Data-Scarce Animal Face Alignment via Bi-Directional Cross-Species Knowledge Transfer

Dan Zeng¹, Shanchuan Hong¹, Shuiwang Li², Qiaomu Shen¹, Bo Tang¹

¹Southern University of Science and Technology,
²Guilin University of Technology
Animal face alignment

• Detect facial landmarks

• Large significance
  • Better understanding animals
  • Promote their health
  • Less expensive to examinations

• It remains largely **unexplored**!

- Facial expression analysis
- Animal pain detection
- Facial tracking
Challenges

• Large intra- and inter- species variations
• Lack large-scale annotated data!!

• Our focus: data-scarce animal face alignment

- 300W
- AFLW
- Menpo
- ...
- AnimalWeb

Challenges

• Large intra- and inter- species variations
• Lack large-scale annotated data!!

Existing solutions
  • Finetuning human face alignment model
  • Animal-specific face alignment (horse, sheep)
  • Utilize auxiliary information

Our solution (Meta-CSKT)
  • Leverage bi-directional cross-species knowledge transfer
Motivation
Motivation
Motivation

NME Difference Confusion Matrix
Motivation

- Pretrained human model vs finetuned animal model
- NME Difference
- Confusion Matrix

- Test data (Family #)
- Finetune data (Family #)

- Images of family members:
  - #0, #17, #26, #35, #65
Motivation
Motivation

Knowledge sharing among animal species, which provides a fundamental premise for Meta-CSKT!!
Motivation

Knowledge sharing among animal species, which provides a fundamental premise for Meta-CSKT!!
Labeled data is SCARCE!
Meta-CSKT

Labeled data is SCARCE!
Two complement networks
Meta-CSKT

Labeled data is SCARCE!

Two complement networks

Positive example mining:
• Augment labeled data
• Purify unlabeled data
Meta-CSKT

Adaptation network:
- $L_u = \left\| \hat{H}_u - A(I_u; \theta_A) \right\|^2$

Base network (1/2):
- $L_s = \left\| H_{fs} - B(I_{fs}; \theta_B) \right\|^2 + \left\| A(I_{pos}; \theta_A) - B(I_{pos}; \theta_B) \right\|^2$
Meta-CSKT

Base network (2/2):

Feedback loss

\[ \mathcal{L}_f = \| \hat{H}_u - B(I_u; \theta_B) \|^2 \]

\[
\begin{align*}
    f &= \eta_A \cdot (\nabla_{\theta_A^{(t+1)}} \text{MSE}(H_{fs}, A(I_{fs}; \theta_A^{(t+1)}))^T \cdot \\
    &\quad \nabla_{\theta_A} \text{MSE}(\hat{H}_u, A(I_u; \theta_A^{(t)})))
\end{align*}
\]
Meta-CSKT

Base network (2/2):

Feedback loss

\[ \mathcal{L}_f = f \cdot \| \hat{H}_u - B(I_u; \theta_B) \|^2 \]

Gradients of the “new” network on few-shot data

\[ f = \eta_A \cdot \left( \nabla \theta_A^{(i+1)} \right) \text{MSE}(H_{f}, A(I_{fs}; \theta_A^{(i+1)})) \cdot \left( \nabla \theta_A \text{MSE}(\hat{H}_u, A(I_u; \theta_A^{(i)})) \right) \]

Gradients of the “old” network on large-scale unlabeled data
Positive example mining

If the **ground truth** is known, it is easy to identify **three** types of unlabeled data
Positive example mining

No ground truth is available!

- If $\| H_{flip} - Flip(H_{orig}) \|^2 > T_{neg}$, the unlabeled data is **hard negative**
- If $\Delta H_{orig} < T_{pos}$, the unlabeled data is **positive**
- Elsewise, the unlabeled data is **semi-hard positives**
Evaluation on Horse Facial Keypoint dataset

Robust to occlusion and large pose variations
Evaluation on Horse Facial Keypoint dataset

- Meta-CSKT (Ours)
- WarpingNet
- TIF

![Graph showing failure rate vs. error threshold for different methods.]

<table>
<thead>
<tr>
<th>Landmarks</th>
<th>LE</th>
<th>RE</th>
<th>Nose</th>
<th>LM</th>
<th>RM</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-CSKT</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>WarpingNet</td>
<td>10</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>TIF</td>
<td>15</td>
<td>25</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

![Bar chart showing failure rate at 0.1NME for specific landmarks.]
Evaluation on Japanese Macaque Species

<table>
<thead>
<tr>
<th>Models</th>
<th>MDMD Base</th>
<th>MDMD 300W</th>
<th>ViTPose+B</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>NME</td>
<td>3.66</td>
<td>3.44</td>
<td>4.69</td>
<td><strong>2.96</strong></td>
</tr>
</tbody>
</table>
## Evaluation on AnimalWeb

<table>
<thead>
<tr>
<th>Models (Known)</th>
<th># Labeled Image</th>
<th>NME</th>
<th>Models (Unknown)</th>
<th># Labeled Image</th>
<th>NME</th>
</tr>
</thead>
<tbody>
<tr>
<td>HG2</td>
<td>17.96K</td>
<td><strong>5.22</strong></td>
<td>HG2</td>
<td>17.62K</td>
<td><strong>6.14</strong></td>
</tr>
<tr>
<td>HG3</td>
<td>17.96K</td>
<td>5.12</td>
<td>HG3</td>
<td>17.62K</td>
<td>5.96</td>
</tr>
<tr>
<td>Ours</td>
<td>40</td>
<td>5.61</td>
<td>Ours</td>
<td>40</td>
<td>7.44</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td><strong>5.55</strong></td>
<td></td>
<td>80</td>
<td><strong>7.21</strong></td>
</tr>
</tbody>
</table>

Our method achieves comparable performance while utilizing significantly fewer labeled data.

- Known species setting: comparable
- Unknown species setting: acceptable gap (1%)
- Images: 40 or 80 versus 17+K; Images per species: 0.11 or 0.22 versus ~50
## Effect of Meta-CSKT design

<table>
<thead>
<tr>
<th>Models</th>
<th>Meta-cskt Loss</th>
<th>Positive Example Mining</th>
<th>NME Known/Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L_s$</td>
<td>$L_u$</td>
<td>$L_f$</td>
</tr>
<tr>
<td>1</td>
<td>✔</td>
<td>✘</td>
<td>✘</td>
</tr>
<tr>
<td>2</td>
<td>✔</td>
<td>✔</td>
<td>✘</td>
</tr>
<tr>
<td>3</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>4</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ours</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

The table above shows the effect of different combinations of losses and positive example mining on the NME (normalized mean error) for known and unknown classes. The **Ours** model demonstrates the best performance with the lowest NME of 5.61/7.44.
Conclusion

• Knowledge sharing among animals motivates Meta-CSKT, the first to leverage bi-directional cross-species knowledge transfer for data-scarce animal face alignment

• We propose positive example mining to effectively utilize unlabeled data: augment labeled data and purify unlabeled data

• Extensive experiments on three datasets demonstrate the superiority of our method for animal face alignment by using only a few labeled images
Please refer to our paper and the project page for more details and analyses: https://www.danzeng.org/about/, https://github.com/danzeng1990/Meta-CSKT

Thanks for attention

Contact: zengd@sustech.edu.cn