

Lecture 7: How to Prepare an AI Poster

https://github.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2025s1

Teerapong Panboonyuen
<https://kaopanboonyuen.github.io>

Crafting a World-Class AI Poster for CVPR, EMNLP, NeurIPS

From Research to Eye-Catching Presentation

Dynamic Conceptional Contrastive Learning for Generalized Category Discovery

Nan Pu, Zhun Zhong, Nicu Sebe

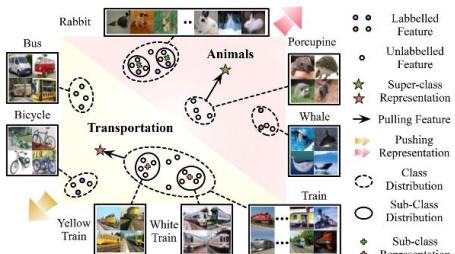
The Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

JUNE 18-22, 2023
CVPR VANCOUVER, CANADA

Highlights

Goal:

Generalized category discovery (GCD) aims to utilize labelled data of seen categories to cluster unlabelled samples from seen and unseen categories.



Challenge:

- GCD needs to **jointly distinguish** the known and unknown categories.
- Discover the novel categories **without any annotations**.

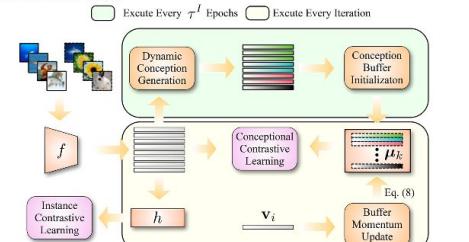
Motivation of Our Solution:

Samples that belong to the same conceptions should be similar to each other in the feature space. The conceptions can be regarded as: classes, super-classes, sub-classes, etc.

Our Solution

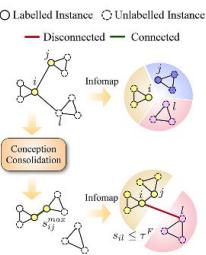
Main Pipeline :

Jointly learning representations and generating dynamic conceptions in an EM-like manner.



Conception Consolidation:

We first Leveraging labelled data to Consolidate the conceptional relationships between samples, and then apply Infomap algorithm to generate rectified conceptions.



$$\mathcal{A}_{ij} = \begin{cases} s_{ij}^{max}, & \text{if } \mathbf{y}_i, \mathbf{y}_j \in \mathcal{Y}^L \text{ and } \mathbf{y}_i = \mathbf{y}_j \\ s_{ij}, & \text{if } \mathbf{y}_i \text{ or } \mathbf{y}_j \in \mathcal{Y}^U \text{ and } s_{ij} > \tau^F \\ 0, & \text{otherwise} \end{cases}$$

$$s_i^{max} = \arg \max_j \{s_{ij} \mid j \in \mathcal{D}\},$$

$$s_{ij} = [(\mathbf{v}_i / \|\mathbf{v}_i\|) \cdot (\mathbf{v}_j / \|\mathbf{v}_j\|) + 1] / 2 \in [0, 1],$$

Experimental Results and Further Exploration

- Achieve a new state-of-the-art performance on three generic and three fine-grained datasets.

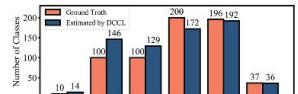
Method	CIFAR-10			CIFAR100			ImageNet-100			Evaluation on Generic datasets			Evaluation on fine-grained datasets				
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	CUB-200	Stanford-Cars	Oxford-Pet		
Kentan	81.6	83.2	82.5	52.6	55.7	51.6	78.1	71.3	74.3	84.9	85.7	82.8	71.1	77.1	70.1	80.7	
RankS+*	46.8	48.2	47.5	58.2	57.6	19.3	57.1	51.6	53.3	51.6	24.2	40.8	-	-	-	-	
UNO*	68.6	98.3	53.5	69.5	80.6	47.2	70.3	95.0	57.9	35.1	49.0	35.5	70.6	18.6	-	-	
GCD	91.5	97.9	88.5	73.0	66.5	74.1	89.8	66.8	51.3	86.6	48.7	39.0	57.6	29.9	80.2	85.1	77.6
DCCL	96.3	98.5	96.9	76.8	76.2	MLS	98.5	76.2	63.2	86.8	64.9	43.1	58.7	36.2	88.1	88.2	

- Exploring the estimated number of the conceptions during training.

- Along with the training, the conceptional generation follow a coarse-to-fine tendency.

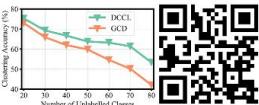


- The number of generated conceptions on generic datasets is larger than ground-truth, while its smaller on fine-grained datasets.



- Exploring the different splits of seen classes and unseen classes.

- Our DCCL can mine efficient cluster-level supervision, even in scenarios where only a limited number of seen classes are accessible.



Please see more details in our paper or by scanning the project QR code.

Why a Poster Matters

- **Purpose of a Poster**
 - Showcase research highlights in *minutes*
 - Attract attendees and spark conversation
 - Build valuable academic connections
- **Poster = Elevator Pitch + Visual Story**
- **Key Thought:** *A great poster gets you noticed before you speak.*

Poster Goals

- Communicate **clearly and visually**
 - Make your **main contribution obvious in 5 seconds**
 - Keep **details available for deeper discussion**
 - **Guide the viewer's eye** from top-left → bottom-right
 - Leave a **memorable impression** after they walk away
-

Poster Structure (Common in CVPR / NeurIPS)

- 1. Title & Authors** – Large, readable from 3 meters
 - 2. Motivation** – Why this research matters
 - 3. Problem Statement** – What gap you're addressing
 - 4. Method** – Architecture diagram, pipeline, or key equation
 - 5. Results** – Tables, metrics, and qualitative examples
 - 6. Conclusion & Future Work** – Short and punchy
 - 7. References & Acknowledgments** – Keep compact
-  *Tip:* Your “teaser figure” from the paper works perfectly as the centerpiece.

Visual Design Principles

- **Fonts:** Use clean sans-serif (Helvetica, Roboto, Calibri)
- **Font Sizes:**
 - Title: 80–120 pt
 - Section Headings: 40–60 pt
 - Body Text: 24–32 pt
- **Colors:** Limit to 2–3 main colors from your university or lab palette
- **Whitespace:** Use margins generously for clarity
- **Icons & Graphics:** Replace text with visuals whenever possible

Content Design Tips

- Use **one big, high-impact figure** to draw attention
 - Turn methods into **step-by-step visual flows**
 - For results: **highlight key numbers in bold or color**
 - Keep text **bullet-based**, avoid long paragraphs
 - Include **QR code** to your paper, code repo, or demo video
-

Presentation Strategy at the Conference

- **Opening Hook:** One-sentence summary of your research
 - **Walkthrough Flow:** Problem → Method → Key Result → Impact
 - **Engagement:** Ask attendees about their research, relate your work
 - **Leave-Behind Material:** Handouts, QR codes, or business cards
-

Common Mistakes to Avoid

- Overloading with text or too many figures
- Using small fonts that can't be read from a distance
- Poor color contrast (light text on light background)
- Crowding all results without storytelling
- Ignoring whitespace

Checklist Before Printing

- Main contribution visible from a few meters away
- All fonts readable from ~1.5 meters
- Colors consistent and print-safe (CMYK)
- No typos or inconsistent terminology
- QR code tested and functional

FINAL THOUGHTS



Final Thought

"Your poster is not just a display — it's a conversation starter, a networking tool, and a visual summary of months or years of research. Make it count."



Sample AI Posters

at a top-tier international conference
(e.g., NeurIPS, ICML, CVPR, AAAI).



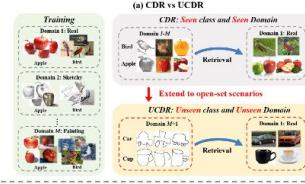
ProS: Prompting-to-simulate Generalized knowledge for UniversalCross-Domain Retrieval

Kaipeng Fang¹, Jingkuan Song¹, Lianli Gao¹, Pengpeng Zeng¹, Zhi-Qi Cheng², Xiya Li³, Heng Tao Shen⁴

¹University of Electronic Science and Technology of China, ²Carnegie Mellon University, ³Kuaishou Technology, ⁴Tongji University



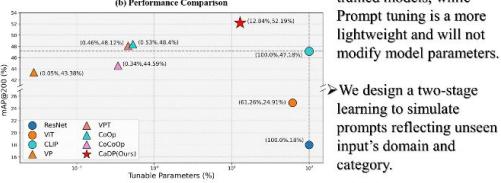
Background and Motivations



➤ Performing cross-domain retrieval well under generalized test scenarios, where query image may belong to strictly unknown domains and categories.

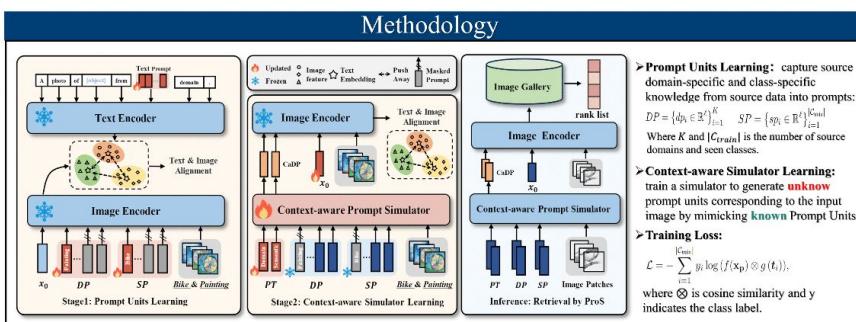
➤ Fine-Tuning will destroy general knowledge in pre-trained models, while Prompt tuning is a more lightweight and will not modify model parameters.

➤ We design a two-stage learning to simulate prompts reflecting unseen input's domain and category.



Contributions

- We are the first to investigate how to adapt CLIP with prompts for UCDR.
- We propose a prompt-based method named Prompting-to-Simulate (ProS), which learns the generalized knowledge to deal with open-set scenarios.
- Extensive experiments on three benchmark datasets show that our ProS achieves new state-of-the-art results compared with prompt-based methods without bringing excessive parameters.



Methodology

➤ **Prompt Units Learning:** capture source domain-specific and class-specific knowledge from source data into prompts:
 $D = \{d_i \in \mathbb{R}^3\}_{i=1}^K \quad SP = \{sp_i \in \mathbb{R}^3\}_{i=1}^{|\mathcal{C}_{train}|}$
 Where K and $|\mathcal{C}_{train}|$ is the number of source domains and seen classes.

➤ **Context-aware Simulator Learning:** train a simulator to generate **unknown** prompt units corresponding to the input image by mimicking **known** Prompt Units.

➤ Training Loss:

$$\mathcal{L} = - \sum_{i=1}^{|\mathcal{C}_t|} y_i \log(f(x_p) \otimes g(t_i)),$$

where \otimes is cosine similarity and y indicates the class label.

Experiments

Query Domain	Method	UCDR			U ⁴ CDR			TU-Berlin		
		Unseen Gallery	Mixed Gallery	Pre@200	Unseen Gallery	Mixed Gallery	Pre@200	Method	Sketchy	TU-Berlin
SaMoNet [32]	0.987	0.432	0.263	0.214	0.3526	0.1857	0.0576	ProS	0.5798	0.5441
RASA [11]	0.925	0.442	0.212	0.021	0.3733	0.2590	0.0576	w/o SP	0.5529	0.5142
Sketch	0.582	0.432	0.212	0.021	0.3733	0.2590	0.0576	w/o DP	0.4984	0.4641
CoOp [10]	0.5812	0.467	0.4985	0.4479	0.6734	0.6245	0.5956	w/o SP	0.5566	0.5109
ProS (Ours)	0.6457	0.601	0.5845	0.5663	0.7038	0.6911	0.6956	w/o Mask	0.5692	0.5324
SaMoNet [32]	0.176	0.181	0.172	0.111	0.1077	0.0589	0.0576	w/o A4	0.5573	0.5053
CLIP-Full	0.201	0.152	0.182	0.116	0.1826	0.0723	0.0576	w/o CLS	0.5241	0.4844
QuickDPR	0.201	0.152	0.182	0.116	0.1826	0.0723	0.0576	ProS (Ours)	0.6991	0.6445
ProS (Ours)	0.6457	0.601	0.5845	0.5663	0.7038	0.6911	0.6956	ProS (Ours)	0.6991	0.6445

Table 3. Different components evaluated on UCDR.

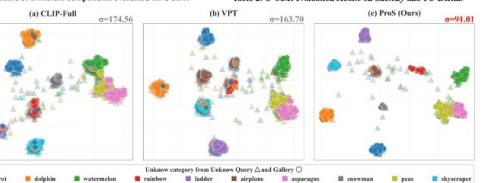


Table 2. UCDR evaluation results on Sketchy and TU-Berlin.

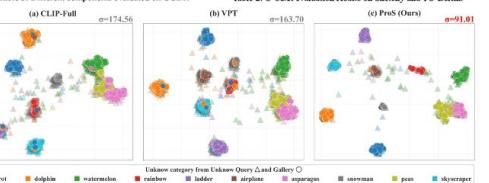


Figure 4. Visualization of image features from 10 randomly selected unseen classes of Real Query and unseen Infograph Gallery set.



Pre-trained Vision and Language Transformers Are Few-Shot Incremental Learners

Keon-Hee Park¹, Kyungwoo Song^{2†}, Gyeong-Moon Park¹¹Kyung Hee University, Republic of Korea ²Yonsei University, Republic of Korea

SAMPLE

Introduction

Few-Shot Class Incremental Learning (FSCIL)



- FSCIL = Few-Shot Learning + Class Incremental Learning.**
- Base Session: Many classes with sufficient training data.
- Incremental Session: Novel classes with few-shot training data.

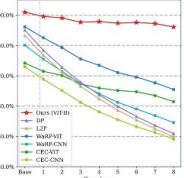
Main Challenges in FSCIL

Catastrophic Forgetting occurs while the network learns new classes sequentially, i.e., the network severely forgets the learned knowledge.

Overfitting arises when the network overly focuses on a limited set of training data, resulting in a degradation of overall performance.

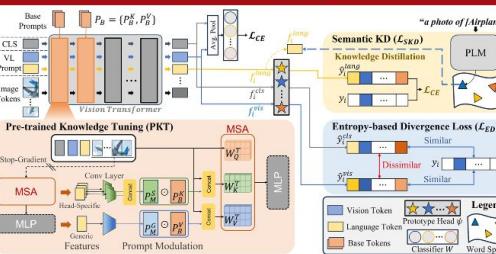
Large Pre-trained Models can be FSCIL Learners

To address challenges in FSCIL, previous studies have used shallow models like ResNet-18. However, their limited capacity hinders effective domain knowledge transfer. Recently, large pre-trained networks like ViT and CLIP have excelled in adaptability and performance but face a trade-off between retaining pre-trained knowledge and capturing novel knowledge, which hinders their use in FSCIL scenario.



We propose a new FSCIL framework based on Pre-trained Vision and Language transformers with prompting functions and knowledge distillation, called **PriViLeg**.

Method: PriViLeg



Pre-trained Knowledge Tuning (PKT)

- PKT adapts pre-trained knowledge for FSCIL by selectively fine-tuning layers with additional prompts for domain-specific knowledge.

Generate M-Prompt

The Process of PKT

$$\begin{aligned} h^{MSA} &= MSA(h_q, h_K, h_V) & h^{MLP} &= MLP(h^{MSA}) \\ P_M^S &= [g_1^S(h^{MSA}); \dots; g_N^S(h^{MSA})] & P_M^K &= P_M^S \odot P_B^K \\ P_M^L &= g^L(h^{MLP}) & P_V^L &= P_M^K \odot P_B^V \\ h^{out} &= MSA([h_q], [P_K^L]; [P_V^L]; h_V) \end{aligned}$$

Entropy-based Divergence Loss (\mathcal{L}_{ED})

- \mathcal{L}_{ED} aims to reinforce the discriminative power of the vision token itself.

$$\mathcal{L}_{ED} = \log \left(\frac{\mathcal{L}_{CE}(\hat{y}_i^{pe}, y_i) + \mathcal{L}_{CE}(\hat{y}_i^{cl}, y_i)}{\mathcal{L}_{KL}(\delta(\hat{y}_i^{pe}), \delta(\hat{y}_i^{cl}))} + 1 \right)$$

Semantic Knowledge Distillation Loss (\mathcal{L}_{SKD})

- \mathcal{L}_{SKD} is designed to provide additional semantic knowledge by distilling language embedding features to complement limited data access.

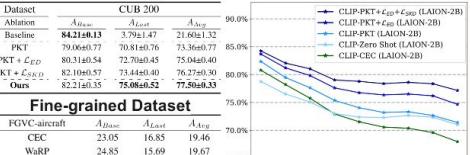
$$\begin{aligned} word_{cn_i} &= "a photo of [cn_i]" & \mathcal{L}_{SKD} &= \mathcal{L}_{KD}(f_i^{lang}, w_{cn_i}) + \gamma \cdot \mathcal{L}_{CE}(f_i^{lang}, y_i) \\ w_{cn_i} &= f_\phi(word_{cn_i}) & & \end{aligned}$$

Experiments

Main Table

Dataset	CUB200			CIFAR-100		
	A_{Base}	A_{Last}	A_{Avg}	A_{Base}	A_{Last}	A_{Avg}
Fine-Tuning + Proto ψ	84.21±0.13	3.79±1.47	21.60±1.32	91.36±0.15	5.19±0.13	37.04±1.06
CEC [CVPR'21]	79.06±0.77	70.14±0.76	73.36±0.77	80.31±0.54	72.07±0.45	75.04±0.40
L2P [CVPR'22]	44.97±2.32	15.41±3.45	24.99±4.30	83.29±0.50	49.87±0.31	64.08±0.39
DualPrompt [ECCV'22]	53.37±1.83	23.25±2.02	26.30±2.39	85.11±0.29	50.93±0.21	65.45±0.27
NC-FSCIL [ICLR'23]	78.49±2.32	38.80±1.14	57.92±1.71	89.51±0.23	53.70±0.14	68.96±0.17
WaRP [ICLR'23]	67.74±5.57	49.36±6.56	55.85±6.06	86.20±1.46	65.48±1.87	74.55±1.67
PriViLeg (Ours)	82.21±0.35	75.08±0.52	77.50±0.33	90.88±0.20	86.06±0.32	88.08±0.20

Ablation Study



Fine-grained Dataset

FGVC-aircraft, CEC, WaRP

PriViLeg (Ours)

82.21±0.35

45.55

50.87

65.00

70.00

75.00

80.00

85.00

90.00

95.00

100.00

105.00

110.00

115.00

120.00

125.00

130.00

135.00

140.00

145.00

150.00

155.00

160.00

165.00

170.00

175.00

180.00

185.00

190.00

195.00

200.00

205.00

210.00

215.00

220.00

225.00

230.00

235.00

240.00

245.00

250.00

255.00

260.00

265.00

270.00

275.00

280.00

285.00

290.00

295.00

300.00

305.00

310.00

315.00

320.00

325.00

330.00

335.00

340.00

345.00

350.00

355.00

360.00

365.00

370.00

375.00

380.00

385.00

390.00

395.00

400.00

405.00

410.00

415.00

420.00

425.00

430.00

435.00

440.00

445.00

450.00

455.00

460.00

465.00

470.00

475.00

480.00

485.00

490.00

495.00

500.00

505.00

510.00

515.00

520.00

525.00

530.00

535.00

540.00

545.00

550.00

555.00

560.00

565.00

570.00

575.00

580.00

585.00

590.00

595.00

600.00

605.00

610.00

615.00

620.00

625.00

630.00

635.00

640.00

645.00

650.00

655.00

660.00

665.00

670.00

675.00

680.00

685.00

690.00

695.00

700.00

705.00

710.00

715.00

720.00

725.00

730.00

735.00

740.00

745.00

750.00

755.00

760.00

765.00

770.00

775.00

780.00

785.00

790.00

795.00

800.00

805.00

810.00

815.00

820.00

825.00

830.00

835.00

840.00

845.00

850.00

855.00

860.00

865.00

870.00

875.00

880.00

885.00

890.00

895.00

900.00

905.00

910.00

915.00

920.00

925.00

930.00

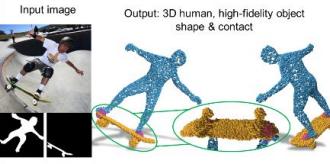
935.00



Template Free Reconstruction of Human-object Interaction with Procedural Interaction Generation

Xianghui Xie Bharat Lal Bhatnagar Jan Eric Lenssen Gerard Pons-Moll
University of Tübingen Tübingen AI Center Max Planck Institute for Informatics

Introduction



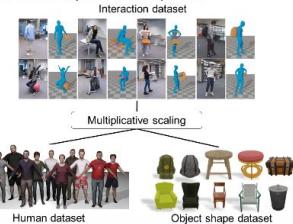
Challenges

- Depth scale ambiguity & occlusion.
- Diverse human & object shapes.
- Template free: lack of interaction data.
- Real data capture is not scalable.

Key Idea

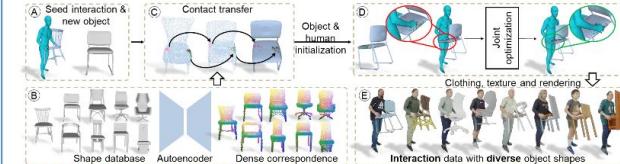
Decompose combinatorial space:

- Generate data: multiply datasets.
- Reconstruction: different sub-models to learn separate subspaces.



ProciGen: Procedural interaction Generation

- Humans interact similarly with objects of the same category.



ProciGen Dataset

- 1M+ interaction images.
- 21k+ different object shapes.
- GT meshes of object + SMPL-D.

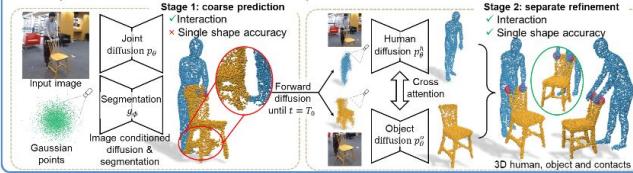


Download
ProciGen
Here!



HDM: Hierarchical Diffusion Model

- Joint model learns interaction space.
- Separate models learn individual shapes.

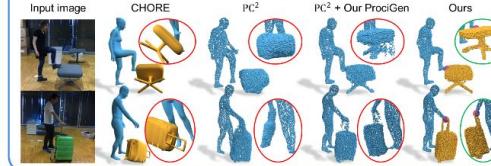


Evaluation

	Method	Human ^f	Object ^f	Comb. ^f
BEHAVE	CHORE [†]	0.3454	0.4258	0.3966
	PC2 [†]	X	X	0.4231
	Ours[‡]	0.3925	0.5049	0.4604
	Ours synz. only [‡]	0.3477	0.4351	0.4110
InterCap	CHORE [†]	0.4064	0.5135	0.4687
	PC2 [†]	X	X	0.5057
	Ours[‡]	0.4399	0.6072	0.5344
	Ours synz. only [‡]	0.3851	0.4928	0.4530

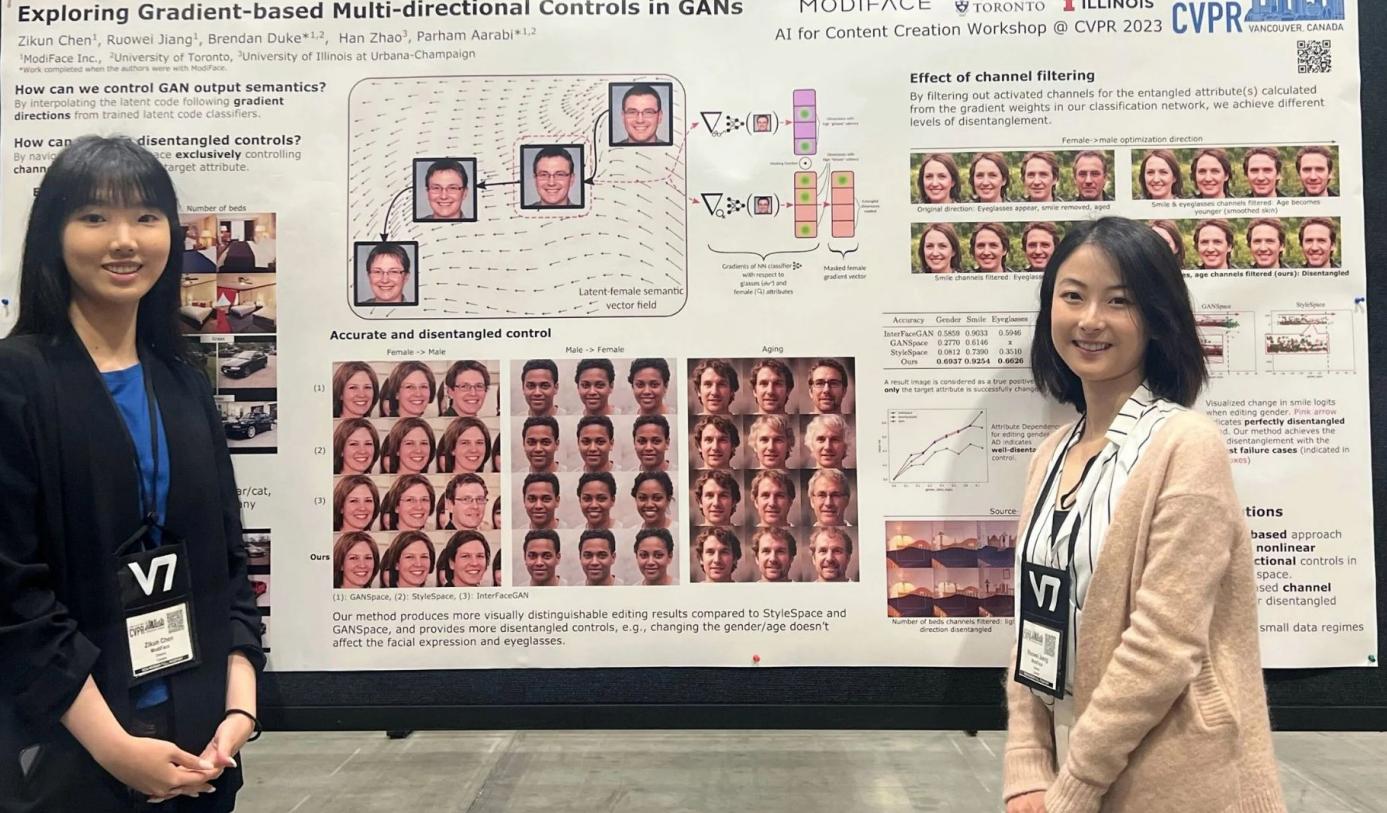
Metric: F-score@0.01m.

†: Template-based. ‡: Template-free.



In the Wild Generalization







Fine-tuned thin-plate spline motion model for manipulating social information in paper-wasp colonies

Kacy Hatfield^a, Akudasuo Ezenyilima, Nitin Verma, Juan José García Mesa, So Eun Moon, Elizabeth Tibbets, Pavan Turaga, Theodore P. Pavlic



^aGeometric Media Lab + Bringing Ecology and Engineering Together Lab (ASU)

*Correspondence: khatfield2@asu.edu

Background: *Polistes* Paper Wasps

- Social Complexity: Species of *Polistes* paper wasp form complex and fluid hierarchical social structures. They are tractable model systems for studying visual individual recognition and the maintenance of social structures in animals [2].



(a) *P. dominula*



(b) *P. fuscatus*

- Facial Recognition: These wasps identify other individual colony members using visual facial features. They leave based on pairwise interactions of others and modulate their own social behaviors based on this recognition [3].

- Nests as Social Laboratories: Researchers can modify wasp faces to study how social hierarchies formed and are reformed. Usually pain to wasps.

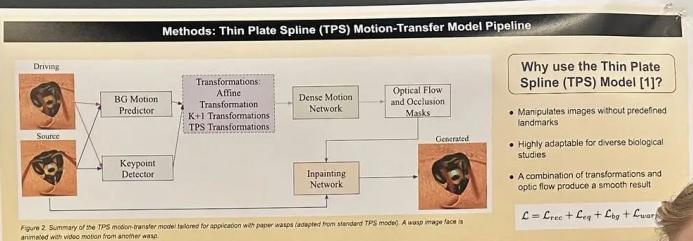
- Other social insects have shown to respond to video:



This promising lead on social-insect recognition through video allows for the digitization and digital manipulation of *Polistes* paper wasp interactions.

Motivation: Generative AI for Studying Wasps

- Richer Set of Behavioral Manipulations: Generative AI technologies, like the TPS motion-transfer model [1], allow for reproducing a wide range of archived complex behaviors with arbitrary wasp faces. This greatly expands laboratory capabilities beyond state-of-the-art methods based on painting wasp faces.
- Broader Impacts: The distinct facial morphology of wasps and other social insects (e.g., concave mandibles and mobile antennae that temporally obscure faces) present new challenges for motion transfer that, if solved, allow for these methods to be used in a wide range of other applications.
- Ecological Advancement: Individual recognition (IR) in animals has historically been tested by manual comparison of appearance or vocalizations [2]. Through effective motion transfer on social species expands the potential for further research into the neurological, behavioral, and social implications of IR.



Preliminary Results: 3 Cases

- A: Pre-Trained Model + Human Motion**
 - Captured broad motion dynamics effectively but struggled with accurate facial feature translation due to anatomical differences.
- B: Pre-Trained Model + Wasp Motion**
 - Displayed limitations in detecting subtle wasp-specific movements, resulting in minimal effective motion transfer.
- C: Custom Model + Wasp Motion**
 - Somewhat detected critical movements like mandible motion, showing promise in more specific facial expression animations.

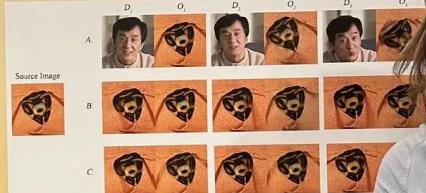


Figure 2: Case A (top) shows results using a model pre-trained with human data and applied to a *P. dominula* source and human target. Case B (middle) uses the same pre-trained model but with *P. dominula* source and driving video. Case C (bottom) uses a custom model trained on wasp faces with *P. dominula* source and driving video. Sample videos are available at https://github.com/PavlicLab/CVPR2024_CV4Competition/tree/main/TPSMotionTransfer_Wasps (GitHub code below).

Conclusion & Future Work

- Defined a pipeline for motion-transfer model application in entomological research: Although currently training on a landmark-less model, investigation into landmarked models continues to boast potential benefits to define precise areas of interest in motion transfer.
- Defined potential for applying motion-transfer models to behavioral research of *Polistes* paper wasps: Given prior research, we intend on testing outcomes of this motion-transfer technology on *Polistes* paper wasps to gain deeper insight into colony hierarchy and individual recognition in non-human animals.
- Illuminated challenges unique to nonhuman data: Further experimentation remains paramount to understand the parameters requiring alteration from human to insect data (e.g. background/antenna removal).

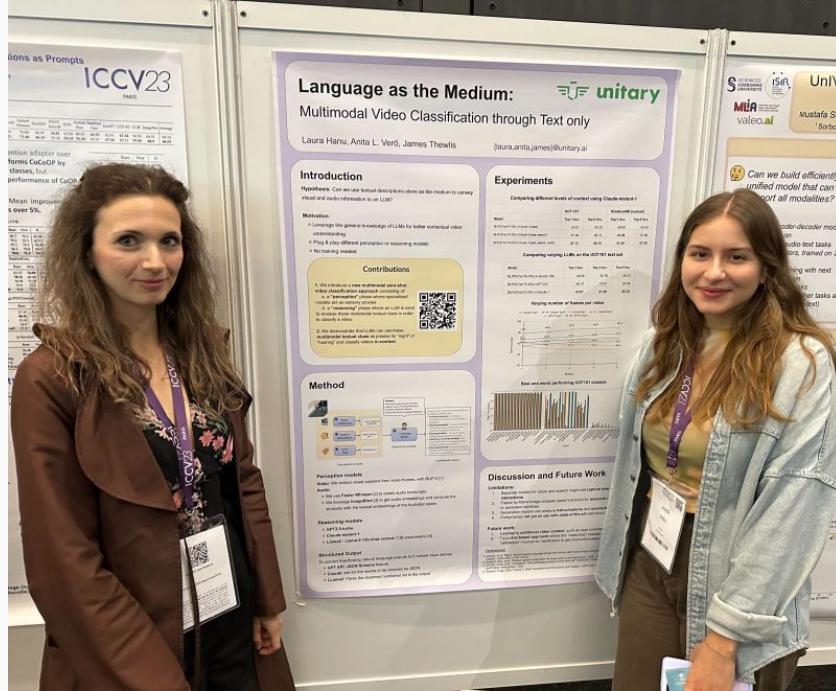
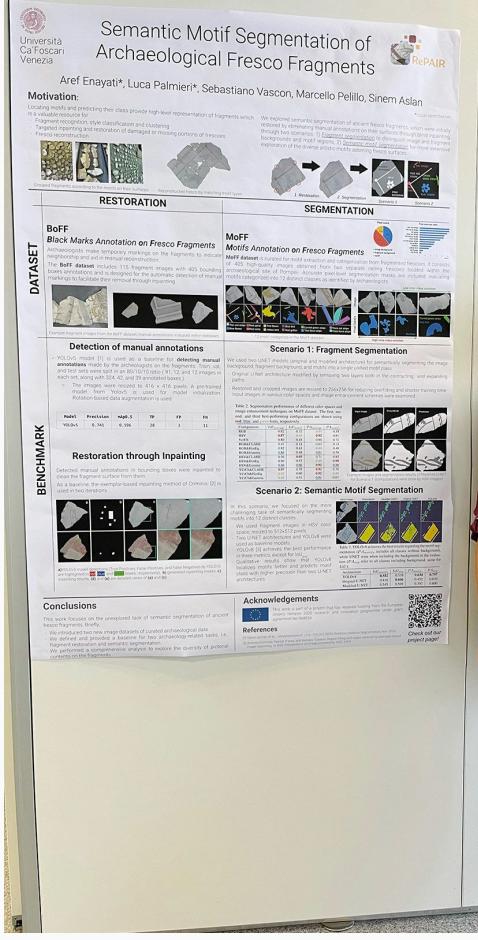
Code



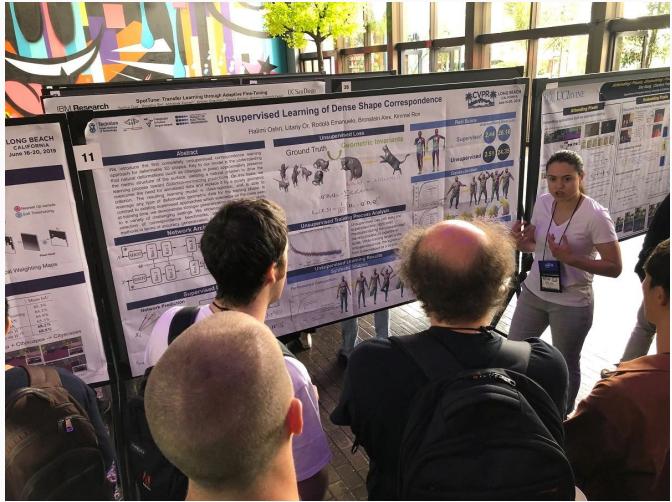
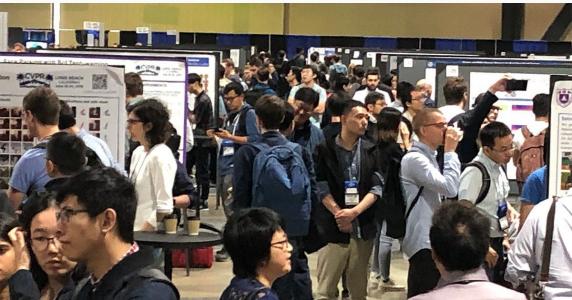
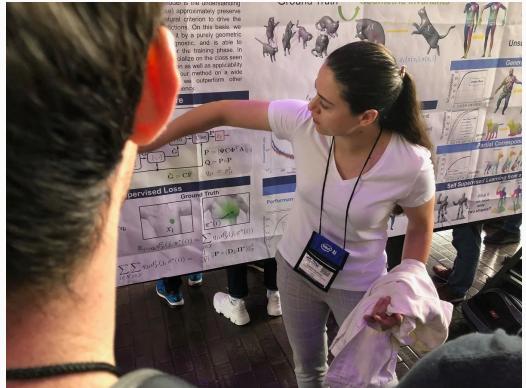
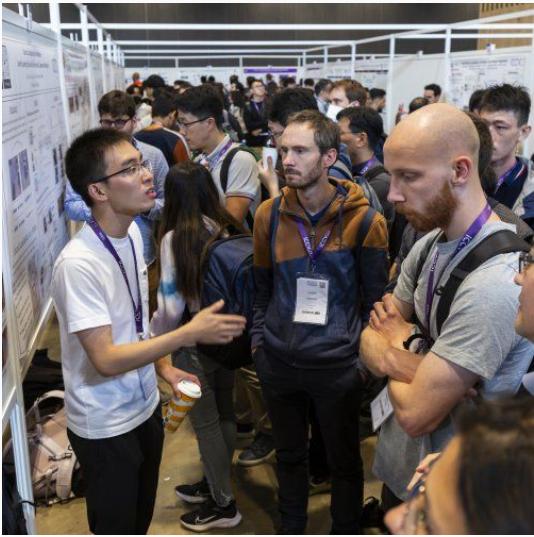
References

- [1] J Zhao and H Zhang. Thin-plate spline motion model for image animation. In: Proceedings of CVPR 2022, pages 3657–3666.
- [2] EA Tibbets and J Dale. Individual recognition: It's good to be different. Trends in Ecology and Evolution, 22(10):529–537, 2007.
- [3] EA Tibbets, J Pardo-Sánchez, J Ramírez-Matías, and A Argáez-Weber. Individual recognition is associated with holistic face processing in *Polistes* paper wasps in a species-specific way. Proceedings of the Royal Society B: Biological Sciences, 288(1943): 20200310, 2020.





SAMPLE



Question?