI-RMSE ↓ (pix)
Synthetic	19.8	67.9	75.75	13.456
Fast-Style-Transfer	15.4	57.6	68.9	17.309
Gc-GAN	11.9	68.7	76.3	13.175
Cycle-GAN	15.0	72.9	78.4	12.302
Ours	40.0	84.7	86.0	8.823
Supervised	72.2	88.8	90.35	6.507

- Pose estimation results for C.elegans and D. rerio:

Metric	Caenorhabditis elegans			Danio rerio		
	PI-PCK ↑ (5 pix)	PI-AUC ↑ (2-20 pix)	PI-RMSE ↓ (pix)	PCK ↑ (10 pix)	AUC ↑ (2-20 pix)	RMSE ↓ (pix)
Synthetic	0.0	0.9	67.29	29.3	37.4	20.15
Fast-Style-Transfer	3.1	25.0	20.50	15.6	20.8	19.25
Gc-GAN	9.7	25.0	27.38	68.2	54.5	27.38
Cycle-GAN	45.3	63.2	14.71	68.7	59.1	9.70
Ours	90.3	87.6	5.36	93.9	83.1	4.50
Supervised	94.6	92.3	3.77	99.6	86.5	3.91