Symbiotic Adversarial Learning for Attribute-based Person Search

Yu-Tong Cao*, Jingya Wang*, and Dacheng Tao

UBTECH Sydney AI Centre, The University of Sydney

*Equal contribution
**Problem** – Attribute-based person search

**Query attribute descriptions**
- Teenager
- Backpack
- Pants
- Short bottom wear
- Short top wear
- Long hair
- Female
- Top white
- Bottom blue

**Gallery images**

**Ranked retrieval results**

Images source: Market-1501 dataset.
Differences with zero-shot learning

1. No class prototypes for unseen classes
2. Large intra-class variations and inter-class similarities

| Attribute-based person search (Market-1501) | Zero-shot learning (AWA2) |
|--------------------------------------------|---------------------------|
| Category: 0                               | Category: Siamese cat     |
| • No bag                                  | • Brown                   |
| • Long hair                               | • Black                   |
| • Female                                  | • Small                   |
| ...                                       | ...                       |
| Category: 1                               | Category: tiger           |
| • Handbag                                 | • Orange                  |
| • Long hair                               | • Black                   |
| • Female                                  | • Big                     |
| ...                                       | ...                       |
| Category: 2                               | Category: Lion            |
| • No bag                                  | • Brown                   |
| • Short hair                              | • Yellow                  |
| • Male                                    | • Big                     |
| ...                                       | ...                       |
| ~508 training categories                  | ~40 training categories   |
| ~26 images per category                   | ~609 images per class     |

Inter-class similarity

Intra-class variation

- Teenager
- Short top
- Bottom black
- Black
- White
- Big
Motivation – Symbiosis

- A close and long-term biological interaction
- Mutualistic symbiosis relationship

Image source: https://icanhas.cheezburger.com/tag/Symbiosis
Method – Symbiotic Adversarial Learning

(a) Embed
(b) Embed + adversarial
(c) SAL

- Embed: Semantic space
- Embed + adversarial: Visual space
- SAL: Common space

: Real features  : Synthetic features  : Common space features  : Semantic space  : Visual space  : Common space
### Method – Symbiotic Adversarial Learning

- **Semantic space** $f_{a}$
- **Visual space** $f_{v}$

#### Attributes $a_{i}$
- Teenager
- Crossbody bag
- Pants
- Female
- Top white
- Bottom black

#### Data flow for paired images and attributes
- **Synthesis adversarial learning**
- **Alignment adversarial learning**

#### Loss function

\[
L_{\text{cat}} = - \sum_{i=1}^{N} \log(p_{\text{cat}}(x_{i}, y_{i}))
\]

\[
L_{\text{att}} = - \sum_{i=1}^{N} \sum_{j=1}^{m} (a_{(i,j)} \log(p_{\text{att}}^{(j)}(x_{i})) + (1 - a_{(i,j)}) \log(1 - p_{\text{att}}^{(j)}(x_{i})))
\]

\[
L_{\text{embed}} = L_{\text{cat}} + L_{\text{att}}.
\]
Method – Symbiotic Adversarial Learning

Three types of inputs to discriminator $D_1$:
1. The fake input pairs $(\tilde{f}_a, f_v)$, where $\tilde{f}_a = G_a(f_v)$.
2. The fake input pairs $(f_a, \tilde{f}_v)$, where $\tilde{f}_v = G_v(f_a)$.
3. The real input pairs $(f_a, f_v)$.

$$L_{\text{gan1}}(G_A, G_V, D_1) = \mathbb{E}_{(f_a, f_v) \sim p(f_a, f_v)}[\log(D_1(f_a, f_v))]$$
$$+ \frac{1}{2} \mathbb{E}_{f_a \sim p(f_a)}[\log(1 - D_1(f_a, \tilde{f}_v))]$$
$$+ \frac{1}{2} \mathbb{E}_{f_v \sim p(f_v)}[\log(1 - D_1(\tilde{f}_a, f_v))]$$

$$L_{\text{cyc}}(G_A, G_V) = \mathbb{E}_{f_a \sim p(f_a)}[||G_A(G_V(f_a), z)) - f_a||_2]$$
$$L_{\text{consis}}(G_A, G_V) = \mathbb{E}_{f_v \sim p(f_v)}[||E_A(\tilde{f}_a) - E_V(f_v)||_2] + \mathbb{E}_{f_a \sim p(f_a)}[||E_V(\tilde{f}_v) - E_A(f_a)||_2]$$
$$+ \mathbb{E}_{(f_a, f_v) \sim p(f_a, f_v)}[||E_A(\tilde{f}_a) - E_A(f_a)||_2] + \mathbb{E}_{(f_a, f_v) \sim p(f_a, f_v)}[||E_V(\tilde{f}_v) - E_V(f_v)||_2]$$
Method – Symbiotic Adversarial Learning

\[
\begin{align*}
L_{\text{gan}_2}(G_C, D_2) &= \mathbb{E}_{f_v \sim p(f_v)}[\log D_2(E_V(f_v))] \\
&+ \mathbb{E}_{f_a \sim p(f_a)}[\log(1 - D_2(E_A(f_a)))] \\
L_{\text{aug}_1}(G_C, D_2) &= \mathbb{E}_{f_a \sim p(f_a)}[\log D_2(E_V(\tilde{f}_v))] \\
&+ \mathbb{E}_{f_v \sim p(f_v)}[\log(1 - D_2(E_A(\tilde{f}_a)))] \\
L_{\text{aug}_2}(E_A, E_V) &= L_{\text{embed}}(\tilde{f}_a) + L_{\text{embed}}(\tilde{f}_v). \\
L_{\text{aug}} &= L_{\text{aug}_1} + L_{\text{aug}_2}. \\
L_{\text{align-adv}} &= L_{\text{gan}_2} + L_{\text{aug}}.
\end{align*}
\]
Method – Symbiotic Adversarial Learning

Attributes $a_i$
- Teenager
- Crossbody bag
- Pants
- Female
- Top white
- Bottom black

Sampled Attributes $a^{unseen}$
- Teenager
- Backpack
- Pants
- Female
- Top red
- Bottom black

Data flow for paired images and attributes

Data flow for synthetic unseen data

Synthesis adversarial learning

Alignment adversarial learning
Table 1. Attribute-based person search performance evaluation. Best results are shown in **bold**. The second-best results are underlined.

| Metric (%)   | Market-1501 Attributes | PETA          |
|--------------|-------------------------|---------------|
| Model        | Reference               | mAP | rank1 | rank5 | rank10 | mAP | rank1 | rank5 | rank10 |
| DeepCCA [1]  | ICML’13                 | 17.5 | 30.0  | 50.7  | 58.1  | 11.5 | 14.4  | 20.8  | 26.3   |
| DeepMAR [23] | ACPR’15                 | 8.9  | 13.1  | 24.9  | 32.9  | 12.7 | 17.8  | 25.6  | 31.1   |
| DeepCCAIE [50]| ICML’15               | 9.7  | 8.1   | 24.0  | 34.6  | 14.5 | 14.2  | 22.1  | 30.0   |
| 2WayNet [8]  | CVPR’17                 | 7.8  | 11.3  | 24.4  | 31.5  | 15.4 | 23.7  | 38.5  | 41.9   |
| CMCE [24]    | ICCV’17                 | 22.8 | 35.0  | 51.0  | 56.5  | 26.2 | 31.7  | 39.2  | 48.4   |
| ReViSE [41]  | ICCV’17                 | 17.7 | 24.2  | 45.2  | 57.6  | 31.1 | 30.5  | 57.0  | 61.5   |
| MMCC [9]     | ECCV’18                 | 22.2 | 34.9  | 58.7  | 70.2  | 33.9 | 33.5  | 57.0  | 69.0   |
| AAI PR [53]  | IJCAI’18               | 20.7 | 40.3  | 49.2  | 58.6  | 27.9 | 39.0  | 53.6  | 62.2   |
| AIHM [7]     | ICCV’19                 | 24.3 | 43.3  | 56.7  | 64.5  | -   | -     | -     | -      |
| **SAL (Ours)**| **ICCV’19**           | **29.8** | **49.0** | **68.6** | **77.5** | **41.2** | **47.0** | **66.5** | **74.0** |
Ablation studies

Table 2. Component analysis of SAL on PETA dataset.

| Metric (%)            | mAP | rank1 | rank5 | rank10 |
|-----------------------|-----|-------|-------|--------|
| Embed                 | 31.3| 34.0  | 57.0  | 64.5   |
| Embed + adv           | 35.0| 37.5  | 60.5  | 66.5   |
| Embed + symb-adv      | 40.6| 44.0  | 64.0  | 70.5   |
| Embed + symb-adv + unseen(SAL) | 41.2| 47.0  | 66.5  | 74.0   |

Table 3. Effect of interactions between two GANs on PETA dataset.

| Metric (%)          | mAP | rank1 | rank5 | rank10 |
|--------------------|-----|-------|-------|--------|
| SAL - $L_{aug}$    | 35.4| 38.0  | 60.0  | 69.0   |
| SAL - $L_{consis}$ | 35.2| 39.5  | 56.5  | 66.0   |
| SAL (Full interaction) | 41.2| 47.0  | 66.5  | 74.0   |

Table 4. Comparing stage-wise training vs. symbiotic training scheme.

| Metric (%)                        | mAP | rank1 | rank5 | rank10 |
|-----------------------------------|-----|-------|-------|--------|
| SAL w/ stage-wise training        | 35.0| 41.0  | 58.0  | 65.0   |
| SAL w/ symbiotic training         | 41.2| 47.0  | 66.5  | 74.0   |
Visualized retrieval results

The green/red border represents correct/wrong selections respectively.
Thank you!

Code at:

https://github.com/ycao5602/SAL