Learning Diverse Human Representation in the Wild

Ziwei Liu

The Chinese University of Hong Kong
Human-Centric AI
Human-Centric AI

Face Representations

Human Representations
Diverse Modalities
Visual-Audio Representation

Diverse Poses & Textures
Colorful 3D Human Representation

Diverse Categories & Relations
Fashion Collocation Representation
Diverse Modalities

Vision-Infused Deep Audio Inpainting,
ICCV 2019
Motivation

• Audio signals often suffer from local distortions where the intervals are corrupted.

• Audio Inpainting: To fill the corrupted information with newly generated samples.
Core Idea

• Formulate audio inpainting into spectrogram inpainting.
Core Idea

- Utilize intact video to guide audio inpainting.
Approach

• Overview: Vision-Infused Audio Inpainter (VIAI)
VIAI–Audio Branch (VIAI-A)

• Using the 2D Time-Frequency representation of Mel-Spectrogram for audios.

• Formulating the problem into inpainting spectrogram with Generative Adversarial Networks
VIAI–Audio-Visual Branch (VIAI-AV)

• Learning synchronization between intact video and audio.

• Concatenate the synchronized features for reconstruction.
VIAI–Audio-Visual Branch (VIAI-AV)

- Probe loss of using intact audio for reconstruction (VIAI-AA').

- Forcing the network to learn from bottleneck features.
WaveNet Decoder

• WaveNet is used to convert Mel-spectrogram back to raw audio.

• Utilizing the given audio for better restoration.
# Experiments

<table>
<thead>
<tr>
<th>Score \ Approach</th>
<th>SampleRNN [33]</th>
<th>Visual2Sound [56]</th>
<th>bi-SampleRNN</th>
<th>bi-Visual2Sound</th>
<th>VIAI-A</th>
<th>VIAI-AV</th>
<th>VIAI-AA'</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>9.1</td>
<td>10.2</td>
<td>12.8</td>
<td>13.6</td>
<td>22.2</td>
<td>23.2</td>
<td>26.6</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.33</td>
<td>0.35</td>
<td>0.38</td>
<td>0.41</td>
<td>0.61</td>
<td>0.64</td>
<td>0.75</td>
</tr>
<tr>
<td>SDR</td>
<td>4.89</td>
<td>3.70</td>
<td>4.20</td>
<td>4.72</td>
<td>6.54</td>
<td>6.63</td>
<td>6.89</td>
</tr>
<tr>
<td>OPS</td>
<td>51.1</td>
<td>51.3</td>
<td>51.2</td>
<td>52.2</td>
<td>52.4</td>
<td>56.3</td>
<td>56.7</td>
</tr>
</tbody>
</table>
Vision-Infused Deep Audio Inpainting

Hang Zhou¹ Ziwei Liu¹ Xudong Xu¹ Ping Luo² Xiaogang Wang¹

1. The Chinese University of Hong Kong
2. The University of Hong Kong
Conclusions

• Discriminative representation’s is capable of distilling and disentangling information from both modalities.

• Audio problems can be easier solved by operating on spectrograms using vision techniques for image processing.

• Synchronization between audio and visual information is the fundamental self-supervision which is crucial for various tasks.
Diverse Poses

Delving Deep into Hybrid Annotations for 3D Human Recovery in the wild, ICCV 2019
3D Human Reconstruction means acquire 3D human representation from given images or videos.

It can facilitate many technologies such as augmented reality and virtual try-on.
We use SMPL, a parametric triangle mesh based 3D model to represent 3D human.

SMPL is parameterized by two parameters: **pose parameters** $\theta \in \mathbb{R}^{72}$ and **shape parameters** $\beta \in \mathbb{R}^{10}$.

To estimate 3D human representation, we only need to predict the pose and shape parameters.
In the experiment, we first study the efficiency of different annotations. We study the efficiency of those annotations when serving as input and serving as supervision. We use per-vertex distance (PVE) as the evaluation metric. The experiments are conducted on COCO-DensePose, UP-3D and 3DPW.

\[
PVE = \sum_{i=1}^{O} \|P_i - \tilde{P}_i\|_2
\]
The overall framework is composed of three parts:
- Input Encoder
- Parameter Estimator
- Loss Calculator
Learning Strategy (I)

- Previous works mainly use 3D annotations and sparse 2D annotations in training.
- Sparse 2D keypoints are too sparse to provide enough guidance.
- 3D annotations are hard to acquire.
- We propose to use dense keypoints in recovering 3D human model.

\[
L_{3D,\text{joints}} = \sum_{i=1}^{M} ||(J_i^{\text{3D}} - \hat{J}_i^{\text{3D}})||_1,
\]

\[
L_{\text{SMPL}} = \sum_{i=1}^{O} ||R(\theta_i) - R(\hat{\theta}_i)||_1 + ||\beta_i - \hat{\beta}_i||_1
\]

\[
L_{3D} = L_{3D,\text{joints}} + L_{\text{SMPL}}
\]

\[
L_{2D} = \sum_{i=1}^{S} ||(J_i^{2D} - \hat{J}_i^{2D})||_1
\]

\[
[v_{i1}, v_{i2}, v_{i3}], [b_{i1}, b_{i2}, b_{i3}] = \phi(D_i),
\]

\[
\hat{X}_i = \sum_{j=1}^{3} \hat{p}_i^{2D} [v_{ij}] \times b_{ij},
\]

\[
L_{\text{dense}} = \sum_{i=1}^{T} ||(X_i - \hat{X}_i)||_1
\]
DensePose build dense correspondence between 2D images and human body surface.

For each dense keypoints, the annotations include \((I, U, V)\). \(I\) indicates which body part this point belongs to. \((U, V)\) indicates the precise position.

Dense keypoints could be annotated by human annotators without using auxiliary equipements.
We use the predicted IUV maps from DensePose model and sample dense keypoints from them.

We conduct refinement using the accurate sparse 2D keypoints to remove erroneous IUV maps.
Table 3. **Influence of different annotations.** The evaluation metrics are PVE, MPJPE and PVE-T, separately. For all metrics, lower is better. “3D” refers to paired in-the-wild 3D annotations. “20% 3D” refers to 20% randomly selected 3D annotations. “Sparse 2D” refers to sparse 2D keypoints. “Dense” refers to dense correspondence, namely, IUV maps generated by DensePose [1, 19].

<table>
<thead>
<tr>
<th>Supervision Input</th>
<th>3D &amp; Dense &amp; Sparse 2D</th>
<th>20% 3D &amp; Dense &amp; Sparse 2D</th>
<th>3D &amp; Sparse 2D</th>
<th>Dense &amp; Sparse 2D</th>
<th>Sparse 2D Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>IUV Only</td>
<td><strong>120.0 / 103.1 / 31.8</strong></td>
<td>125.0 / 107.2 / 32.6</td>
<td>125.2 / 106.4 / 32.1</td>
<td>138.7 / 121.2 / 54.7</td>
<td>204.3 / 177.0 / 92.1</td>
</tr>
<tr>
<td>Segment Only</td>
<td>123.0 / 105.1 / 32.7</td>
<td>126.7 / 110.0 / 33.2</td>
<td>124.8 / 107.8 / 31.7</td>
<td>147.4 / 130.1 / 55.9</td>
<td>203.8 / 176.7 / 93.3</td>
</tr>
<tr>
<td>Image Only</td>
<td>123.7 / 105.9 / 30.9</td>
<td>127.5 / 110.6 / 32.2</td>
<td>127.4 / 108.5 / 30.7</td>
<td>137.7 / 120.3 / 51.7</td>
<td>203.2 / 178.5 / 106.2</td>
</tr>
<tr>
<td>Image &amp; IUV</td>
<td><strong>122.4 / 105.1 / 30.2</strong></td>
<td>125.0 / 107.6 / 32.1</td>
<td>125.5 / 107.3 / 30.7</td>
<td>133.8 / 117.2 / 52.5</td>
<td>197.3 / 172.8 / 107.9</td>
</tr>
<tr>
<td>Image &amp; Segment</td>
<td>121.5 / 104.3 / 31.0</td>
<td>126.4 / 107.0 / 31.6</td>
<td>125.8 / 106.8 / 31.5</td>
<td>142.2 / 124.2 / 56.6</td>
<td>201.2 / 177.5 / 101.7</td>
</tr>
</tbody>
</table>
Delving Deep into Hybrid Annotations for 3D Human Recovery

Paper ID 2209

This video is composed of two parts:
I. Influence of different annotations
II. Comparison with previous state-of-the-arts.
Diverse Textures

Learning to Synthesis Fashion Textures,
(in submission)
Fashion Texture Synthesis

• Use Gram matrix as texture feature to synthesize images
  • Flexible
  • Visually pleasing
Fashion Texture Synthesis

• Two-step generation
Generative Framework

• Training Gram-WAE-GAN
  • Reconstruct the input Gram matrix
  • Match the latent distribution with the prior

Recursive Structure

• Model a set of Gram matrices from multi-granularity levels
GMM Sampling

Training GMM

Sampling

Gram Transformation

- Transform the Gram matrix to a low dimensional vector
  - Number of parameters: 184M -> 10.8M
Results

<table>
<thead>
<tr>
<th>Method</th>
<th>FID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>41.97</td>
</tr>
<tr>
<td>DistGAN [87]</td>
<td>44.38</td>
</tr>
<tr>
<td>PSGAN [5]</td>
<td>77.10</td>
</tr>
<tr>
<td>TextureGAN [93]</td>
<td></td>
</tr>
<tr>
<td>Ablation Study</td>
<td></td>
</tr>
<tr>
<td>FC transformation</td>
<td>37.52</td>
</tr>
<tr>
<td>MLP structure</td>
<td>45.72</td>
</tr>
<tr>
<td>No GMM sampling</td>
<td>40.83</td>
</tr>
<tr>
<td>Ours</td>
<td>37.74</td>
</tr>
</tbody>
</table>

The table shows the FID (Fréchet Inception Distance) scores for different methods. The images below illustrate the results of each method.
Results
Diverse Categories

Large-Scale Long-Tailed Recognition in an Open World,
CVPR 2019
Train

<table>
<thead>
<tr>
<th>Cat</th>
<th>Fox</th>
<th>Panda</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Cat" /></td>
<td><img src="image2" alt="Fox" /></td>
<td><img src="image3" alt="Panda" /></td>
</tr>
<tr>
<td><img src="image4" alt="Cat" /></td>
<td><img src="image5" alt="Fox" /></td>
<td><img src="image6" alt="Panda" /></td>
</tr>
<tr>
<td><img src="image7" alt="Cat" /></td>
<td><img src="image8" alt="Fox" /></td>
<td><img src="image9" alt="Panda" /></td>
</tr>
</tbody>
</table>

Test

<table>
<thead>
<tr>
<th>Cat</th>
<th>Fox</th>
<th>Panda</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image10" alt="Cat" /></td>
<td><img src="image11" alt="Fox" /></td>
<td><img src="image12" alt="Panda" /></td>
</tr>
<tr>
<td><img src="image13" alt="Cat" /></td>
<td><img src="image14" alt="Fox" /></td>
<td><img src="image15" alt="Panda" /></td>
</tr>
<tr>
<td><img src="image16" alt="Cat" /></td>
<td><img src="image17" alt="Fox" /></td>
<td><img src="image18" alt="Panda" /></td>
</tr>
</tbody>
</table>

Cat (many-shot class)

Fox (medium-shot class)

Panda (few-shot class)

? (open class)
Faces [Zhang et al. 2017]

Places [Wang et al. 2017]

Species [Van Horn et al. 2019]

Actions [Zhang et al. 2019]
Open Long-Tailed Recognition

Head Classes

Tail Classes

Open Classes

Open World
Open Long-Tailed Recognition

Head Classes

Tail Classes

Open Classes

Knowledge Transfer

Open World
Open Long-Tailed Recognition

Head Classes  Tail Classes  Open Classes

Knowledge Transfer

Open World

Sensitivity to Novelty
Open Long-Tailed Recognition

- Avoid Forgetting
- Knowledge Transfer

Head Classes | Tail Classes | Open Classes

Open World
Sensitivity to Novelty
Open Long-Tailed Recognition

Imbalanced Classification

Few-shot Learning

Head Classes

Tail Classes

Open Set Recognition

Open World

Open Classes
Imbalanced Classification
(metric learning, re-sampling, re-weighting)

Few-Shot Learning
(meta learning, classifier dynamics)

Open Set Recognition
(distribution rectification, out-of-distribution detection)

Open Long-Tailed Recognition
(dynamic meta-embedding)
Open Long-Tailed Recognition
(dynamic meta-embedding)

Knowledge Transfer
Avoid Forgetting
Sensitivity to Novelty

train

test
visual memory

FLY

top-down attention

bottom-up attention

direct embedding

enhanced embedding

familiarity

FLY
Head Classes

Tail Classes

Open Classes

Avoid Forgetting

Knowledge Transfer

Sensitivity to Novelty

bottom-up attention

visual memory

top-down attention

familiarity
Head Classes

Tail Classes

Open Classes

bottom-up attention

top-down attention

visual memory

direct embedding

enhanced embedding

associative memory

feature selection

familiarity
ImageNet-LT Benchmark
Absolute Performance Gain: ~20%

Places-LT Benchmark
Absolute Performance Gain: ~10%

MS1M-LT Benchmark
Absolute Performance Gain: ~2%
**Overall F1 Score** on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
**Overall F1 Score** on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
### Overall F1 Score on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
**Overall F1 Score** on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
### Overall F1 Score on ImageNet-LT, Places-LT and MS1M-LT Benchmarks

<table>
<thead>
<tr>
<th>Methods</th>
<th>ImageNet-LT</th>
<th>Places-LT</th>
<th>MS1M-LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Model</td>
<td>0.295</td>
<td>0.366</td>
<td>0.738</td>
</tr>
<tr>
<td>Sample Re-weighting (Focal Loss)</td>
<td>0.371</td>
<td>0.453</td>
<td>-</td>
</tr>
<tr>
<td>Metric Learning (Range Loss)</td>
<td>0.373</td>
<td>0.457</td>
<td>0.722</td>
</tr>
<tr>
<td>Open Set Recognition (OpenMax)</td>
<td>0.368</td>
<td>0.458</td>
<td>-</td>
</tr>
<tr>
<td>Few-shot Learning (FSLwF)</td>
<td>0.347</td>
<td>0.375</td>
<td>-</td>
</tr>
<tr>
<td><strong>Dynamic Meta-Embedding</strong></td>
<td><strong>0.474</strong></td>
<td><strong>0.464</strong></td>
<td><strong>0.745</strong></td>
</tr>
</tbody>
</table>
Diverse Relations

Learning Diverse Fashion Collocation by Neural Graph Filtering,
(in submission)
Motivation

• Increasing demand for intelligent fashion recommendation system

• A successful fashion collocation framework should be featured with two desired properties: **Flexibility** and **Diversity**.

• Existing work can only accept fashion sets with **fixed length**, e.g., the four-garment set{tops, outerwear, bottoms and shoes} and **limited categories**, e.g., discarding accessories, bags and hats.
Overall Framework of Diverse Fashion Graph Filtering

We firstly use the convolutional neural networks to extract the visual embeddings of the input garment sets with flexible length, and then consider each visual embedding as a node input to the neural graph network, which not only computes the node features, but also implements edge feature aggregation. Note that one node could appear in several collocations. Afterwards a compatibility predictor calculates the compatibility scores for diverse styled garment sets.
• The graph network architecture constructed using **edge feature aggregation** operations.

• In the last layer, edge information gathered at all the nodes are pooled to compute a compatibility score, and an optional fashion style distribution for a compatible garment set.
Graph edge Filtering at **one layer**: aggregates all the edge information connecting to the node under consideration.
## Quantitative Evaluation

<table>
<thead>
<tr>
<th>dataset</th>
<th>Polyvore</th>
<th>Polyvore-D</th>
<th>H. (%)</th>
<th>Polyvore</th>
<th>Polyvore-D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>FITB</td>
<td>AUC</td>
<td>FITB</td>
<td>AUC</td>
</tr>
<tr>
<td>Bi-LSTM (Han et al. 2017)</td>
<td>0.65</td>
<td>39.7</td>
<td>0.62</td>
<td>39.4</td>
<td>5.0</td>
</tr>
<tr>
<td>CSN (Veit, Belongie, and Karaletsos 2017)</td>
<td>0.83</td>
<td>54.0</td>
<td>0.82</td>
<td>52.5</td>
<td>0</td>
</tr>
<tr>
<td>TransNFCM (Xun Yang 2019)</td>
<td>0.75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wardrobe (Wei-Lin Hsiao 2018)</td>
<td>0.88</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.5</td>
</tr>
<tr>
<td>Type Aware (Vasileva et al. 2018)</td>
<td>0.86</td>
<td>56.2</td>
<td>0.84</td>
<td>54.9</td>
<td>5.0</td>
</tr>
<tr>
<td>Neural Graph Filtering (Ours)</td>
<td>0.94</td>
<td>58.8</td>
<td>0.88</td>
<td>55.1</td>
<td>82.5</td>
</tr>
</tbody>
</table>

| Neural Graph Filtering (Ours) | 0.94     | 58.8       | 0.88   | 55.1     | 82.5       | Neural Graph Filtering (Ours) |
| Unnamed                       | 0.85     | 54.7       | 0.82   | 53.4     |            |                          |
| Unnamed                       | 0.85     | 55.1       | 0.83   | 54.2     |            |                          |
| Unnamed                       | 0.92     | 55.3       | 0.84   | 47.8     |            |                          |
| Unnamed                       | 0.93     | 57.7       | 0.87   | 52.8     |            |                          |
| Unnamed                       | 0.93     | 58.0       | 0.86   | 53.8     |            |                          |
| Unnamed                       | 0.94     | 58.8       | 0.88   | 55.1     |            |                          |
Fill-in-blank

given a sequence of fashion items, ask for the most compatible one from the four choices
Fashion Compatibility Prediction
score a candidate outfit, higher score means more compatibility

- Compatible: 0.994
- Not Compatible: 0.041
Diverse Fashion Collocations

Given 1 query item, generate fashion sets of **diverse** styles and **flexible** length

Dataset: Polyvore
Diverse Fashion Collocations
Given 1 query item, generate fashion sets of diverse styles and flexible length

query item

- Analogous
- Complementary
- Triadic
- Same
- Monochromatic
- Other
Diverse Fashion Collocations

Given 1 query item, generate fashion sets of **diverse** styles and **flexible** length
Diverse Fashion Collocations

Dataset: Amazon Fashion
Diverse Fashion Collocations

Dataset: Amazon Fashion

query item

Analogous  Complementary  Triadic  Same  Monochromatic  Other
Conclusions

• The concept of **flexible** and **diverse** fashion collocations:
  • support both inputs/outputs with flexible lengths;
  • generate fashion sets with diverse styles

• Novel framework of **neural graph filtering**
  • the graph structure that explores the inter-garment relationship is more suitable for fashion compatibility learning.

• Newly proposed benchmark and evaluation protocols
  • *AmazonFashion* Dataset: comprises of different styles for diversity learning and evaluation
Database and Toolbox
Two New Datasets:
- Fashion Parsing Benchmark
- Fashion Recommendation Benchmark
Open-source toolbox for visual fashion analysis based on PyTorch: https://github.com/open-mmlab/mmfashion

Features

- **Flexible**: modular design and easy to extend
- **Friendly**: off-the-shelf models for layman users
- **Comprehensive**: support a wide spectrum of fashion analysis tasks
  - ✔️ Fashion Attribute Prediction
  - ✔️ Fashion Recognition and Retrieval
  - ✔️ Fashion Landmark Detection
  - ☐ Fashion Parsing and Segmentation
  - ☐ Fashion Compatibility and Recommendation
Thanks!

Science is what we understand well enough to explain to a computer. Art is everything else we do.

Homepage: https://liuziwei7.github.io/