



# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



Xuan Ju<sup>\*12</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



**50K** Images  
**123K** Instances  
**5** Natural Scenarios  
**15** Artificial Scenarios

Xuan Ju<sup>\*12</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

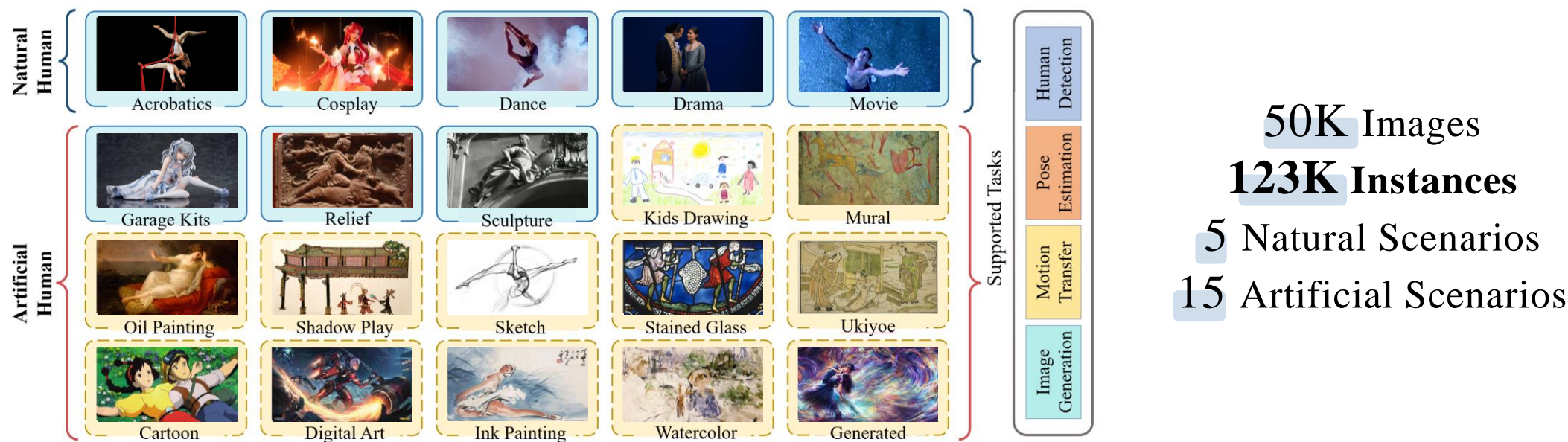
\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



Xuan Ju<sup>\*12</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



50K Images  
 123K Instances  
 5 Natural Scenarios  
 15 Artificial Scenarios

Xuan Ju<sup>\*12</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



**50K** Images  
**123K** Instances  
**5** Natural Scenarios  
**15** Artificial Scenarios

Xuan Ju<sup>\*12</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

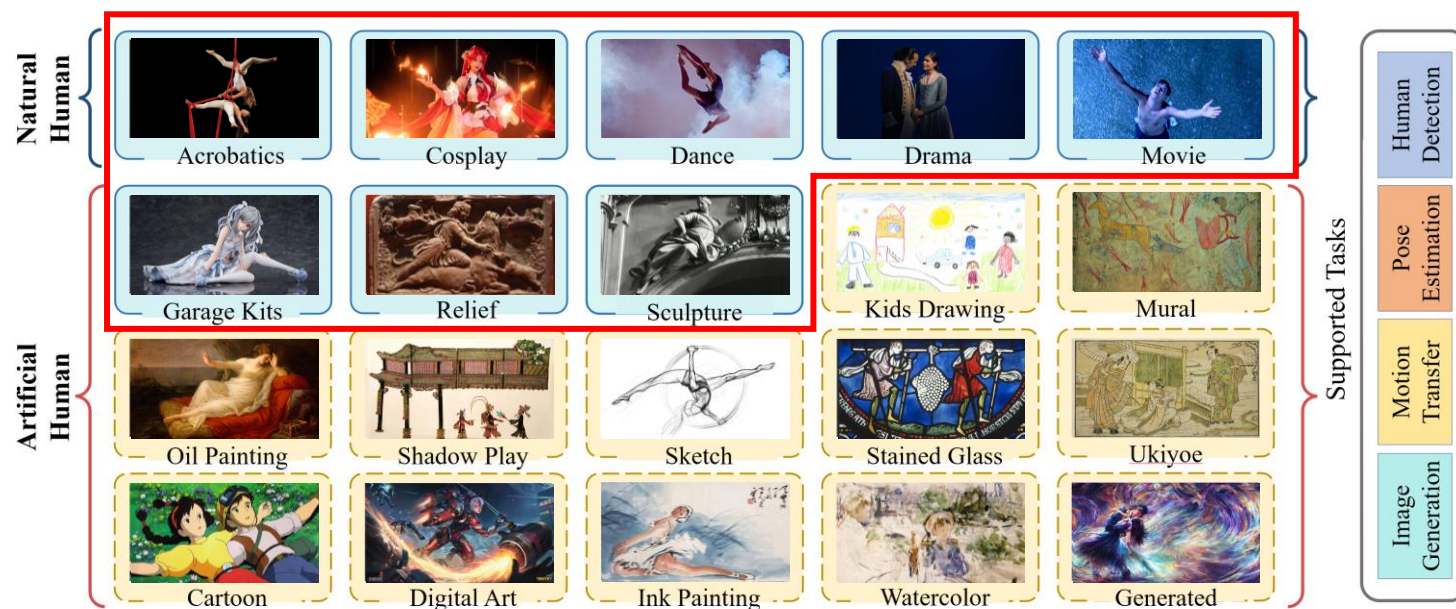
\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



50K Images  
 123K Instances  
 5 Natural Scenarios  
 15 Artificial Scenarios

Xuan Ju<sup>\*12</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

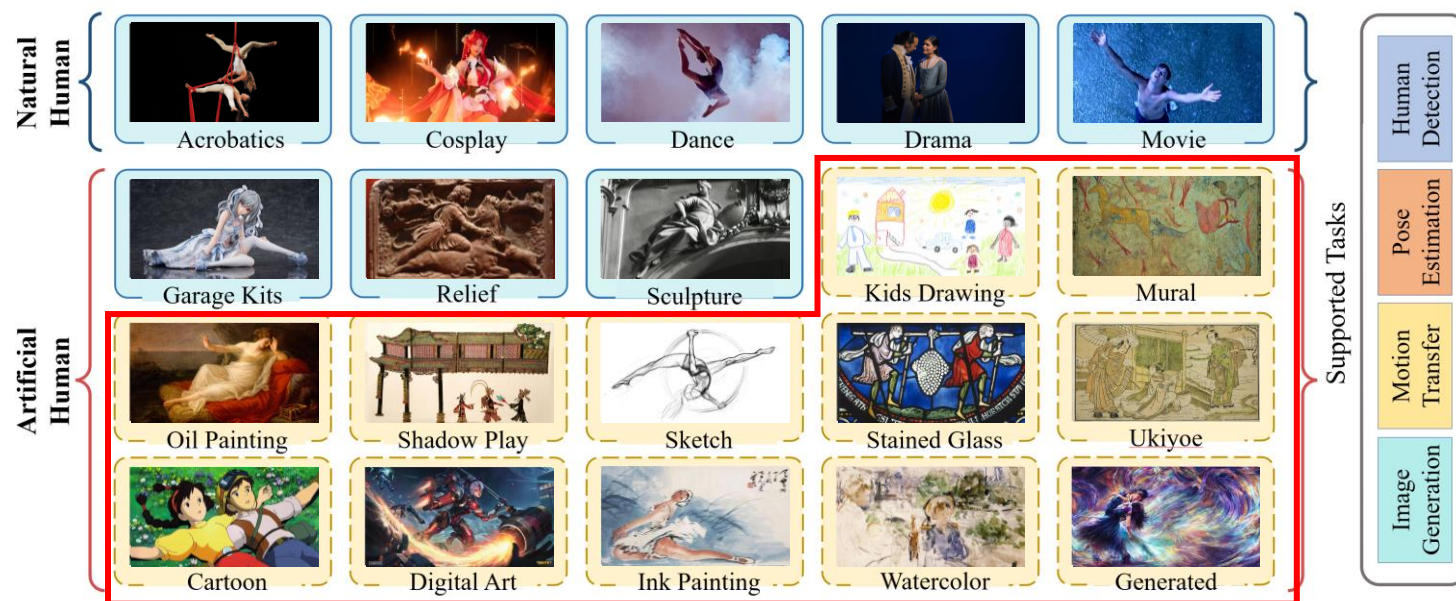
\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



50K Images  
 123K Instances  
 5 Natural Scenarios  
 15 Artificial Scenarios

Xuan Ju<sup>\*12</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



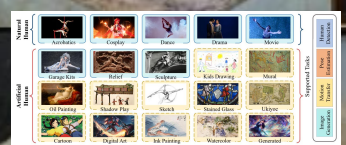
Xuan Ju<sup>\*1,2</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



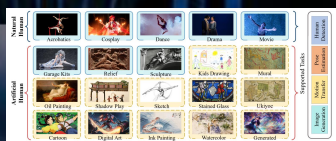
Xuan Ju<sup>\*1,2</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



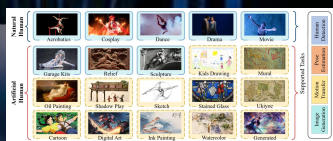
Xuan Ju<sup>\*1,2</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



**Bounding box**

**Self-contact**

**Picture Name:** Black Canary (Heroic Pose)  
**Author:** jettmanas  
**Description:** Black Canary with Blonde hair and blue eyes is wearing a blue leotard and boots, standing with arms crossed.  
**Category:** cartoon

0: Nose	11: L-hip
1: L-eye	12: R-hip
2: R-eye	13: L-knee
3: L-ear	14: R-knee
4: R-ear	15: L-ankle
5: L-shoulder	16: R-ankle
6: R-shoulder	17: L-finger
7: L-elbow	18: R-finger
8: R-elbow	19: L-toe
9: L-wrist	20: R-toe
10: R-wrist	



Xuan Ju<sup>\*12</sup>, Ailing Zeng<sup>\*1</sup>, Jianan Wang<sup>1</sup>, Qiang Xu<sup>2</sup>, Lei Zhang<sup>1</sup>

\* Equal contribution <sup>1</sup>International Digital Economy Academy <sup>2</sup>The Chinese University of Hong Kong



# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



**Bounding box**

**Self-contact**

**Picture Name:** Black Canary  
(Heroic Pose)  
**Author:** jettmanas  
**Description:** Black Canary with Blonde hair and blue eyes is wearing a blue leotard and boots, standing with arms crossed.  
**Category:** cartoon

0: Nose	11: L-hip
1: L-eye	12: R-hip
2: R-eye	13: L-knee
3: L-ear	14: R-knee
4: R-ear	15: L-ankle
5: L-shoulder	16: R-ankle
6: R-shoulder	17: L-finger
7: L-elbow	18: R-finger
8: R-elbow	19: L-toe
9: L-wrist	20: R-toe
10: R-wrist	



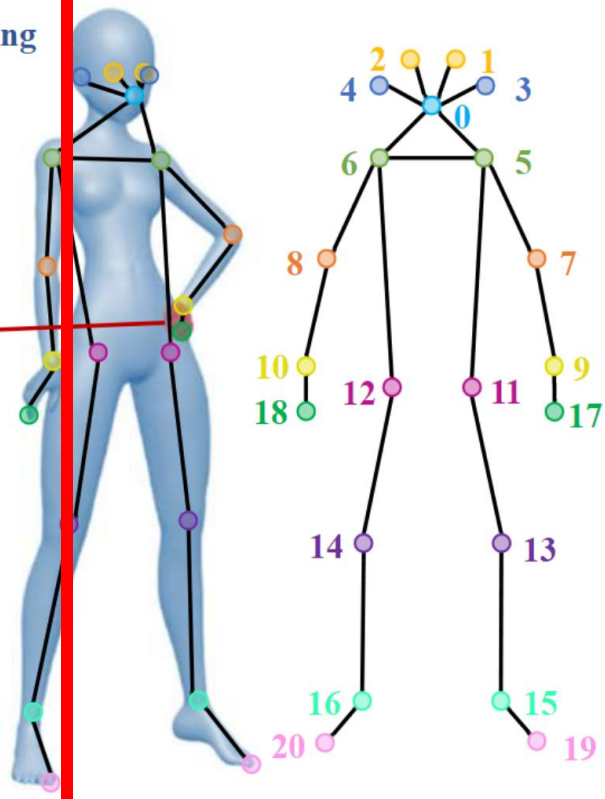


# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



Bounding box

Self-contact



**Picture Name:** Black Canary  
(Heroic Pose)  
**Author:** jettmanas  
**Description:** Black Canary with  
Blonde hair and blue eyes is  
wearing a blue leotard and boots,  
standing with arms crossed.  
**Category:** cartoon


- |               |              |
|---------------|--------------|
| 0: Nose       | 11: L-hip    |
| 1: L-eye      | 12: R-hip    |
| 2: R-eye      | 13: L-knee   |
| 3: L-ear      | 14: R-knee   |
| 4: R-ear      | 15: L-ankle  |
| 5: L-shoulder | 16: R-ankle  |
| 6: R-shoulder | 17: L-finger |
| 7: L-elbow    | 18: R-finger |
| 8: R-elbow    | 19: L-toe    |
| 9: L-wrist    | 20: R-toe    |
| 10: R-wrist   |              |



# Human-Art: A Versatile Human-Centric Dataset

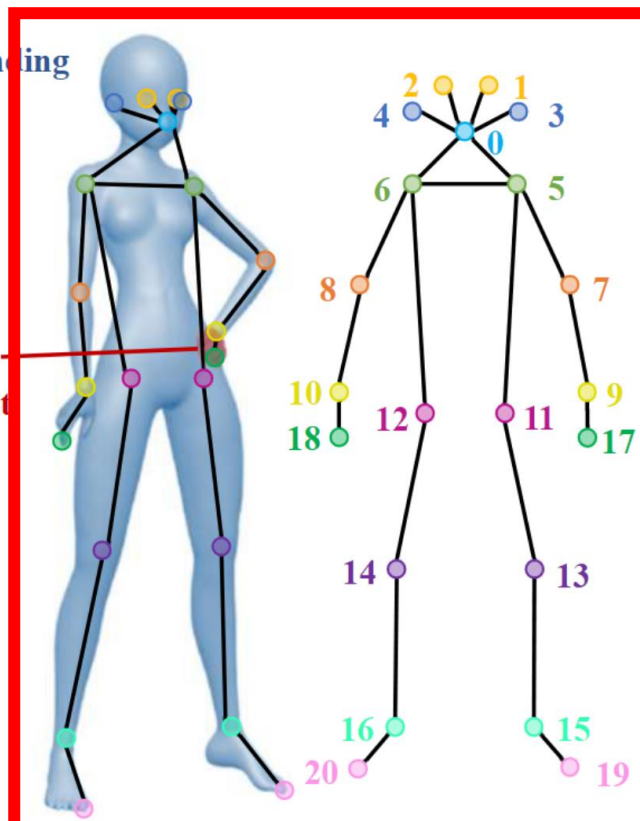
## Bridging Natural and Artificial Scenes





**Bounding box**

**Self-contact**



**Picture Name:** Black Canary  
(Heroic Pose)

**Author:** jettmanas

**Description:** Black Canary with Blonde hair and blue eyes is wearing a blue leotard and boots, standing with arms crossed.

**Category:** cartoon

---

<b>0: Nose</b>	<b>11: L-hip</b>
<b>1: L-eye</b>	<b>12: R-hip</b>
<b>2: R-eye</b>	<b>13: L-knee</b>
<b>3: L-ear</b>	<b>14: R-knee</b>
<b>4: R-ear</b>	<b>15: L-ankle</b>
<b>5: L-shoulder</b>	<b>16: R-ankle</b>
<b>6: R-shoulder</b>	<b>17: L-finger</b>
<b>7: L-elbow</b>	<b>18: R-finger</b>
<b>8: R-elbow</b>	<b>19: L-toe</b>
<b>9: L-wrist</b>	<b>20: R-toe</b>
<b>10: R-wrist</b>	





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes

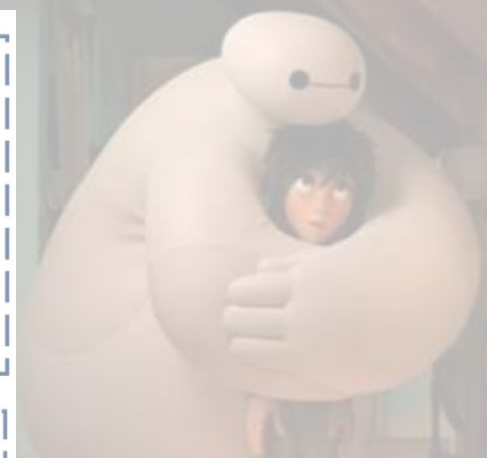


**Bounding box**

**Self-contact**

**Picture Name:** Black Canary  
(Heroic Pose)  
**Author:** jettmanas  
**Description:** Black Canary with Blonde hair and blue eyes is wearing a blue leotard and boots, standing with arms crossed.  
**Category:** cartoon

0: Nose	11: L-hip
1: L-eye	12: R-hip
2: R-eye	13: L-knee
3: L-ear	14: R-knee
4: R-ear	15: L-ankle
5: L-shoulder	16: R-ankle
6: R-shoulder	17: L-finger
7: L-elbow	18: R-finger
8: R-elbow	19: L-toe
9: L-wrist	20: R-toe
10: R-wrist	





# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes

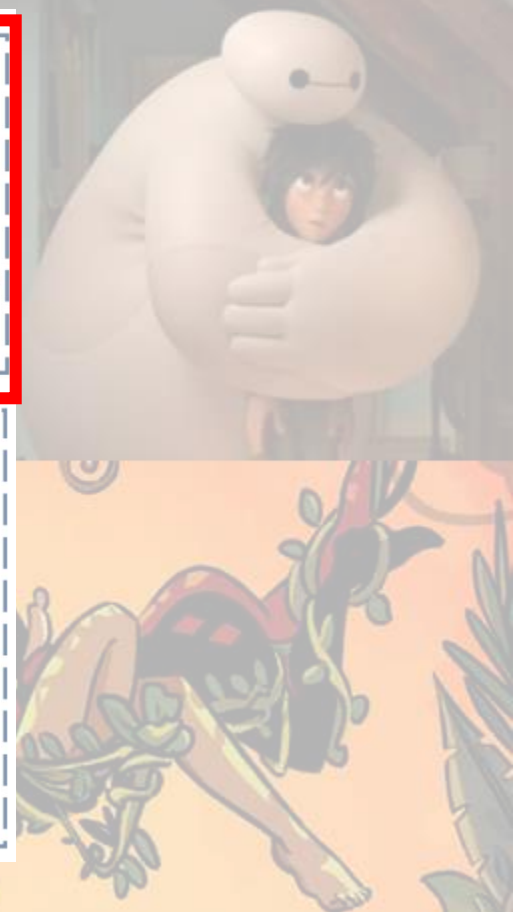


**Bounding box**

**Self-contact**

**Picture Name:** Black Canary  
(Heroic Pose)  
**Author:** jettmanas  
**Description:** Black Canary with Blonde hair and blue eyes is wearing a blue leotard and boots, standing with arms crossed.  
**Category:** cartoon

0: Nose	11: L-hip
1: L-eye	12: R-hip
2: R-eye	13: L-knee
3: L-ear	14: R-knee
4: R-ear	15: L-ankle
5: L-shoulder	16: R-ankle
6: R-shoulder	17: L-finger
7: L-elbow	18: R-finger
8: R-elbow	19: L-toe
9: L-wrist	20: R-toe
10: R-wrist	







# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



**Bounding box**

**Self-contact**

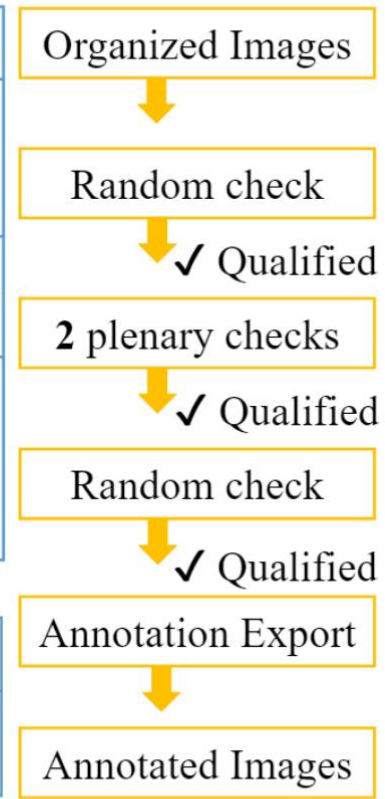
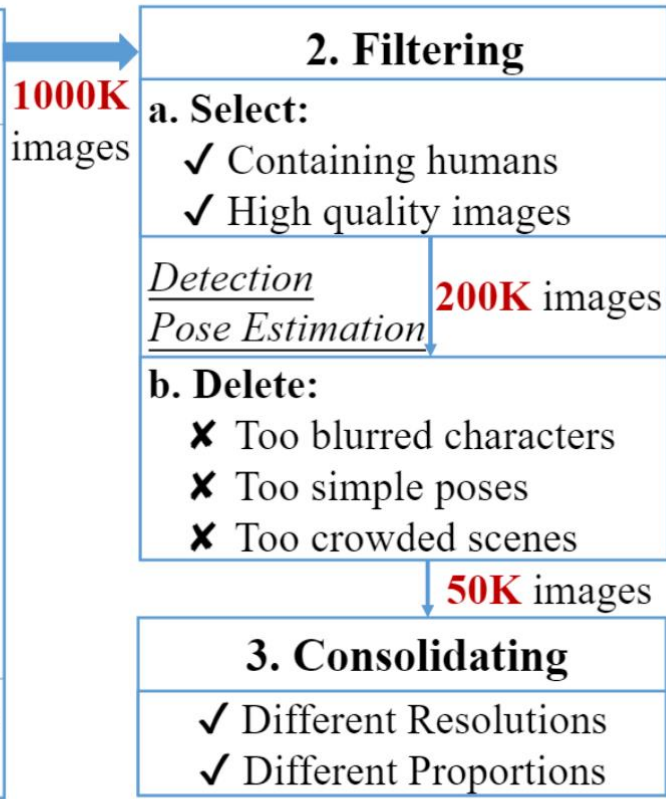
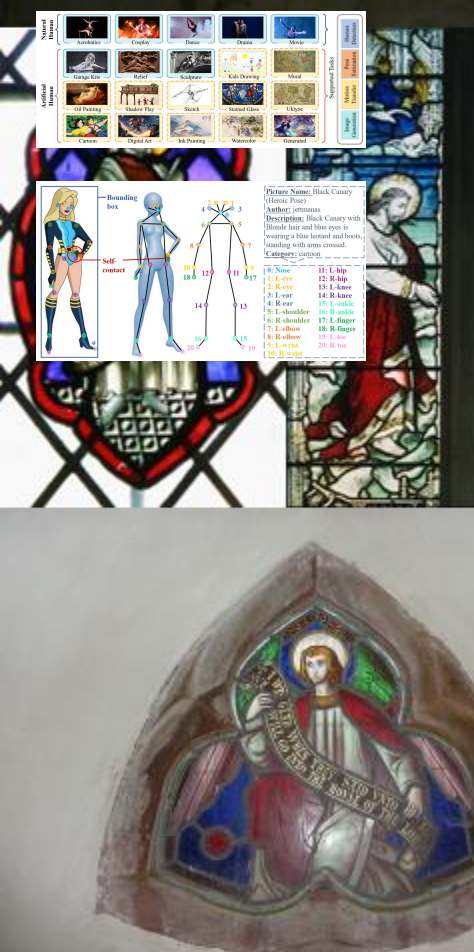
**Picture Name:** Black Canary  
(Heroic Pose)  
**Author:** jettmanas  
**Description:** Black Canary with Blonde hair and blue eyes is wearing a blue leotard and boots, standing with arms crossed.  
**Category:** cartoon

0: Nose	11: L-hip
1: L-eye	12: R-hip
2: R-eye	13: L-knee
3: L-ear	14: R-knee
4: R-ear	15: L-ankle
5: L-shoulder	16: R-ankle
6: R-shoulder	17: L-finger
7: L-elbow	18: R-finger
8: R-elbow	19: L-toe
9: L-wrist	20: R-toe
10: R-wrist	





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



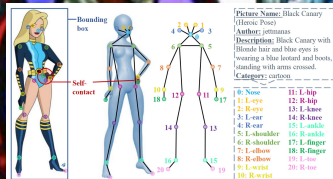
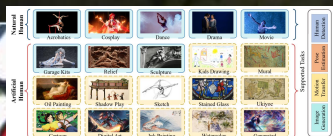
(a) Data collection processes

(b) Annotation





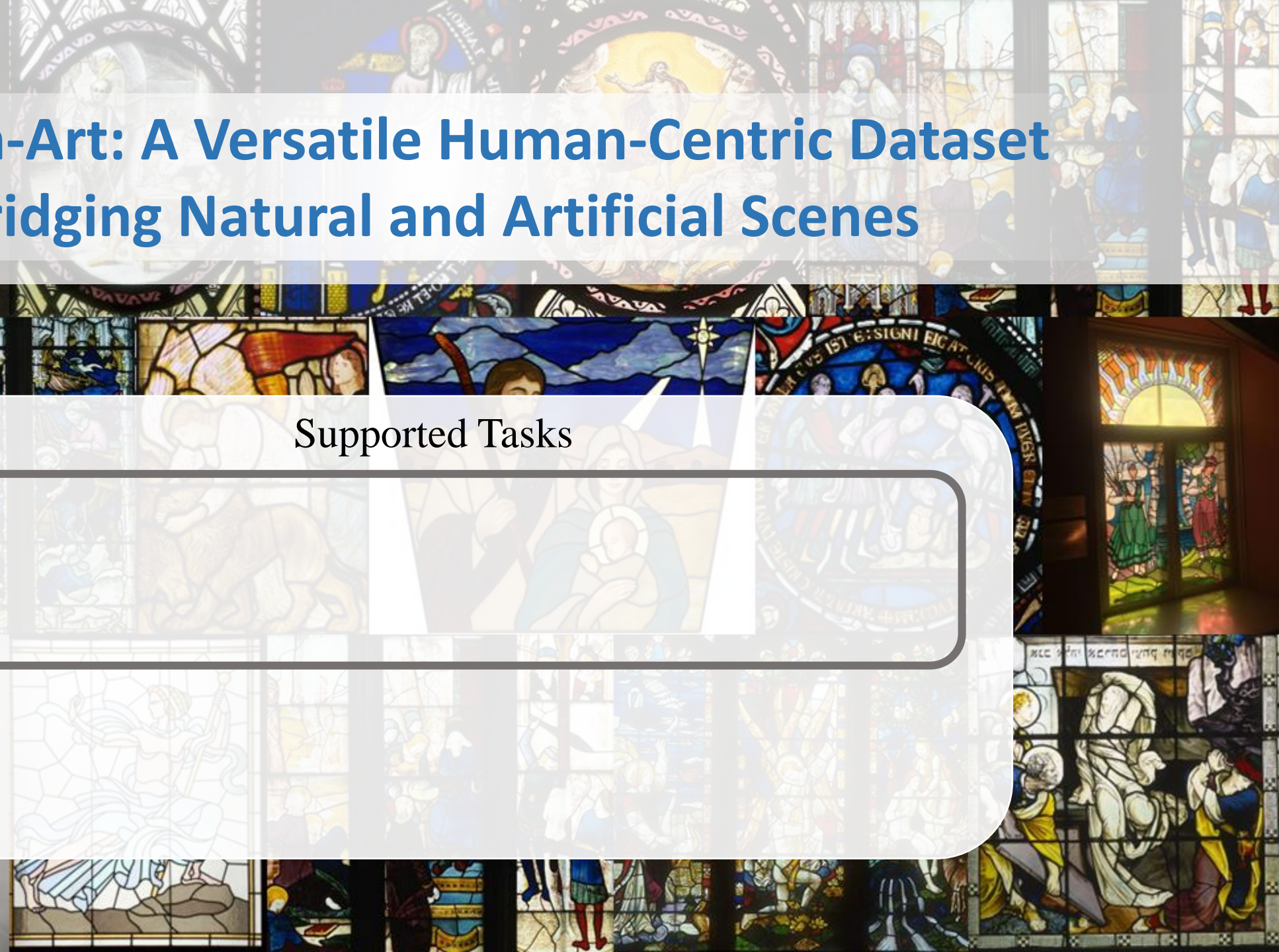
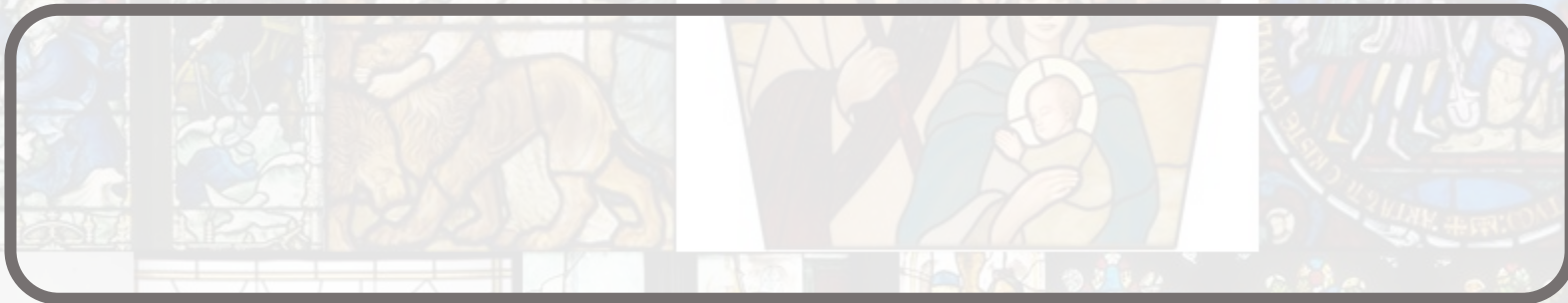
# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



1. Manual Collection & Classification	2. Filtering	Organized Images
1000k images Ukiyo-e.org TheMet Sonia Halliday 28 image collection websites and 3 search engines	a. Select: ✓ Containing humans ✓ High quality images b. Detect: ✗ Too blurred characters ✗ Too simple poses ✗ Too crowded scenes c. Consolidating: ✓ Different Resolutions ✓ Different Proportions	Random check ✓ Qualified 2 plenary checks Random check ✓ Qualified Annotation Export Annotated Images

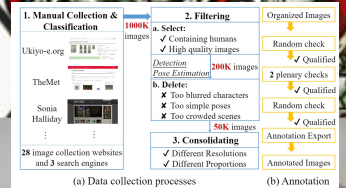
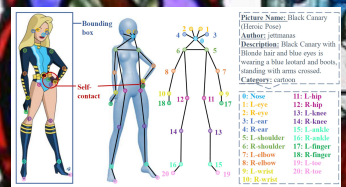
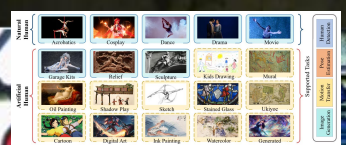
(a) Data collection processes (b) Annotation

## Supported Tasks





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



## Supported Tasks

Human  
Detection





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes

**1. Manual Collection & Classification**

- 1000k images
- Ukiyo-e.org
- TheMet
- Sonia Halliday
- 28 image collection websites and 3 search engines

**2. Filtering**

a. Select:

- ✓ Containing humans
- ✓ High quality images

b. Detect:

- ✓ 200k images
- ✓ Pose Estimation
- ✗ Too blurred characters
- ✗ Too simple poses
- ✗ Too crowded scenes
- 50k images

**3. Consolidating**

- ✓ Different Resolutions
- ✓ Different Proportions

**Organized Images**

- ✓ Random check
- ✓ Qualified
- 2 plenary checks
- ✓ Qualified
- Random check
- ✓ Qualified

**Annotation Export**

- ✓ Annotated Images

(a) Data collection processes (b) Annotation

## Supported Tasks

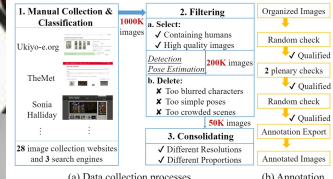
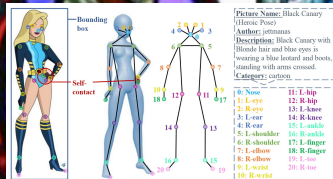
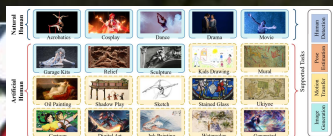
Human  
Detection

Pose  
Estimation





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



## Supported Tasks

Human  
Detection

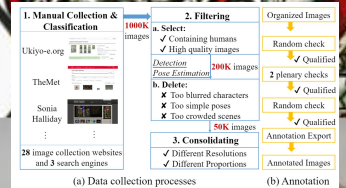
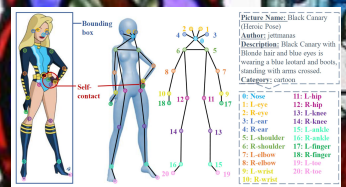
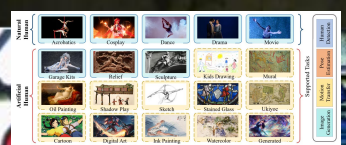
Pose  
Estimation

Motion  
Transfer





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



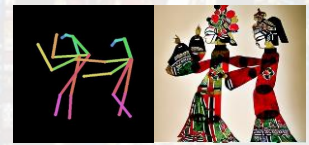
## Supported Tasks

Human  
Detection

Pose  
Estimation

Motion  
Transfer

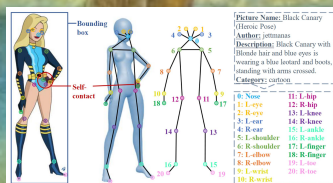
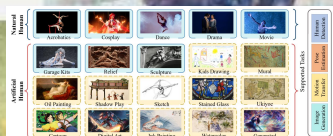
Image  
Generation



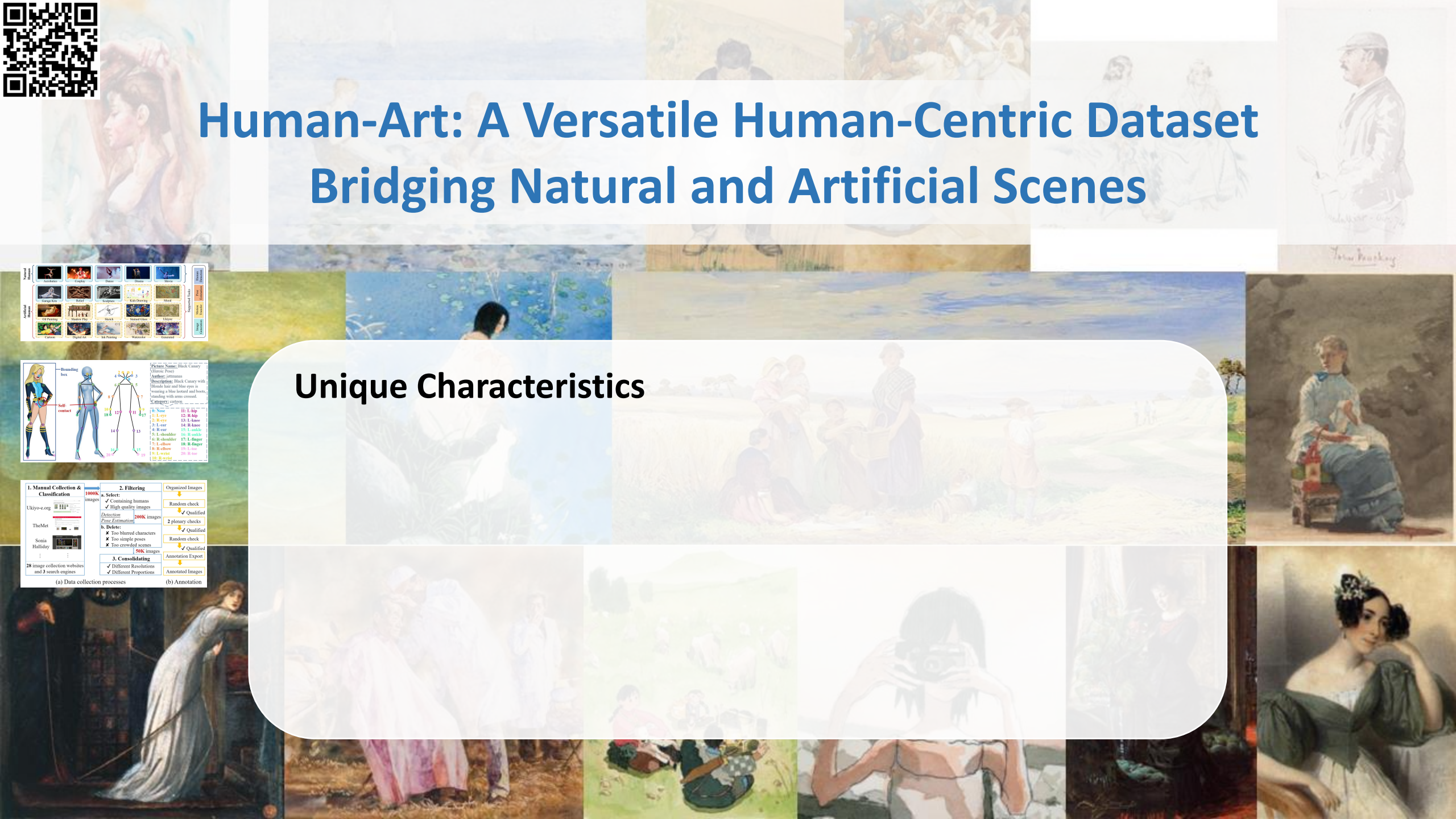
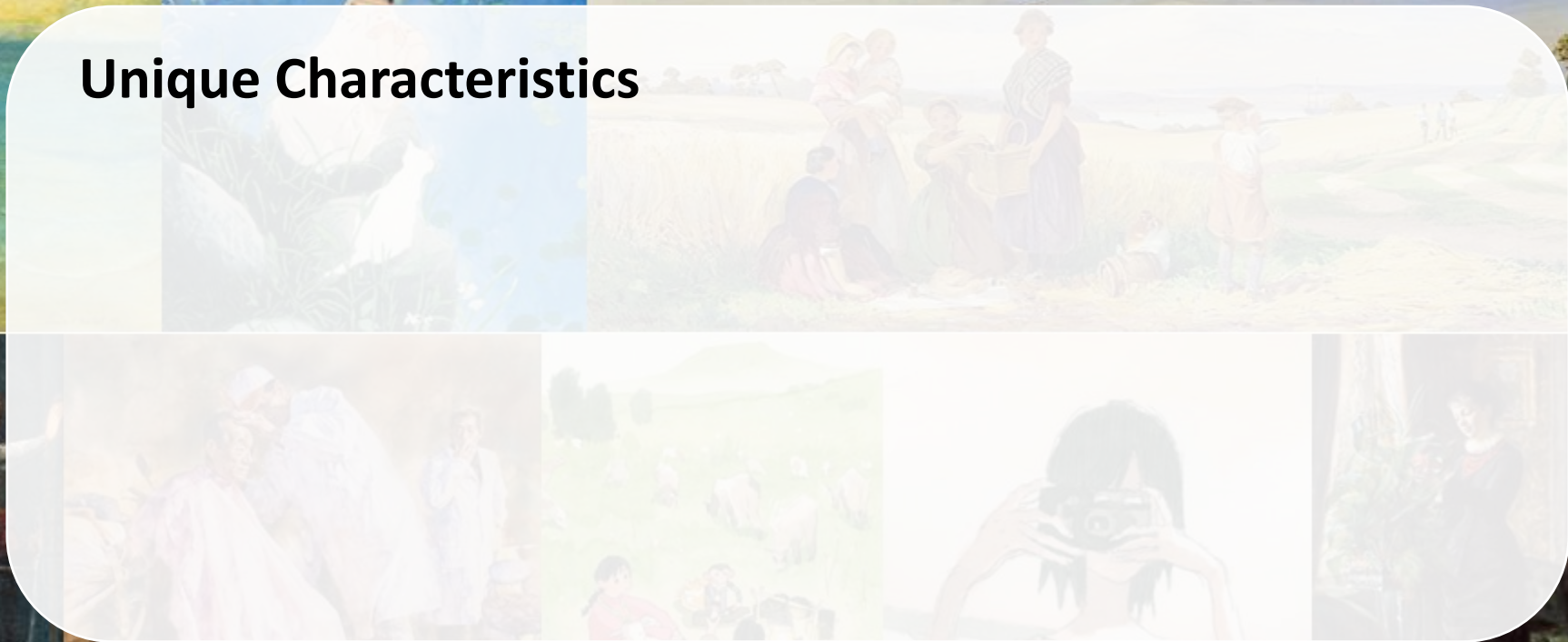
Shadow play, two people dressed orientally dancing



# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



1. Manual Collection & Classification	2. Filtering	Organized Images
<ul style="list-style-type: none"> <li>1000k images</li> <li>Ukiyo-e.org</li> <li>TheMet</li> <li>Sonia Halliday</li> <li>28 image collection websites and 3 search engines</li> </ul>	<p>a. Select:</p> <ul style="list-style-type: none"> <li>✓ Containing humans</li> <li>✓ High quality images</li> </ul> <p>b. Detect:</p> <ul style="list-style-type: none"> <li>✗ Too blurred characters</li> <li>✗ Too simple poses</li> <li>✗ Too crowded scenes</li> </ul>	<ul style="list-style-type: none"> <li>Random check</li> <li>2 plenary checks</li> <li>Random check</li> <li>Annotation Export</li> <li>Annotated Images</li> </ul>
(a) Data collection processes	(b) Annotation	





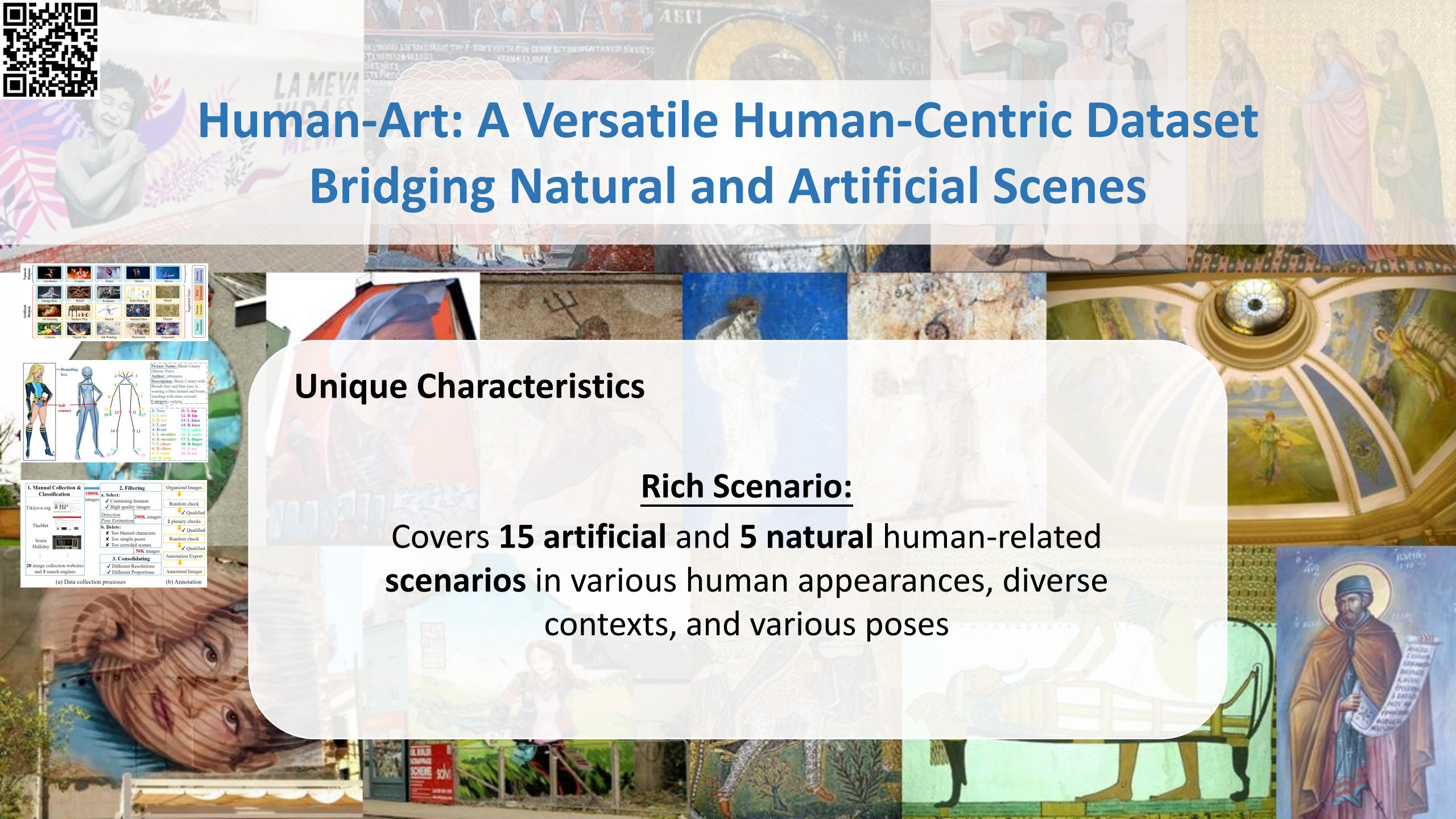
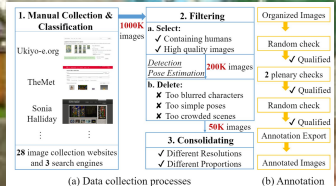
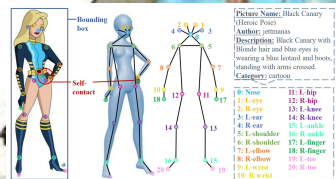
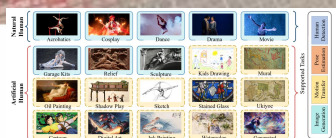


# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes

## Unique Characteristics

## Rich Scenario:

Covers **15 artificial** and **5 natural** human-related scenarios in various human appearances, diverse contexts, and various poses





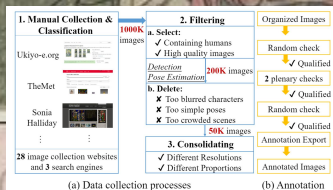
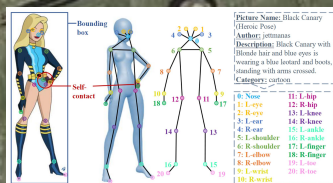
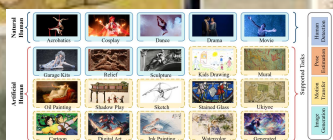


# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes

## Unique Characteristics

### Versatile annotations:

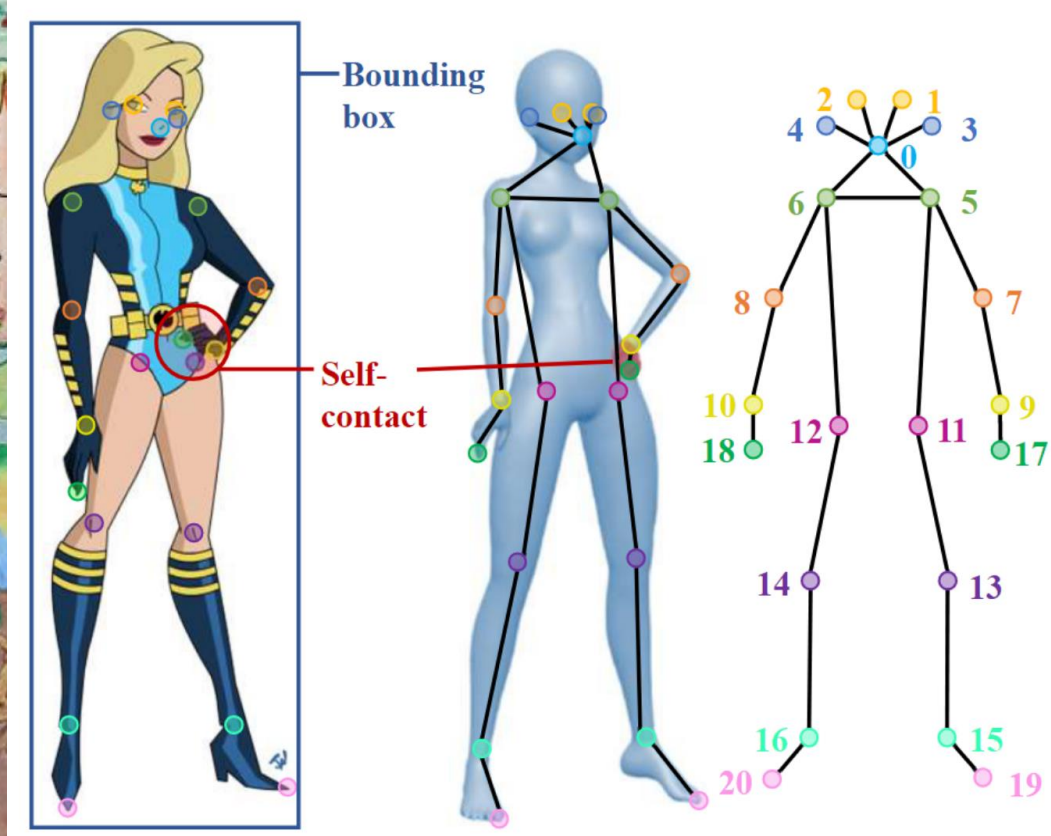
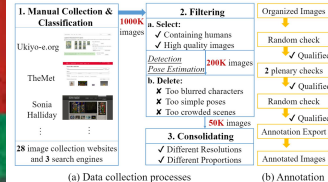
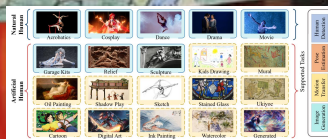
Provides manual annotations of 2D human keypoints, human bounding boxes, and self-contact keypoints for downstream tasks.





# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



**Picture Name:** Black Canary  
(Heroic Pose)

**Author:** jettmanas

**Description:** Black Canary with Blonde hair and blue eyes is wearing a blue leotard and boots, standing with arms crossed.

**Category:** cartoon

0: Nose	11: L-hip
1: L-eye	12: R-hip
2: R-eye	13: L-knee
3: L-ear	14: R-knee
4: R-ear	15: L-ankle
5: L-shoulder	16: R-ankle
6: R-shoulder	17: L-finger
7: L-elbow	18: R-finger
8: R-elbow	19: L-toe
9: L-wrist	20: R-toe
10: R-wrist	



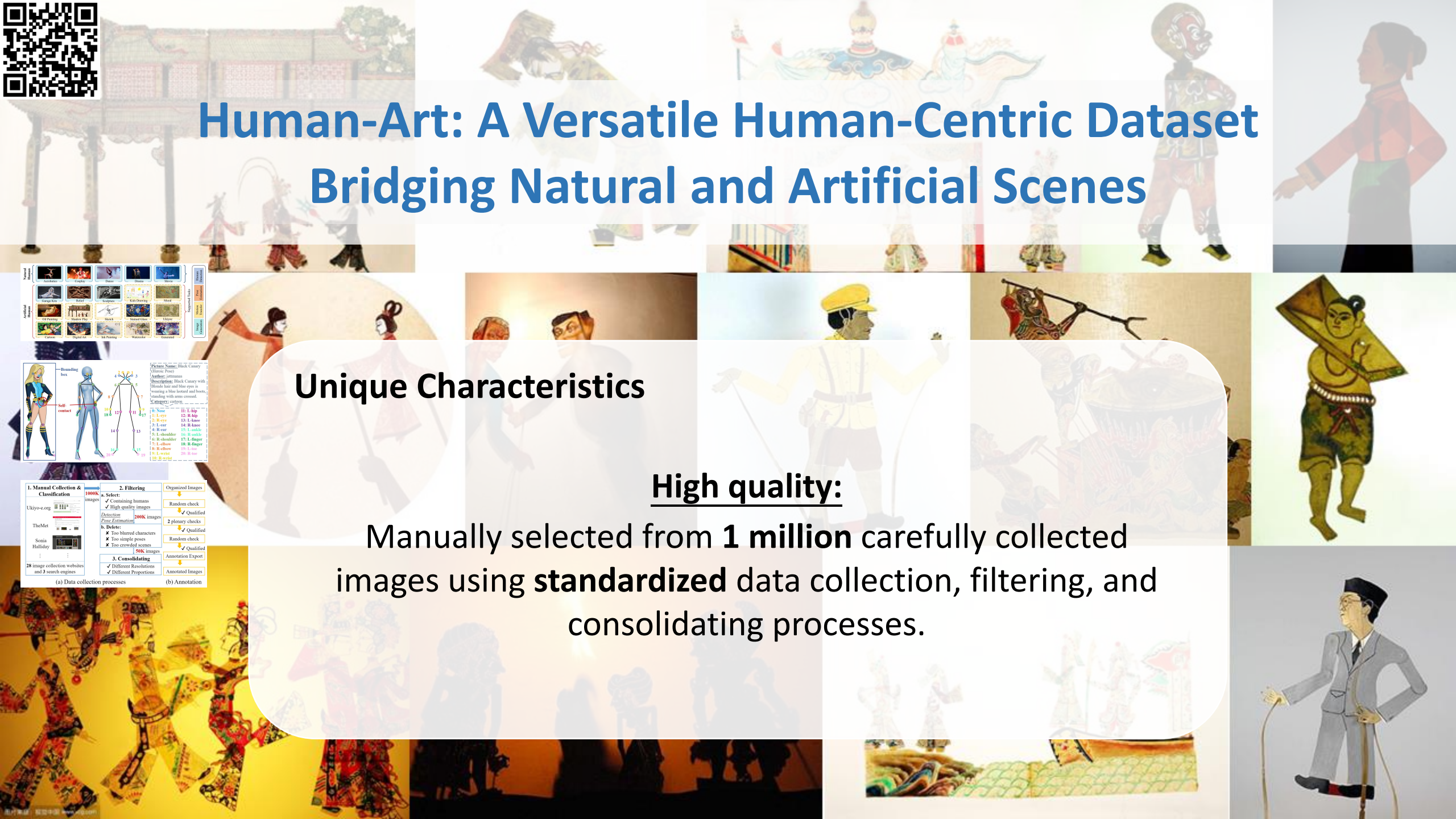
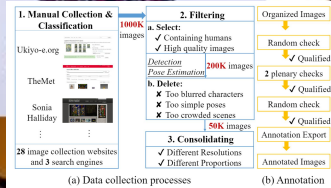
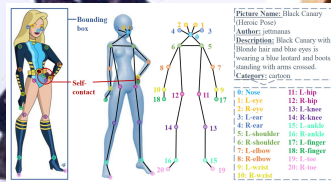
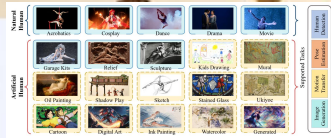


# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes

## Unique Characteristics

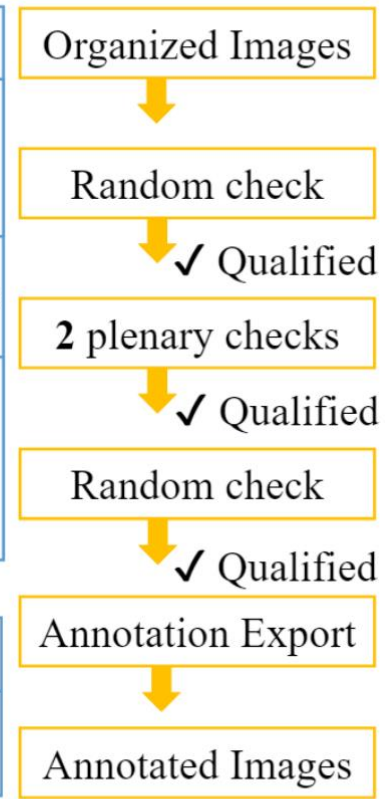
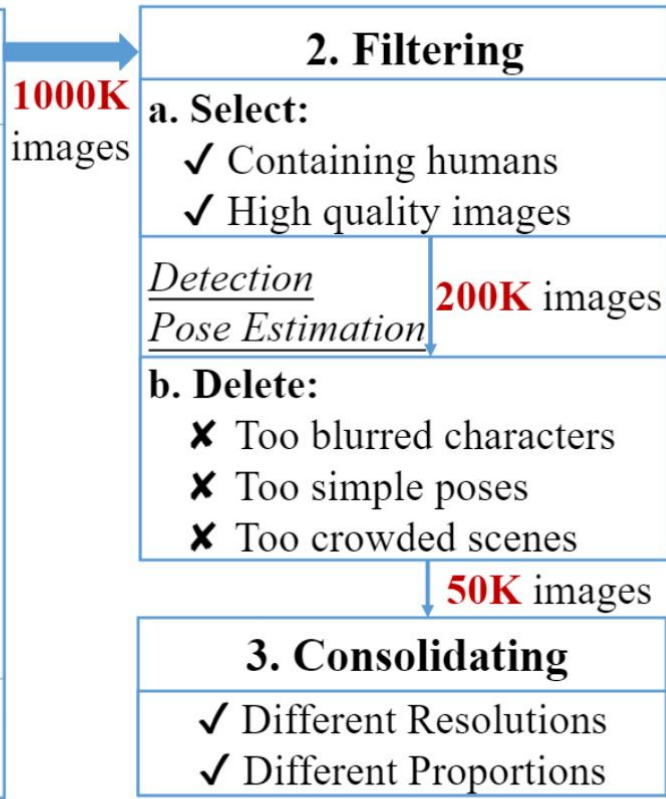
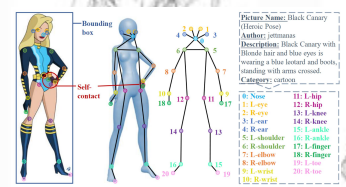
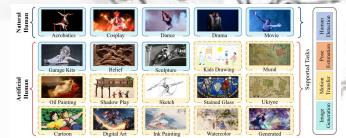
## High quality:

Manually selected from **1 million** carefully collected images using **standardized** data collection, filtering, and consolidating processes.



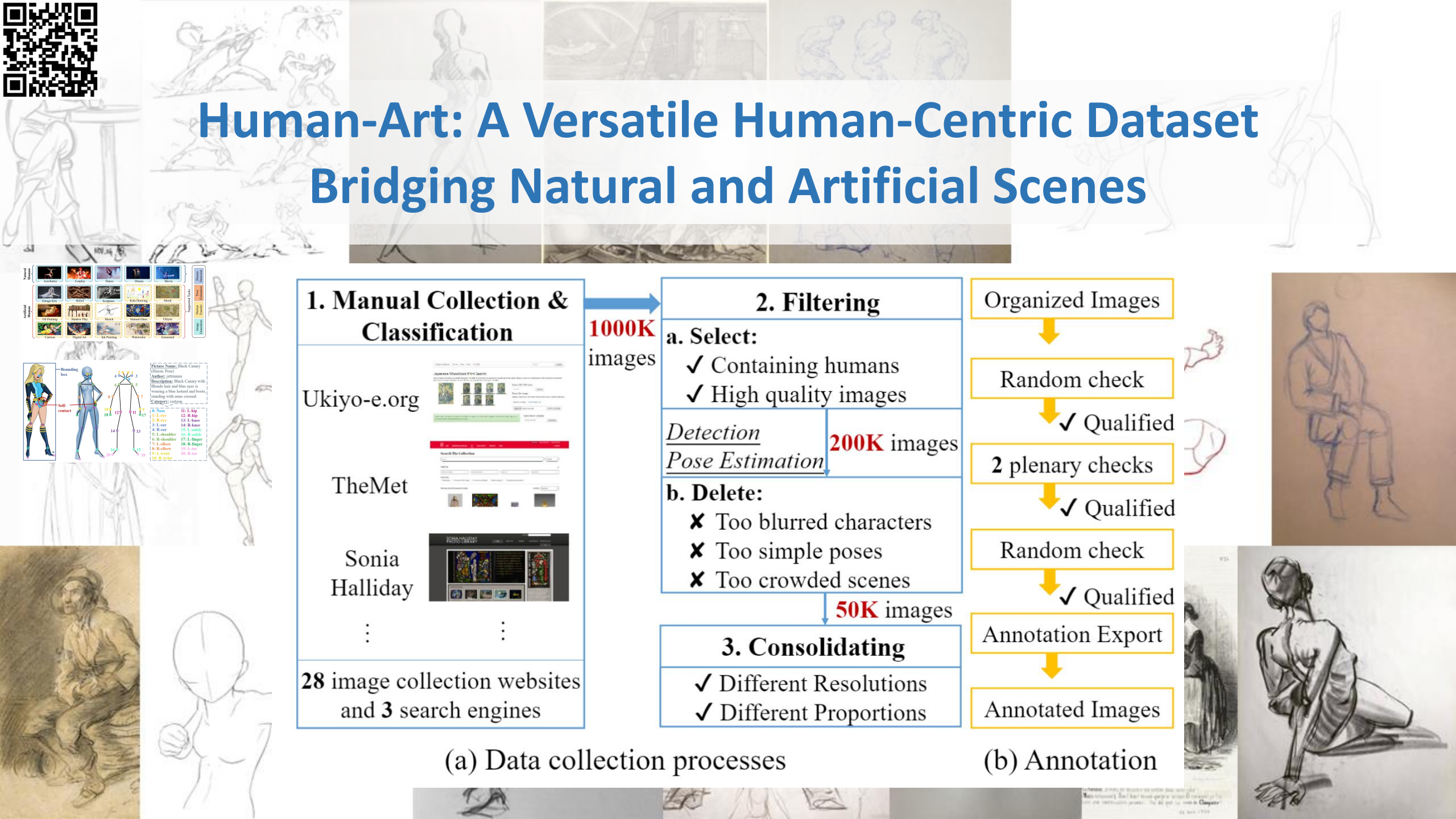
# Human-Art: A Versatile Human-Centric Dataset

## Bridging Natural and Artificial Scenes



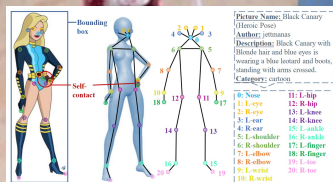
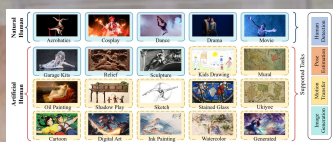
(a) Data collection processes

(b) Annotation

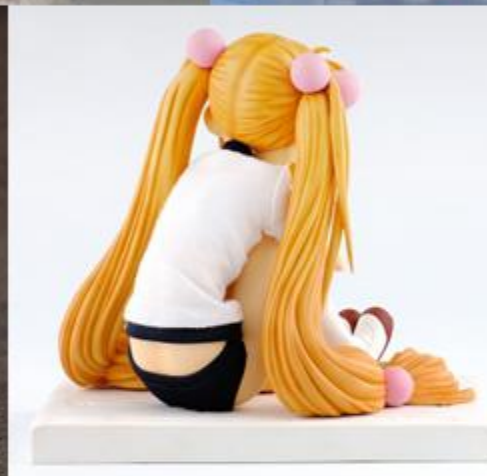




# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes

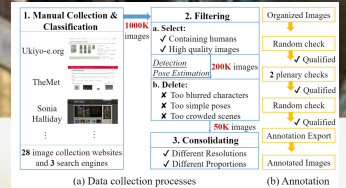
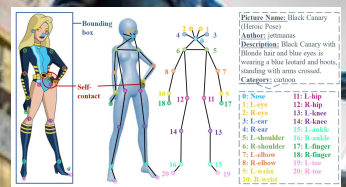
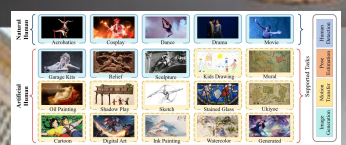


1. Manual Collection & Classification		2. Filtering		3. Consolidating	
1000K images	Ukyo-e-eg	200K images	50K images	28 image collection websites and 3 search engines	
		<ul style="list-style-type: none"> <li>✓ Consulting humans</li> <li>✓ High quality images</li> </ul>	<ul style="list-style-type: none"> <li>✓ Qualified</li> <li>2 primary checks</li> <li>✓ Qualified</li> </ul>	<ul style="list-style-type: none"> <li>✓ Different Resolutions</li> <li>✓ Different Proportions</li> </ul>	<ul style="list-style-type: none"> <li>Annotation Export</li> <li>Annotated Images</li> </ul>
(a) Data collection processes		(b) Annotation			

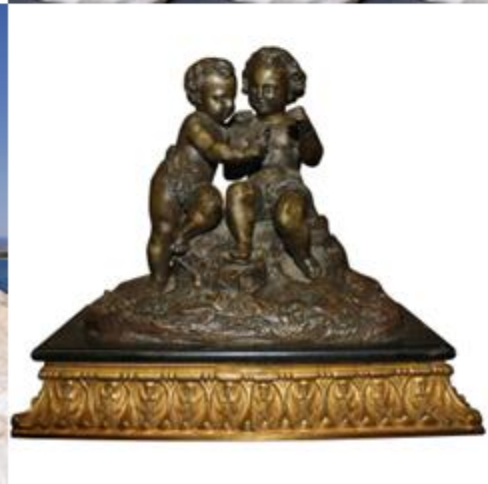
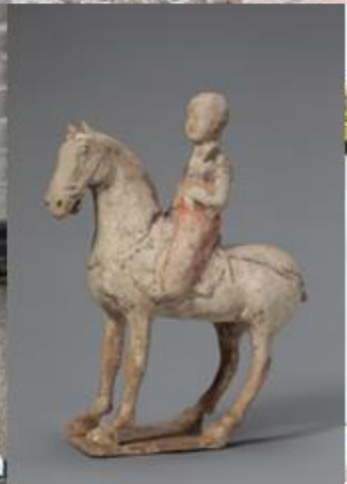




# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



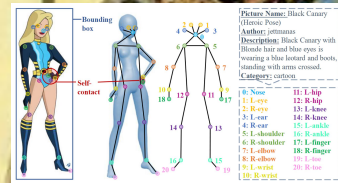
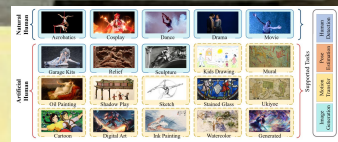
(a) Data collection processes (b) Annotation







# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



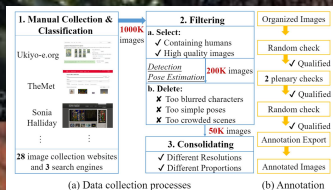
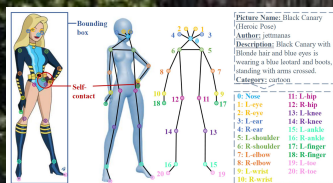
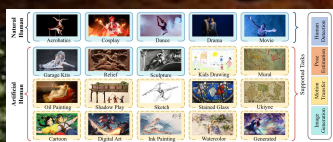
<b>1. Manual Collection &amp; Classification</b>	<b>2. Filtering</b>	<b>3. Consolidating</b>
1000k images Ukiyo-e.org TheMet Sonia Halliday 28 image collection websites and 3 search engines	<b>a. Select:</b> ✓ Containing humans ✓ High quality images <b>200k images</b> <b>b. Detect:</b> ✗ Too blurred characters ✗ Too simple poses ✗ Too crowded scenes <b>50k images</b>	✓ Different Resolutions ✓ Different Proportions <b>Annotated Images</b>

(a) Data collection processes (b) Annotation



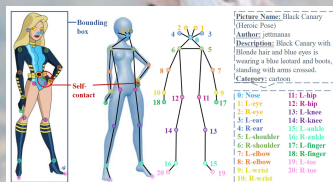
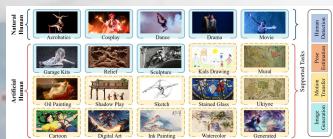


# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes

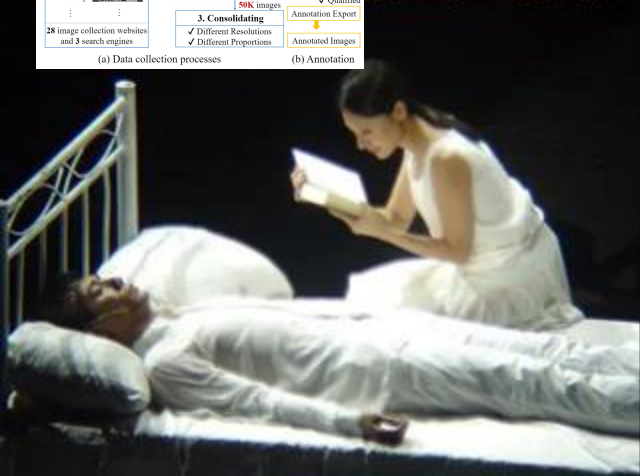
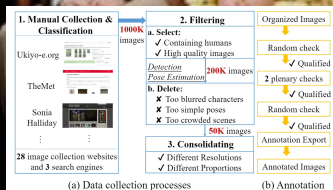
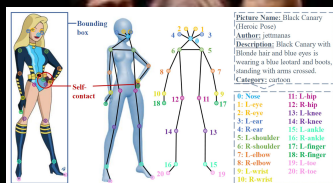
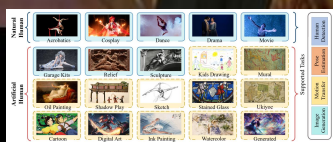


1. Manual Collection & Classification	2. Filtering	Organized Images
1000K images Ukayo-eog TheMet Sonia Halliday ... 28 image collection websites and 3 search engines	a. Select: <input checked="" type="checkbox"/> Containing humans <input checked="" type="checkbox"/> High quality images b. Detect: <input checked="" type="checkbox"/> Too blurred characters <input checked="" type="checkbox"/> Too simple poses <input checked="" type="checkbox"/> Too crowded scenes 200K images 50K images	Random check <input checked="" type="checkbox"/> Qualified 2 plenary checks <input checked="" type="checkbox"/> Qualified Random check <input checked="" type="checkbox"/> Qualified Annotation Export <input checked="" type="checkbox"/> Annotated Images
(a) Data collection processes	(b) Annotation	





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes

**1. Manual Collection & Classification**

- 1000k images
- Ukiyo-e.org
- TheMet
- Sonia Halliday
- 28 image collection websites and 3 search engines

**2. Filtering**

a. Select:

- Containing humans
- High quality images

Detection: 200k images

b. Delete:

- Too blurred characters
- Too simple poses
- Too crowded scenes

50k images

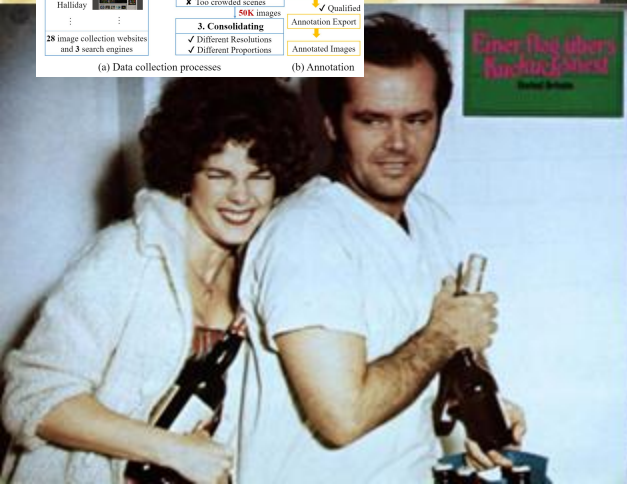
**3. Consolidating**

- Different Resolutions
- Different Proportions

Organized Images

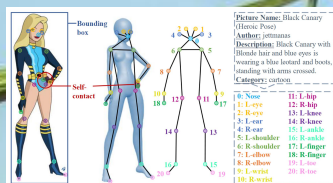
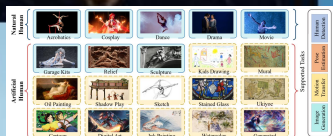
- Random check
- 2 plenary checks
- Random check
- Annotation Export
- Annotated Images

(a) Data collection processes (b) Annotation





# Human-Art: A Versatile Human-Centric Dataset Bridging Natural and Artificial Scenes



1. Manual Collection & Classification		2. Filtering	
1000k images	<ul style="list-style-type: none"> <li>✓ Containing humans</li> <li>✓ High quality images</li> </ul>	<ul style="list-style-type: none"> <li>✓ Qualified</li> <li>2 plenary checks</li> <li>✓ Qualified</li> <li>Random check</li> <li>✓ Qualified</li> </ul>	<ul style="list-style-type: none"> <li>Annotation Export</li> <li>Annotated Images</li> </ul>
28 image collection websites and 3 search engines	<p><b>3. Consolidating</b></p> <ul style="list-style-type: none"> <li>✓ Different Resolutions</li> <li>✓ Different Proportions</li> </ul>	<p><b>3. Consolidating</b></p> <ul style="list-style-type: none"> <li>✓ Different Resolutions</li> <li>✓ Different Proportions</li> </ul>	

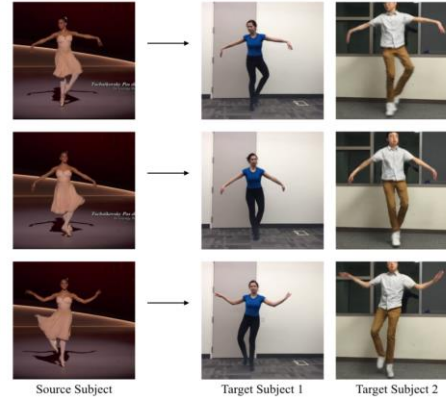
(a) Data collection processes (b) Annotation



## Human-centric computer vision tasks



Human Pose Estimation [1]



Motion Transfer [2]

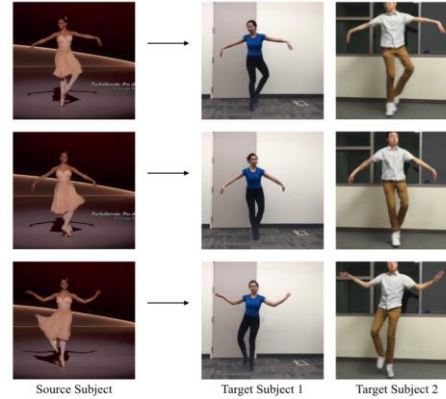


Human Image Generation [3]

## Human-centric computer vision tasks



Human Pose Estimation [1]



Motion Transfer [2]



Human Image Generation [3]

## Datasets



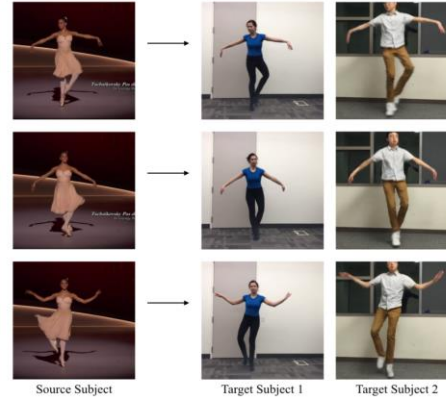
Existing Datasets [4]



## Human-centric computer vision tasks



Human Pose Estimation [1]



Motion Transfer [2]



Human Image Generation [3]

## Datasets



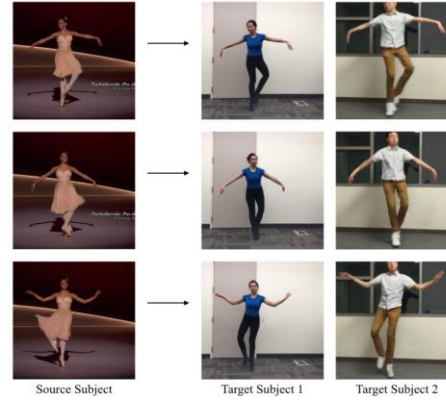
Existing Datasets [4]

**Only Natural Human**

## Human-centric computer vision tasks



Human Pose Estimation [1]



Motion Transfer [2]



Human Image Generation [3]

## Datasets



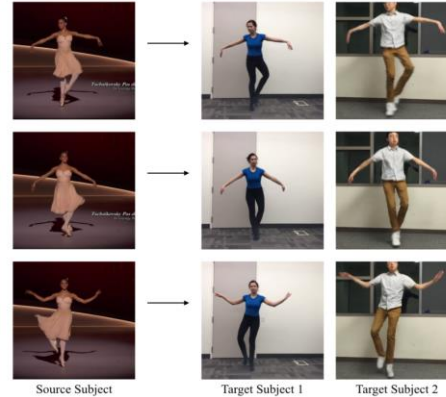
Existing Datasets [4]

## More Scenarios

## Human-centric computer vision tasks



Human Pose Estimation [1]



Motion Transfer [2]



Human Image Generation [3]

## Datasets



Existing Datasets [4]



More Scenarios

## Previous Datasets

Dataset	Image	Instance	Keypoint Number	Bbox	Pose	Self-Contact	Natural Scenario	Artificial Scenario
VOC2012 <sup>1</sup> [11]	8,174	17,132	-	✓	-	-	✓	-
MSCOCO <sup>1</sup> [32]	66,808	273,469	-	✓	-	-	✓	-
BodyHands [46]	20,490	63,095	-	✓	-	-	✓	-
People-Art [69]	1,490	3,870	-	✓	-	-	✓	✓
MSCOCO <sup>2</sup> [32]	58,945	156,165	17	✓	✓	-	✓	-
MPII [3]	24,920	40,522	16	✓	✓	-	✓	-
AI Challenger [71]	240,000	448,776	14	✓	✓	-	✓	-
CrowdPose [27]	20,000	~80,000	14	✓	✓	-	✓	-
OCHuman [85]	4,731	8,110	17	✓	✓	-	✓	-
PoseTrack <sup>3</sup> [2]	66,374	153,615	15	✓	✓	-	✓	-
HiEve <sup>3</sup> [33]	49,820	1,099,357	14	✓	✓	-	✓	-
ClassArch [40]	1513	1728	17	✓	✓	-	-	✓
Sketch2Pose [4]	808	14,772	18	✓	✓	✓	-	✓
<b>Human-Art (Ours)</b>	50,000	123,131	21	✓	✓	✓	✓	✓

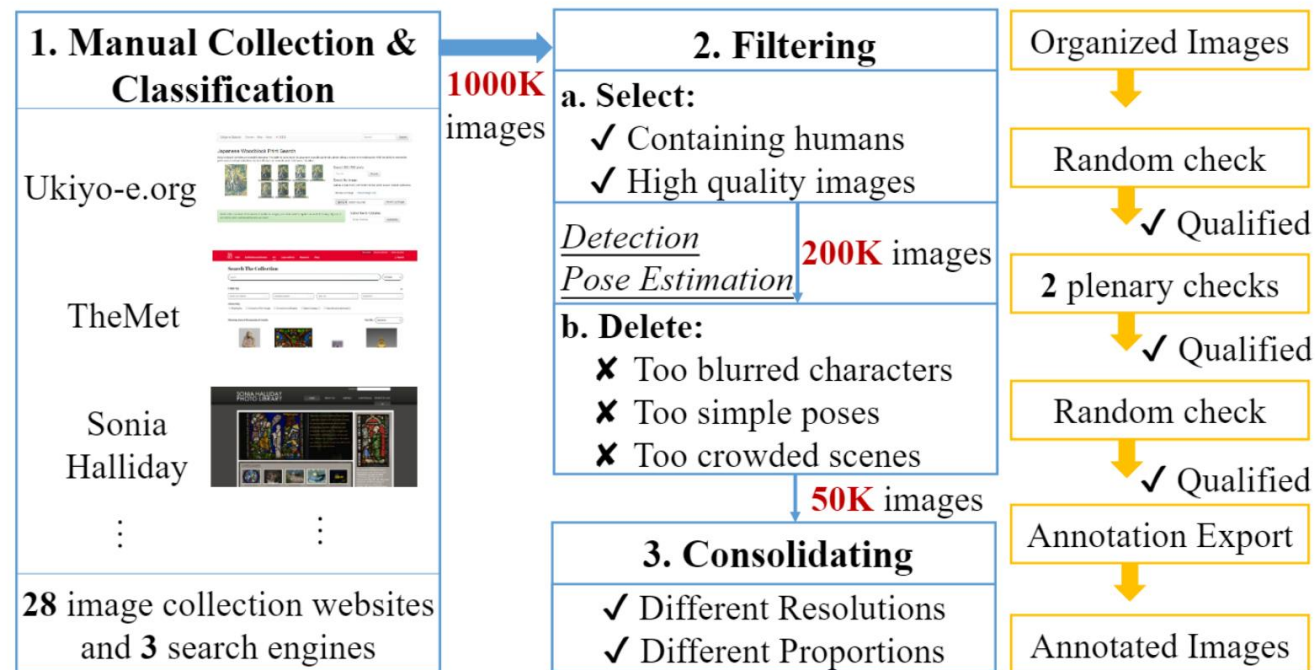
<sup>1</sup> Only calculate statistics of images that contain human bounding box annotation for detection;

<sup>2</sup> Only calculate statistics of images that contain human keypoint annotation for human pose estimation.

<sup>3</sup> Video-based datasets.

Comparison of human-centric recognition datasets, including human detection and pose estimation tasks.

- **5** types of **natural human** scenes: Acrobatics, Cosplay, Dance, Drama, and Movie;
- **3** types of **3D artificial human** scenes: Garage Kits, Relief, and Sculpture;
- **12** types of **2D artificial human** scenes: Kids Drawing, Mural, Oil Painting, Shadow Play, Sketch, Stained Glass, Ukiyoe, Cartoon, Digital Art, Ink Painting, Watercolor, and Generated Images;



(a) Data collection processes

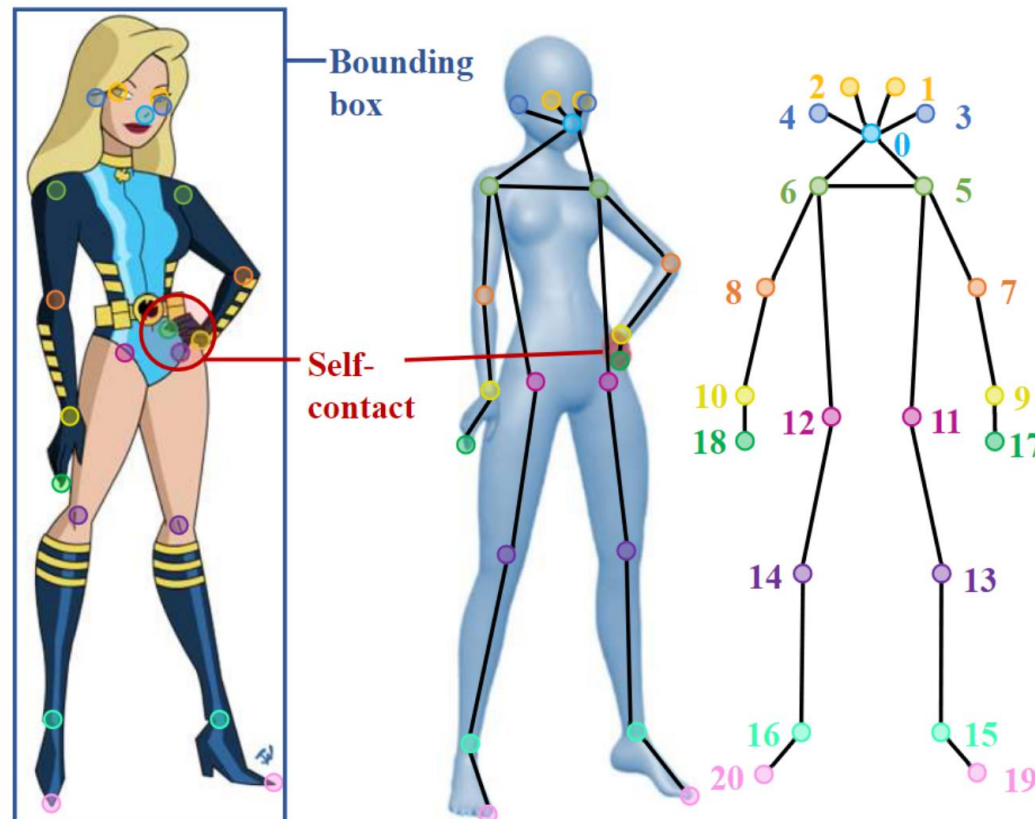
(b) Annotation

## Annotations

human bounding-box  
 21 human keypoints  
 self-contact keypoints  
 text information.

## Annotators

35 data annotators  
 12 data auditors



**Picture Name:** Black Canary

(Heroic Pose)

**Author:** jettmanas

**Description:** Black Canary with Blonde hair and blue eyes is wearing a blue leotard and boots, standing with arms crossed.

**Category:** cartoon

**0: Nose**

**1: L-eye**

**2: R-eye**

**3: L-ear**

**4: R-ear**

**5: L-shoulder**

**6: R-shoulder**

**7: L-elbow**

**8: R-elbow**

**9: L-wrist**

**10: R-wrist**

**11: L-hip**

**12: R-hip**

**13: L-knee**

**14: R-knee**

**15: L-ankle**

**16: R-ankle**

**17: L-finger**

**18: R-finger**

**19: L-toe**

**20: R-toe**

## Human-Centric Recognition

## Human Detection

Detector		Faster R-CNN				YOLOX		Deformable DETR		DINO		
Setting		val	val *	test	test *	val	test	val	test	val	test	
MSCOCO [32]		<b>52.2</b>	51.6	-	-	<b>61.9</b>	-	<b>57.2</b>	-	<b>63.2</b>	-	
Artificial Scene	2D Representation	Cartoon	8.8	37.9	7.0	<u>33.5</u>	10.8	9.2	7.9	6.7	8.7	8.1
		Digital Art	18.8	46.4	17.8	44.2	24.1	22.9	17.6	15.5	18.6	18.1
		Ink Painting	11.0	37.7	9.1	37.2	15.5	13.0	11.9	10.0	14.5	11.6
		Kids Drawing	6.6	54.2	8.0	53.6	6.8	11.5	5.6	7.2	6.8	8.2
		Mural	9.7	35.5	9.3	34.5	12.2	12.2	9.3	8.1	10.2	9.5
		Oil Painting	15.9	41.1	13.7	37.5	20.8	18.3	17.1	14.2	17.0	15.0
		Shadow Play	7.5	<b>64.1</b>	8.2	<b>63.7</b>	5.4	7.5	5.3	5.1	6.4	7.9
		Sketch	<u>2.6</u>	48.8	<u>2.4</u>	55.7	<u>4.6</u>	<u>5.2</u>	5.8	9.2	<u>3.6</u>	7.1
		Stained Glass	8.8	<u>35.0</u>	8.1	34.7	8.2	7.8	5.1	<u>4.6</u>	7.8	7.8
		Ukiyoe	12.7	51.9	12.7	50.3	13.1	12.8	8.5	8.4	11.4	11.4
Watercolor	14.8	42.8	14.2	42.2	19.7	18.2	15.6	13.6	15.4	14.3		
Natural Scene	3D Representation	Garage Kits	22.9	60.0	22.5	62.5	22.3	19.9	17.9	14.6	22.8	19.5
		Relief	4.9	37.5	4.7	33.4	8.4	9.1	<u>4.7</u>	5.7	4.4	<u>5.9</u>
		Sculpture	17.7	48.6	14.4	47.0	15.8	13.2	9.4	7.1	10.1	8.5
		Acrobatics	17.0	49.7	17.0	53.4	20.0	19.4	17.3	17.6	19.4	18.9
		Cosplay	31.2	52.8	<b>31.3</b>	56.7	38.0	<b>37.2</b>	34.6	<b>34.5</b>	37.2	<b>36.7</b>
		Dance	17.0	46.6	18.4	49.3	20.3	21.1	17.8	18.5	19.3	19.6
Drama	24.3	46.0	24.8	48.7	27.4	27.5	15.4	25.8	27.8	16.7		
Movie	26.3	36.5	25.0	37.2	28.0	26.8	26.6	26.2	27.2	26.3		
Average		12.0	44.2	12.5	43.0	14.4	14.7	11.7	11.7	12.6	12.7	

\* the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Detection

Detector		Faster R-CNN				YOLOX		Deformable DETR		DINO		
Setting		val	val*	test	test*	val	test	val	test	val	test	
MSCOCO [32]		<b>52.2</b>	51.6	-	-	<b>61.9</b>	-	<b>57.2</b>	-	<b>63.2</b>	-	
Artificial Scene	2D Representation	Cartoon	8.8	37.9	7.0	<u>33.5</u>	10.8	9.2	7.9	6.7	8.7	8.1
		Digital Art	18.8	46.4	17.8	44.2	24.1	22.9	17.6	15.5	18.6	18.1
		Ink Painting	11.0	37.7	9.1	37.2	15.5	13.0	11.9	10.0	14.5	11.6
		Kids Drawing	6.6	54.2	8.0	53.6	6.8	11.5	5.6	7.2	6.8	8.2
		Mural	9.7	35.5	9.3	34.5	12.2	12.2	9.3	8.1	10.2	9.5
		Oil Painting	15.9	41.1	13.7	37.5	20.8	18.3	17.1	14.2	17.0	15.0
		Shadow Play	7.5	<b>64.1</b>	8.2	<b>63.7</b>	5.4	7.5	5.3	5.1	6.4	7.9
		Sketch	<u>2.6</u>	48.8	<u>2.4</u>	55.7	<u>4.6</u>	<u>5.2</u>	5.8	9.2	<u>3.6</u>	7.1
		Stained Glass	8.8	<u>35.0</u>	8.1	34.7	8.2	7.8	5.1	<u>4.6</u>	7.8	7.8
		Ukiyoe	12.7	51.9	12.7	50.3	13.1	12.8	8.5	8.4	11.4	11.4
Watercolor	14.8	42.8	14.2	42.2	19.7	18.2	15.6	13.6	15.4	14.3		
Natural Scene	3D Representation	Garage Kits	22.9	60.0	22.5	62.5	22.3	19.9	17.9	14.6	22.8	19.5
		Relief	4.9	37.5	4.7	33.4	8.4	9.1	<u>4.7</u>	5.7	4.4	<u>5.9</u>
		Sculpture	17.7	48.6	14.4	47.0	15.8	13.2	9.4	7.1	10.1	8.5
		Acrobatics	17.0	49.7	17.0	53.4	20.0	19.4	17.3	17.6	19.4	18.9
		Cosplay	31.2	52.8	<b>31.3</b>	56.7	38.0	<b>37.2</b>	34.6	<b>34.5</b>	37.2	<b>36.7</b>
		Dance	17.0	46.6	18.4	49.3	20.3	21.1	17.8	18.5	19.3	19.6
Drama	24.3	46.0	24.8	48.7	27.4	27.5	15.4	25.8	27.8	16.7		
Movie	26.3	36.5	25.0	37.2	28.0	26.8	26.6	26.2	27.2	26.3		
Average		12.0	44.2	12.5	43.0	14.4	14.7	11.7	11.7	12.6	12.7	

\* the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.



## Human-Centric Recognition

## Human Detection

Detector		Faster R-CNN				YOLOX		Deformable DETR		DINO		
Setting		val	val *	test	test *	val	test	val	test	val	test	
MSCOCO [32]		<b>52.2</b>	51.6	-	-	<b>61.9</b>	-	<b>57.2</b>	-	<b>63.2</b>	-	
Artificial Scene	2D Representation	Cartoon	8.8	37.9	7.0	<u>33.5</u>	10.8	9.2	7.9	6.7	8.7	8.1
		Digital Art	18.8	46.4	17.8	44.2	24.1	22.9	17.6	15.5	18.6	18.1
		Ink Painting	11.0	37.7	9.1	37.2	15.5	13.0	11.9	10.0	14.5	11.6
		Kids Drawing	6.6	54.2	8.0	53.6	6.8	11.5	5.6	7.2	6.8	8.2
		Mural	9.7	35.5	9.3	34.5	12.2	12.2	9.3	8.1	10.2	9.5
		Oil Painting	15.9	41.1	13.7	37.5	20.8	18.3	17.1	14.2	17.0	15.0
		Shadow Play	7.5	<b>64.1</b>	8.2	<b>63.7</b>	5.4	7.5	5.3	5.1	6.4	7.9
		Sketch	<u>2.6</u>	48.8	<u>2.4</u>	55.7	<u>4.6</u>	<u>5.2</u>	5.8	9.2	<u>3.6</u>	7.1
		Stained Glass	8.8	<u>35.0</u>	8.1	34.7	8.2	7.8	5.1	<u>4.6</u>	7.8	7.8
		Ukiyoe	12.7	51.9	12.7	50.3	13.1	12.8	8.5	8.4	11.4	11.4
Watercolor	14.8	42.8	14.2	42.2	19.7	18.2	15.6	13.6	15.4	14.3		
Natural Scene	3D Representation	Garage Kits	22.9	60.0	22.5	62.5	22.3	19.9	17.9	14.6	22.8	19.5
		Relief	4.9	37.5	4.7	33.4	8.4	9.1	<u>4.7</u>	5.7	4.4	<u>5.9</u>
		Sculpture	17.7	48.6	14.4	47.0	15.8	13.2	9.4	7.1	10.1	8.5
		Acrobatics	17.0	49.7	17.0	53.4	20.0	19.4	17.3	17.6	19.4	18.9
		Cosplay	31.2	52.8	<b>31.3</b>	56.7	38.0	<b>37.2</b>	34.6	<b>34.5</b>	37.2	<b>36.7</b>
		Dance	17.0	46.6	18.4	49.3	20.3	21.1	17.8	18.5	19.3	19.6
Drama	24.3	46.0	24.8	48.7	27.4	27.5	15.4	25.8	27.8	16.7		
Movie	26.3	36.5	25.0	37.2	28.0	26.8	26.6	26.2	27.2	26.3		
Average		12.0	44.2	12.5	43.0	14.4	14.7	11.7	11.7	12.6	12.7	

\* the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Detection

Detector		Faster R-CNN				YOLOX		Deformable DETR		DINO		
Setting		val	val*	test	test*	val	test	val	test	val	test	
MSCOCO [32]		<b>52.2</b>	51.6	-	-	<b>61.9</b>	-	<b>57.2</b>	-	<b>63.2</b>	-	
Artificial Scene	2D Representation	Cartoon	8.8	37.9	7.0	33.5	10.8	9.2	7.9	6.7	8.7	8.1
		Digital Art	18.8	46.4	17.8	44.2	24.1	22.9	17.6	15.5	18.6	18.1
		Ink Painting	11.0	37.7	9.1	37.2	15.5	13.0	11.9	10.0	14.5	11.6
		Kids Drawing	6.6	54.2	8.0	53.6	6.8	11.5	5.6	7.2	6.8	8.2
		Mural	9.7	35.5	9.3	34.5	12.2	12.2	9.3	8.1	10.2	9.5
		Oil Painting	15.9	41.1	13.7	37.5	20.8	18.3	17.1	14.2	17.0	15.0
		Shadow Play	7.5	<b>64.1</b>	8.2	<b>63.7</b>	5.4	7.5	5.3	5.1	6.4	7.9
		Sketch	<u>2.6</u>	48.8	<u>2.4</u>	55.7	<u>4.6</u>	<u>5.2</u>	5.8	9.2	<u>3.6</u>	7.1
		Stained Glass	8.8	<u>35.0</u>	8.1	34.7	8.2	7.8	5.1	<u>4.6</u>	7.8	7.8
		Ukiyoe	12.7	51.9	12.7	50.3	13.1	12.8	8.5	8.4	11.4	11.4
Watercolor	14.8	42.8	14.2	42.2	19.7	18.2	15.6	13.6	15.4	14.3		
Natural Scene	3D Representation	Garage Kits	22.9	60.0	22.5	62.5	22.3	19.9	17.9	14.6	22.8	19.5
		Relief	4.9	37.5	4.7	33.4	8.4	9.1	<u>4.7</u>	5.7	4.4	<u>5.9</u>
		Sculpture	17.7	48.6	14.4	47.0	15.8	13.2	9.4	7.1	10.1	8.5
		Acrobatics	17.0	49.7	17.0	53.4	20.0	19.4	17.3	17.6	19.4	18.9
		Cosplay	31.2	52.8	<b>31.3</b>	56.7	38.0	<b>37.2</b>	34.6	<b>34.5</b>	37.2	<b>36.7</b>
		Dance	17.0	46.6	18.4	49.3	20.3	21.1	17.8	18.5	19.3	19.6
Drama	24.3	46.0	24.8	48.7	27.4	27.5	15.4	25.8	27.8	16.7		
Movie	26.3	36.5	25.0	37.2	28.0	26.8	26.6	26.2	27.2	26.3		
Average		<b>12.0</b>	44.2	12.5	43.0	<b>14.4</b>	14.7	<b>11.7</b>	11.7	<b>12.6</b>	12.7	

\* the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Detection

Detector		Faster R-CNN				YOLOX		Deformable DETR		DINO		
		val	val *	test	test *	val	test	val	test	val	test	
MSCOCO [32]		<b>52.2</b>	51.6	-	-	<b>61.9</b>	-	<b>57.2</b>	-	<b>63.2</b>	-	
Artificial Scene	2D Representation	Cartoon	8.8	37.9	7.0	<u>33.5</u>	10.8	9.2	7.9	6.7	8.7	8.1
		Digital Art	18.8	46.4	17.8	44.2	24.1	22.9	17.6	15.5	18.6	18.1
		Ink Painting	11.0	37.7	9.1	37.2	15.5	13.0	11.9	10.0	14.5	11.6
		Kids Drawing	6.6	54.2	8.0	53.6	6.8	11.5	5.6	7.2	6.8	8.2
		Mural	9.7	35.5	9.3	34.5	12.2	12.2	9.3	8.1	10.2	9.5
		Oil Painting	15.9	41.1	13.7	37.5	20.8	18.3	17.1	14.2	17.0	15.0
		Shadow Play	7.5	<b>64.1</b>	8.2	<b>63.7</b>	5.4	7.5	5.3	5.1	6.4	7.9
		Sketch	<u>2.6</u>	48.8	<u>2.4</u>	55.7	<u>4.6</u>	<u>5.2</u>	5.8	9.2	<u>3.6</u>	7.1
		Stained Glass	8.8	<u>35.0</u>	8.1	34.7	8.2	7.8	5.1	<u>4.6</u>	7.8	7.8
		Ukiyoe	12.7	51.9	12.7	50.3	13.1	12.8	8.5	8.4	11.4	11.4
Watercolor	14.8	42.8	14.2	42.2	19.7	18.2	15.6	13.6	15.4	14.3		
Natural Scene	3D Representation	Garage Kits	22.9	60.0	22.5	62.5	22.3	19.9	17.9	14.6	22.8	19.5
		Relief	4.9	37.5	4.7	33.4	8.4	9.1	<u>4.7</u>	5.7	4.4	<u>5.9</u>
		Sculpture	17.7	48.6	14.4	47.0	15.8	13.2	9.4	7.1	10.1	8.5
		Acrobatics	17.0	49.7	17.0	53.4	20.0	19.4	17.3	17.6	19.4	18.9
		Cosplay	31.2	52.8	<b>31.3</b>	56.7	38.0	<b>37.2</b>	34.6	<b>34.5</b>	37.2	<b>36.7</b>
		Dance	17.0	46.6	18.4	49.3	20.3	21.1	17.8	18.5	19.3	19.6
Drama	24.3	46.0	24.8	48.7	27.4	27.5	15.4	25.8	27.8	16.7		
Movie	26.3	36.5	25.0	37.2	28.0	26.8	26.6	26.2	27.2	26.3		
Average		12.0	44.2	12.5	43.0	14.4	14.7	11.7	11.7	12.6	12.7	

\* the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Detection

Detector		Faster R-CNN				YOLOX		Deformable DETR		DINO		
Setting		val	val *	test	test *	val	test	val	test	val	test	
MSCOCO [32]		<b>52.2</b>	51.6	-	-	<b>61.9</b>	-	<b>57.2</b>	-	<b>63.2</b>	-	
Artificial Scene	2D Representation	Cartoon	8.8	37.9	7.0	<u>33.5</u>	10.8	9.2	7.9	6.7	8.7	8.1
		Digital Art	18.8	46.4	17.8	44.2	24.1	22.9	17.6	15.5	18.6	18.1
		Ink Painting	11.0	37.7	9.1	37.2	15.5	13.0	11.9	10.0	14.5	11.6
		Kids Drawing	6.6	54.2	8.0	53.6	6.8	11.5	5.6	7.2	6.8	8.2
		Mural	9.7	35.5	9.3	34.5	12.2	12.2	9.3	8.1	10.2	9.5
		Oil Painting	15.9	41.1	13.7	37.5	20.8	18.3	17.1	14.2	17.0	15.0
		Shadow Play	7.5	<b>64.1</b>	8.2	<b>63.7</b>	5.4	7.5	5.3	5.1	6.4	7.9
		Sketch	<u>2.6</u>	48.8	<u>2.4</u>	55.7	<u>4.6</u>	<u>5.2</u>	5.8	9.2	<u>3.6</u>	7.1
		Stained Glass	8.8	<u>35.0</u>	8.1	34.7	8.2	7.8	5.1	<u>4.6</u>	7.8	7.8
		Ukiyoe	12.7	51.9	12.7	50.3	13.1	12.8	8.5	8.4	11.4	11.4
Watercolor	14.8	42.8	14.2	42.2	19.7	18.2	15.6	13.6	15.4	14.3		
Natural Scene	3D Representation	Garage Kits	22.9	60.0	22.5	62.5	22.3	19.9	17.9	14.6	22.8	19.5
		Relief	4.9	37.5	4.7	33.4	8.4	9.1	<u>4.7</u>	5.7	4.4	<u>5.9</u>
		Sculpture	17.7	48.6	14.4	47.0	15.8	13.2	9.4	7.1	10.1	8.5
		Acrobatics	17.0	49.7	17.0	53.4	20.0	19.4	17.3	17.6	19.4	18.9
		Cosplay	31.2	52.8	<b>31.3</b>	56.7	38.0	<b>37.2</b>	34.6	<b>34.5</b>	37.2	<b>36.7</b>
		Dance	17.0	46.6	18.4	49.3	20.3	21.1	17.8	18.5	19.3	19.6
Drama	24.3	46.0	24.8	48.7	27.4	27.5	15.4	25.8	27.8	16.7		
Movie	26.3	36.5	25.0	37.2	28.0	26.8	26.6	26.2	27.2	26.3		
Average		12.0	44.2	12.5	43.0	14.4	14.7	11.7	11.7	12.6	12.7	

\* the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Pose Estimation

Detector		Faster R-CNN + HRNet						YOLOX + ViTPose				HigherHRNet		ED-Pose				
Setting		val	val <sup>‡</sup>	val <sup>*‡</sup>	test	test <sup>‡</sup>	test <sup>*‡</sup>	val	val <sup>‡</sup>	test	test <sup>‡</sup>	val	test	val	val <sup>*</sup>	test	test <sup>*</sup>	
MSCOCO [32]		<b>75.6</b>	77.6	77.2	<b>73.4</b>	-	-	79.8	82.3	81.1	-	68.6	70.5	71.6	72.4	69.8	-	
Artificial Scene	2D Representation	Cartoon	9.7	37.6	64.7	7.3	34.4	61.0	16.3	55.1	13.0	50.5	15.7	12.0	22.2	60.4	18.0	57.1
		Digital Art	22.6	59.6	74.4	25.8	61.2	75.7	29.0	69.9	31.9	72.2	42.5	44.3	43.5	71.4	45.6	75.1
		Ink Painting	6.3	51.4	72.1	5.6	48.0	72.4	8.9	59.8	9.2	58.2	26.8	20.9	28.2	56.8	24.9	55.4
		Kids Drawing	10.5	40.8	86.1	10.0	44.6	85.9	14.0	59.2	13.2	62.6	12.6	13.8	20.7	76.7	23.2	78.8
		Mural	11.6	54.0	71.1	12.3	53.6	71.6	15.9	50.1	16.4	51.6	30.6	32.0	34.6	64.7	35.1	65.4
		Oil Painting	31.6	65.7	78.1	28.5	62.2	75.6	39.6	73.4	36.7	70.5	54.4	51.1	56.2	75.2	51.7	71.4
		Shadow Play	<u>5.4</u>	<u>15.9</u>	<u>59.8</u>	<u>5.0</u>	<u>17.2</u>	<u>58.4</u>	<u>8.1</u>	<u>29.2</u>	<u>8.4</u>	<u>26.0</u>	<u>4.4</u>	<u>6.5</u>	<u>6.0</u>	<u>38.5</u>	<u>7.7</u>	<u>39.7</u>
		Sketch	6.3	44.1	73.1	6.7	57.2	79.4	9.1	61.3	10.9	71.3	13.6	<u>6.3</u>	12.0	66.8	12.2	75.8
		Stained Glass	10.4	46.0	74.8	9.7	45.1	73.1	12.0	59.1	12.1	58.1	26.6	23.1	27.6	74.4	25.6	71.5
		Ukiyoe	17.8	48.1	82.4	18.8	47.7	81.8	23.8	61.2	26.8	63.1	20.2	19.4	25.0	83.6	25.8	83.6
Watercolor	26.7	60.1	73.9	25.5	57.6	73.4	36.4	71.0	36.1	69.0	48.9	43.4	50.6	73.5	45.6	71.3		
Natural Scene	3D Representation	Garage Kits	45.2	57.5	86.7	44.5	61.4	<b>89.2</b>	52.5	76.2	50.6	77.0	37.4	34.7	47.9	87.7	44.1	90.1
		Relief	10.5	57.3	78.7	7.9	53.4	76.0	16.2	70.8	14.9	67.1	32.5	29.8	28.0	70.6	27.1	67.6
		Sculpture	36.4	65.9	81.0	38.5	64.0	78.5	34.9	78.5	34.2	73.7	33.5	35.2	45.9	76.9	46.7	74.7
		Acrobatics	45.8	68.0	85.2	46.6	68.4	83.2	69.1	86.8	66.3	83.9	58.6	57.4	41.4	80.0	44.4	78.9
		Cosplay	71.0	<b>81.1</b>	<b>87.2</b>	72.6	<b>81.9</b>	87.0	<b>80.0</b>	<b>90.3</b>	<b>81.7</b>	<b>88.8</b>	<b>78.1</b>	<b>77.8</b>	<b>79.6</b>	<b>89.1</b>	<b>79.7</b>	<b>90.4</b>
		Dance	43.1	67.3	77.2	49.2	70.1	80.1	57.3	81.5	61.5	83.8	51.4	62.4	53.6	76.5	61.2	82.2
Drama	45.3	75.1	82.0	46.7	75.8	83.1	54.2	83.9	56.9	84.8	59.6	72.2	75.0	85.9	76.0	86.1		
Movie	49.5	71.5	77.2	50.4	72.2	76.2	57.6	76.8	56.5	78.6	54.9	65.8	69.2	82.2	68.2	80.4		
Average		22.2	55.2	76.4	24.1	55.4	76.0	28.7	67.7	30.7	67.5	34.6	36.3	37.5	72.3	39.2	72.7	

<sup>‡</sup> the top-down pose estimation results that use ground truth bounding box;

<sup>\*</sup> the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Pose Estimation

Detector		Faster R-CNN + HRNet					YOLOX + ViTPose				HigherHRNet		ED-Pose					
Setting		val	val <sup>‡</sup>	val <sup>*‡</sup>	test	test <sup>‡</sup>	test <sup>*‡</sup>	val	val <sup>‡</sup>	test	test	val	test	val	val <sup>*</sup>	test	test <sup>*</sup>	
MSCOCO [32]		<b>75.6</b>	77.6	77.2	<b>73.4</b>	-	-	79.8	82.3	81.1	-	68.6	70.5	71.6	72.4	69.8	-	
Artificial Scene	2D Representation	Cartoon	9.7	37.6	64.7	7.3	34.4	61.0	16.3	55.1	13.0	50.5	15.7	12.0	22.2	60.4	18.0	57.1
		Digital Art	22.6	59.6	74.4	25.8	61.2	75.7	29.0	69.9	31.9	72.2	42.5	44.3	43.5	71.4	45.6	75.1
		Ink Painting	6.3	51.4	72.1	5.6	48.0	72.4	8.9	59.8	9.2	58.2	26.8	20.9	28.2	56.8	24.9	55.4
		Kids Drawing	10.5	40.8	86.1	10.0	44.6	85.9	14.0	59.2	13.2	62.6	12.6	13.8	20.7	76.7	23.2	78.8
		Mural	11.6	54.0	71.1	12.3	53.6	71.6	15.9	50.1	16.4	51.6	30.6	32.0	34.6	64.7	35.1	65.4
		Oil Painting	31.6	65.7	78.1	28.5	62.2	75.6	39.6	73.4	36.7	70.5	54.4	51.1	56.2	75.2	51.7	71.4
		Shadow Play	<u>5.4</u>	<u>15.9</u>	<u>59.8</u>	<u>5.0</u>	<u>17.2</u>	<u>58.4</u>	<u>8.1</u>	<u>29.2</u>	<u>8.4</u>	<u>26.0</u>	<u>4.4</u>	<u>6.5</u>	<u>6.0</u>	<u>38.5</u>	<u>7.7</u>	<u>39.7</u>
		Sketch	6.3	44.1	73.1	6.7	57.2	79.4	9.1	61.3	10.9	71.3	13.6	<u>6.3</u>	12.0	66.8	12.2	75.8
		Stained Glass	10.4	46.0	74.8	9.7	45.1	73.1	12.0	59.1	12.1	58.1	26.6	23.1	27.6	74.4	25.6	71.5
		Ukiyoe	17.8	48.1	82.4	18.8	47.7	81.8	23.8	61.2	26.8	63.1	20.2	19.4	25.0	83.6	25.8	83.6
Watercolor	26.7	60.1	73.9	25.5	57.6	73.4	36.4	71.0	36.1	69.0	48.9	43.4	50.6	73.5	45.6	71.3		
Natural Scene	3D Representation	Garage Kits	45.2	57.5	86.7	44.5	61.4	<b>89.2</b>	52.5	76.2	50.6	77.0	37.4	34.7	47.9	87.7	44.1	90.1
		Relief	10.5	57.3	78.7	7.9	53.4	76.0	16.2	70.8	14.9	67.1	32.5	29.8	28.0	70.6	27.1	67.6
		Sculpture	36.4	65.9	81.0	38.5	64.0	78.5	34.9	78.5	34.2	73.7	33.5	35.2	45.9	76.9	46.7	74.7
		Acrobatics	45.8	68.0	85.2	46.6	68.4	83.2	69.1	86.8	66.3	83.9	58.6	57.4	41.4	80.0	44.4	78.9
		Cosplay	71.0	<b>81.1</b>	<b>87.2</b>	72.6	<b>81.9</b>	87.0	<b>80.0</b>	<b>90.3</b>	<b>81.7</b>	<b>88.8</b>	<b>78.1</b>	<b>77.8</b>	<b>79.6</b>	<b>89.1</b>	<b>79.7</b>	<b>90.4</b>
		Dance	43.1	67.3	77.2	49.2	70.1	80.1	57.3	81.5	61.5	83.8	51.4	62.4	53.6	76.5	61.2	82.2
Drama	45.3	75.1	82.0	46.7	75.8	83.1	54.2	83.9	56.9	84.8	69.6	72.2	75.0	85.9	76.0	86.1		
Movie	49.5	71.5	77.2	50.4	72.2	76.2	57.6	76.8	56.5	78.6	64.9	65.8	69.2	82.2	68.2	80.4		
Average		22.2	55.2	76.4	24.1	55.4	76.0	28.7	67.7	30.7	67.5	34.6	36.3	37.5	72.3	39.2	72.7	

<sup>‡</sup> the top-down pose estimation results that use ground truth bounding box;

<sup>\*</sup> the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Pose Estimation

Detector		Faster R-CNN + HRNet					YOLOX + ViTPose				HigherHRNet		ED-Pose					
Setting		val	val <sup>‡</sup>	val <sup>*‡</sup>	test	test <sup>‡</sup>	test <sup>*‡</sup>	val	val <sup>‡</sup>	test	test <sup>‡</sup>	val	test	val	val <sup>*</sup>	test	test <sup>*</sup>	
MSCOCO [32]		<b>75.6</b>	77.6	77.2	<b>73.4</b>	-	-	79.8	82.3	81.1	-	68.6	70.5	71.6	72.4	69.8	-	
Artificial Scene	2D Representation	Cartoon	9.7	37.6	64.7	7.3	34.4	61.0	16.3	55.1	13.0	50.5	15.7	12.0	22.2	60.4	18.0	57.1
		Digital Art	22.6	59.6	74.4	25.8	61.2	75.7	29.0	69.9	31.9	72.2	42.5	44.3	43.5	71.4	45.6	75.1
		Ink Painting	6.3	51.4	72.1	5.6	48.0	72.4	8.9	59.8	9.2	58.2	26.8	20.9	28.2	56.8	24.9	55.4
		Kids Drawing	10.5	40.8	86.1	10.0	44.6	85.9	14.0	59.2	13.2	62.6	12.6	13.8	20.7	76.7	23.2	78.8
		Mural	11.6	54.0	71.1	12.3	53.6	71.6	15.9	50.1	16.4	51.6	30.6	32.0	34.6	64.7	35.1	65.4
		Oil Painting	31.6	65.7	78.1	28.5	62.2	75.6	39.6	73.4	36.7	70.5	54.4	51.1	56.2	75.2	51.7	71.4
		Shadow Play	<u>5.4</u>	<u>15.9</u>	<u>59.8</u>	<u>5.0</u>	<u>17.2</u>	<u>58.4</u>	<u>8.1</u>	<u>29.2</u>	<u>8.4</u>	<u>26.0</u>	<u>4.4</u>	<u>6.5</u>	<u>6.0</u>	<u>38.5</u>	<u>7.7</u>	<u>39.7</u>
		Sketch	6.3	44.1	73.1	6.7	57.2	79.4	9.1	61.3	10.9	71.3	13.6	<u>6.3</u>	12.0	66.8	12.2	75.8
		Stained Glass	10.4	46.0	74.8	9.7	45.1	73.1	12.0	59.1	12.1	58.1	26.6	23.1	27.6	74.4	25.6	71.5
		Ukiyoe	17.8	48.1	82.4	18.8	47.7	81.8	23.8	61.2	26.8	63.1	20.2	19.4	25.0	83.6	25.8	83.6
Watercolor	26.7	60.1	73.9	25.5	57.6	73.4	36.4	71.0	36.1	69.0	48.9	43.4	50.6	73.5	45.6	71.3		
Natural Scene	3D Representation	Garage Kits	45.2	57.5	86.7	44.5	61.4	<b>89.2</b>	52.5	76.2	50.6	77.0	37.4	34.7	47.9	87.7	44.1	90.1
		Relief	10.5	57.3	78.7	7.9	53.4	76.0	16.2	70.8	14.9	67.1	32.5	29.8	28.0	70.6	27.1	67.6
		Sculpture	36.4	65.9	81.0	38.5	64.0	78.5	34.9	78.5	34.2	73.7	33.5	35.2	45.9	76.9	46.7	74.7
		Acrobatics	45.8	68.0	85.2	46.6	68.4	83.2	69.1	86.8	66.3	83.9	58.6	57.4	41.4	80.0	44.4	78.9
		Cosplay	71.0	<b>81.1</b>	<b>87.2</b>	72.6	<b>81.9</b>	87.0	<b>80.0</b>	<b>90.3</b>	<b>81.7</b>	<b>88.8</b>	<b>78.1</b>	<b>77.8</b>	<b>79.6</b>	<b>89.1</b>	<b>79.7</b>	<b>90.4</b>
		Dance	43.1	67.3	77.2	49.2	70.1	80.1	57.3	81.5	61.5	83.8	51.4	62.4	53.6	76.5	61.2	82.2
		Drama	45.3	75.1	82.0	46.7	75.8	83.1	54.2	83.9	56.9	84.8	69.6	72.2	75.0	85.9	76.0	86.1
Movie	49.5	71.5	77.2	50.4	72.2	76.2	57.6	76.8	56.5	78.6	64.9	65.8	69.2	82.2	68.2	80.4		
Average		22.2	55.2	76.4	24.1	55.4	76.0	28.7	67.7	30.7	67.5	34.6	36.3	37.5	72.3	39.2	72.7	

<sup>‡</sup> the top-down pose estimation results that use ground truth bounding box;

<sup>\*</sup> the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Pose Estimation

Detector		Faster R-CNN + HRNet					YOLOX + ViTPose				HigherHRNet		ED-Pose					
Setting		val	val <sup>‡</sup>	val <sup>*‡</sup>	test	test <sup>‡</sup>	test <sup>*‡</sup>	val	val <sup>‡</sup>	test	test <sup>‡</sup>	val	test	val	val <sup>*</sup>	test	test <sup>*</sup>	
MSCOCO [32]		<b>75.6</b>	77.6	77.2	<b>73.4</b>	-	-	79.8	82.3	81.1	-	68.6	70.5	71.6	72.4	69.8	-	
Artificial Scene	2D Representation	Cartoon	9.7	37.6	64.7	7.3	34.4	61.0	16.3	55.1	13.0	50.5	15.7	12.0	22.2	60.4	18.0	57.1
		Digital Art	22.6	59.6	74.4	25.8	61.2	75.7	29.0	69.9	31.9	72.2	42.5	44.3	43.5	71.4	45.6	75.1
		Ink Painting	6.3	51.4	72.1	5.6	48.0	72.4	8.9	59.8	9.2	58.2	26.8	20.9	28.2	56.8	24.9	55.4
		Kids Drawing	10.5	40.8	86.1	10.0	44.6	85.9	14.0	59.2	13.2	62.6	12.6	13.8	20.7	76.7	23.2	78.8
		Mural	11.6	54.0	71.1	12.3	53.6	71.6	15.9	50.1	16.4	51.6	30.6	32.0	34.6	64.7	35.1	65.4
		Oil Painting	31.6	65.7	78.1	28.5	62.2	75.6	39.6	73.4	36.7	70.5	54.4	51.1	56.2	75.2	51.7	71.4
		Shadow Play	<u>5.4</u>	<u>15.9</u>	<u>59.8</u>	<u>5.0</u>	<u>17.2</u>	<u>58.4</u>	<u>8.1</u>	<u>29.2</u>	<u>8.4</u>	<u>26.0</u>	<u>4.4</u>	<u>6.5</u>	<u>6.0</u>	<u>38.5</u>	<u>7.7</u>	<u>39.7</u>
		Sketch	6.3	44.1	73.1	6.7	57.2	79.4	9.1	61.3	10.9	71.3	13.6	<u>6.3</u>	12.0	66.8	12.2	75.8
		Stained Glass	10.4	46.0	74.8	9.7	45.1	73.1	12.0	59.1	12.1	58.1	26.6	23.1	27.6	74.4	25.6	71.5
		Ukiyoe	17.8	48.1	82.4	18.8	47.7	81.8	23.8	61.2	26.8	63.1	20.2	19.4	25.0	83.6	25.8	83.6
Watercolor	26.7	60.1	73.9	25.5	57.6	73.4	36.4	71.0	36.1	69.0	48.9	43.4	50.6	73.5	45.6	71.3		
Natural Scene	3D Representation	Garage Kits	45.2	57.5	86.7	44.5	61.4	<b>89.2</b>	52.5	76.2	50.6	77.0	37.4	34.7	47.9	87.7	44.1	90.1
		Relief	10.5	57.3	78.7	7.9	53.4	76.0	16.2	70.8	14.9	67.1	32.5	29.8	28.0	70.6	27.1	67.6
		Sculpture	36.4	65.9	81.0	38.5	64.0	78.5	34.9	78.5	34.2	73.7	33.5	35.2	45.9	76.9	46.7	74.7
		Acrobatics	45.8	68.0	85.2	46.6	68.4	83.2	69.1	86.8	66.3	83.9	58.6	57.4	41.4	80.0	44.4	78.9
		Cosplay	71.0	<b>81.1</b>	<b>87.2</b>	72.6	<b>81.9</b>	87.0	<b>80.0</b>	<b>90.3</b>	<b>81.7</b>	<b>88.8</b>	<b>78.1</b>	<b>77.8</b>	<b>79.6</b>	<b>89.1</b>	<b>79.7</b>	<b>90.4</b>
		Dance	43.1	67.3	77.2	49.2	70.1	80.1	57.3	81.5	61.5	83.8	51.4	62.4	53.6	76.5	61.2	82.2
		Drama	45.3	75.1	82.0	46.7	75.8	83.1	54.2	83.9	56.9	84.8	69.6	72.2	75.0	85.9	76.0	86.1
Movie	49.5	71.5	77.2	50.4	72.2	76.2	57.6	76.8	56.5	78.6	64.9	65.8	69.2	82.2	68.2	80.4		
Average		22.2	<b>55.2</b>	<b>76.4</b>	24.1	<b>55.4</b>	<b>76.0</b>	28.7	67.7	30.7	67.5	34.6	36.3	37.5	72.3	39.2	72.7	

<sup>‡</sup> the top-down pose estimation results that use ground truth bounding box;

<sup>\*</sup> the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.



## Human-Centric Recognition

## Human Pose Estimation

Detector		Faster R-CNN + HRNet					YOLOX + ViTPose				HigherHRNet		ED-Pose					
Setting		val	val <sup>‡</sup>	val <sup>*‡</sup>	test	test <sup>‡</sup>	test <sup>*‡</sup>	val	val <sup>‡</sup>	test	test <sup>‡</sup>	val	test	val	val <sup>*</sup>	test	test <sup>*</sup>	
MSCOCO [32]		<b>75.6</b>	77.6	77.2	<b>73.4</b>	-	-	79.8	82.3	81.1	-	68.6	70.5	<b>71.6</b>	<b>72.4</b>	69.8	-	
Artificial Scene	2D Representation	Cartoon	9.7	37.6	64.7	7.3	34.4	61.0	16.3	55.1	13.0	50.5	15.7	12.0	22.2	60.4	18.0	57.1
		Digital Art	22.6	59.6	74.4	25.8	61.2	75.7	29.0	69.9	31.9	72.2	42.5	44.3	43.5	71.4	45.6	75.1
		Ink Painting	6.3	51.4	72.1	5.6	48.0	72.4	8.9	59.8	9.2	58.2	26.8	20.9	28.2	56.8	24.9	55.4
		Kids Drawing	10.5	40.8	86.1	10.0	44.6	85.9	14.0	59.2	13.2	62.6	12.6	13.8	20.7	76.7	23.2	78.8
		Mural	11.6	54.0	71.1	12.3	53.6	71.6	15.9	50.1	16.4	51.6	30.6	32.0	34.6	64.7	35.1	65.4
		Oil Painting	31.6	65.7	78.1	28.5	62.2	75.6	39.6	73.4	36.7	70.5	54.4	51.1	56.2	75.2	51.7	71.4
		Shadow Play	<u>5.4</u>	<u>15.9</u>	<u>59.8</u>	<u>5.0</u>	<u>17.2</u>	<u>58.4</u>	<u>8.1</u>	<u>29.2</u>	<u>8.4</u>	<u>26.0</u>	<u>4.4</u>	<u>6.5</u>	<u>6.0</u>	<u>38.5</u>	<u>7.7</u>	<u>39.7</u>
		Sketch	6.3	44.1	73.1	6.7	57.2	79.4	9.1	61.3	10.9	71.3	13.6	<u>6.3</u>	12.0	66.8	12.2	75.8
		Stained Glass	10.4	46.0	74.8	9.7	45.1	73.1	12.0	59.1	12.1	58.1	26.6	23.1	27.6	74.4	25.6	71.5
		Ukiyoe	17.8	48.1	82.4	18.8	47.7	81.8	23.8	61.2	26.8	63.1	20.2	19.4	25.0	83.6	25.8	83.6
Watercolor	26.7	60.1	73.9	25.5	57.6	73.4	36.4	71.0	36.1	69.0	48.9	43.4	50.6	73.5	45.6	71.3		
Natural Scene	3D Representation	Garage Kits	45.2	57.5	86.7	44.5	61.4	<b>89.2</b>	52.5	76.2	50.6	77.0	37.4	34.7	47.9	87.7	44.1	90.1
		Relief	10.5	57.3	78.7	7.9	53.4	76.0	16.2	70.8	14.9	67.1	32.5	29.8	28.0	70.6	27.1	67.6
		Sculpture	36.4	65.9	81.0	38.5	64.0	78.5	34.9	78.5	34.2	73.7	33.5	35.2	45.9	76.9	46.7	74.7
		Acrobatics	45.8	68.0	85.2	46.6	68.4	83.2	69.1	86.8	66.3	83.9	58.6	57.4	41.4	80.0	44.4	78.9
		Cosplay	71.0	<b>81.1</b>	<b>87.2</b>	72.6	<b>81.9</b>	87.0	<b>80.0</b>	<b>90.3</b>	<b>81.7</b>	<b>88.8</b>	<b>78.1</b>	<b>77.8</b>	<b>79.6</b>	<b>89.1</b>	<b>79.7</b>	<b>90.4</b>
		Dance	43.1	67.3	77.2	49.2	70.1	80.1	57.3	81.5	61.5	83.8	51.4	62.4	53.6	76.5	61.2	82.2
		Drama	45.3	75.1	82.0	46.7	75.8	83.1	54.2	83.9	56.9	84.8	69.6	72.2	75.0	85.9	76.0	86.1
Movie	49.5	71.5	77.2	50.4	72.2	76.2	57.6	76.8	56.5	78.6	64.9	65.8	69.2	82.2	68.2	80.4		
Average		22.2	55.2	76.4	24.1	55.4	76.0	28.7	67.7	30.7	67.5	34.6	36.3	37.5	72.3	39.2	72.7	

<sup>‡</sup> the top-down pose estimation results that use ground truth bounding box;

<sup>\*</sup> the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

## Human Pose Estimation

Detector		Faster R-CNN + HRNet					YOLOX + ViTPose				HigherHRNet		ED-Pose					
Setting		val	val <sup>‡</sup>	val <sup>*‡</sup>	test	test <sup>‡</sup>	test <sup>*‡</sup>	val	val <sup>‡</sup>	test	test <sup>‡</sup>	val	test	val	val <sup>*</sup>	test	test <sup>*</sup>	
MSCOCO [32]		<b>75.6</b>	77.6	77.2	<b>73.4</b>	-	-	79.8	82.3	81.1	-	68.6	70.5	71.6	72.4	69.8	-	
Artificial Scene	2D Representation	Cartoon	9.7	37.6	64.7	7.3	34.4	61.0	16.3	55.1	13.0	50.5	15.7	12.0	22.2	60.4	18.0	57.1
		Digital Art	22.6	59.6	74.4	25.8	61.2	75.7	29.0	69.9	31.9	72.2	42.5	44.3	43.5	71.4	45.6	75.1
		Ink Painting	6.3	51.4	72.1	5.6	48.0	72.4	8.9	59.8	9.2	58.2	26.8	20.9	28.2	56.8	24.9	55.4
		Kids Drawing	10.5	40.8	86.1	10.0	44.6	85.9	14.0	59.2	13.2	62.6	12.6	13.8	20.7	76.7	23.2	78.8
		Mural	11.6	54.0	71.1	12.3	53.6	71.6	15.9	50.1	16.4	51.6	30.6	32.0	34.6	64.7	35.1	65.4
		Oil Painting	31.6	65.7	78.1	28.5	62.2	75.6	39.6	73.4	36.7	70.5	54.4	51.1	56.2	75.2	51.7	71.4
		Shadow Play	<u>5.4</u>	<u>15.9</u>	<u>59.8</u>	<u>5.0</u>	<u>17.2</u>	<u>58.4</u>	<u>8.1</u>	<u>29.2</u>	<u>8.4</u>	<u>26.0</u>	<u>4.4</u>	<u>6.5</u>	<u>6.0</u>	<u>38.5</u>	<u>7.7</u>	<u>39.7</u>
		Sketch	6.3	44.1	73.1	6.7	57.2	79.4	9.1	61.3	10.9	71.3	13.6	<u>6.3</u>	12.0	66.8	12.2	75.8
		Stained Glass	10.4	46.0	74.8	9.7	45.1	73.1	12.0	59.1	12.1	58.1	26.6	23.1	27.6	74.4	25.6	71.5
		Ukiyoe	17.8	48.1	82.4	18.8	47.7	81.8	23.8	61.2	26.8	63.1	20.2	19.4	25.0	83.6	25.8	83.6
Watercolor	26.7	60.1	73.9	25.5	57.6	73.4	36.4	71.0	36.1	69.0	48.9	43.4	50.6	73.5	45.6	71.3		
Natural Scene	3D Representation	Garage Kits	45.2	57.5	86.7	44.5	61.4	<b>89.2</b>	52.5	76.2	50.6	77.0	37.4	34.7	47.9	87.7	44.1	90.1
		Relief	10.5	57.3	78.7	7.9	53.4	76.0	16.2	70.8	14.9	67.1	32.5	29.8	28.0	70.6	27.1	67.6
		Sculpture	36.4	65.9	81.0	38.5	64.0	78.5	34.9	78.5	34.2	73.7	33.5	35.2	45.9	76.9	46.7	74.7
		Acrobatics	45.8	68.0	85.2	46.6	68.4	83.2	69.1	86.8	66.3	83.9	58.6	57.4	41.4	80.0	44.4	78.9
		Cosplay	71.0	<b>81.1</b>	<b>87.2</b>	72.6	<b>81.9</b>	87.0	<b>80.0</b>	<b>90.3</b>	<b>81.7</b>	<b>88.8</b>	<b>78.1</b>	<b>77.8</b>	<b>79.6</b>	<b>89.1</b>	<b>79.7</b>	<b>90.4</b>
		Dance	43.1	67.3	77.2	49.2	70.1	80.1	57.3	81.5	61.5	83.8	51.4	62.4	53.6	76.5	61.2	82.2
Drama	45.3	75.1	82.0	46.7	75.8	83.1	54.2	83.9	56.9	84.8	69.6	72.2	75.0	85.9	76.0	86.1		
Movie	49.5	71.5	77.2	50.4	72.2	76.2	57.6	76.8	56.5	78.6	64.9	65.8	69.2	82.2	68.2	80.4		
Average		22.2	55.2	76.4	24.1	55.4	76.0	28.7	67.7	30.7	67.5	34.6	36.3	37.5	72.3	39.2	72.7	

<sup>‡</sup> the top-down pose estimation results that use ground truth bounding box;

<sup>\*</sup> the baseline results we provide by training on the joint of MSCOCO [32] and *Human-Art*.

## Human-Centric Recognition

### Human Mesh Recovery



Oil Painting



(a)



Watercolor



(c)



Digital Art



(e)

**Illustration of how the annotated self-contact points can benefit 3D human mesh recovery.**

(a), (c), (e): **without** self-contact optimization

**Depth Ambiguities**

## Human-Centric Recognition

### Human Mesh Recovery



Oil Painting



(a)



(b)



Watercolor



(c)



(d)



Digital Art



(e)



(f)

**Illustration of how the annotated self-contact points can benefit 3D human mesh recovery.**

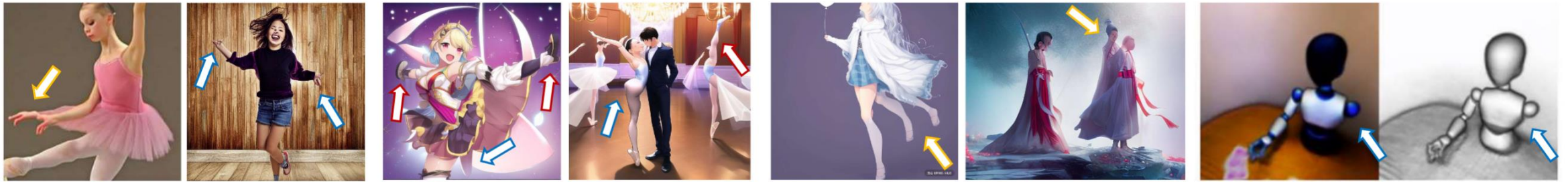
(a), (c), (e): **without** self-contact optimization

(b), (d), (f): **with** self-contact points.

## Human-Centric Recognition

### Controllable Image Generation

#### Failure Case



(a) Stable Diffusion v1.5

(b) NovelAI

(c) ERNIE-ViLG

(d) DreamFusion3D

#### Text-to-Image



Kids Drawing

Garage Kits

Digital Art

Oil Painting

Shadow Play

#### Text&Pose-to-Image



*Shadow play, two people  
dressed orientally dancing*

*Kids drawing, a family  
with children and flowers*

## Human-Art



Rich scenario

High quality

Versatile annotations

## Future Work

- cross-domain human recognition algorithms
  - trustworthy image generation
- Inclusive motion transfer algorithms across different scenes