Dual Feature Augmentation Network for Generalized Zero-shot Learning

Lei Xiang[1], Yuan Zhou[1], Haoran Duan[2], Yang Long[2]
[1]Nanjing University of Information Science and Technology [2]Durham University

BMVC 2023

Motivation

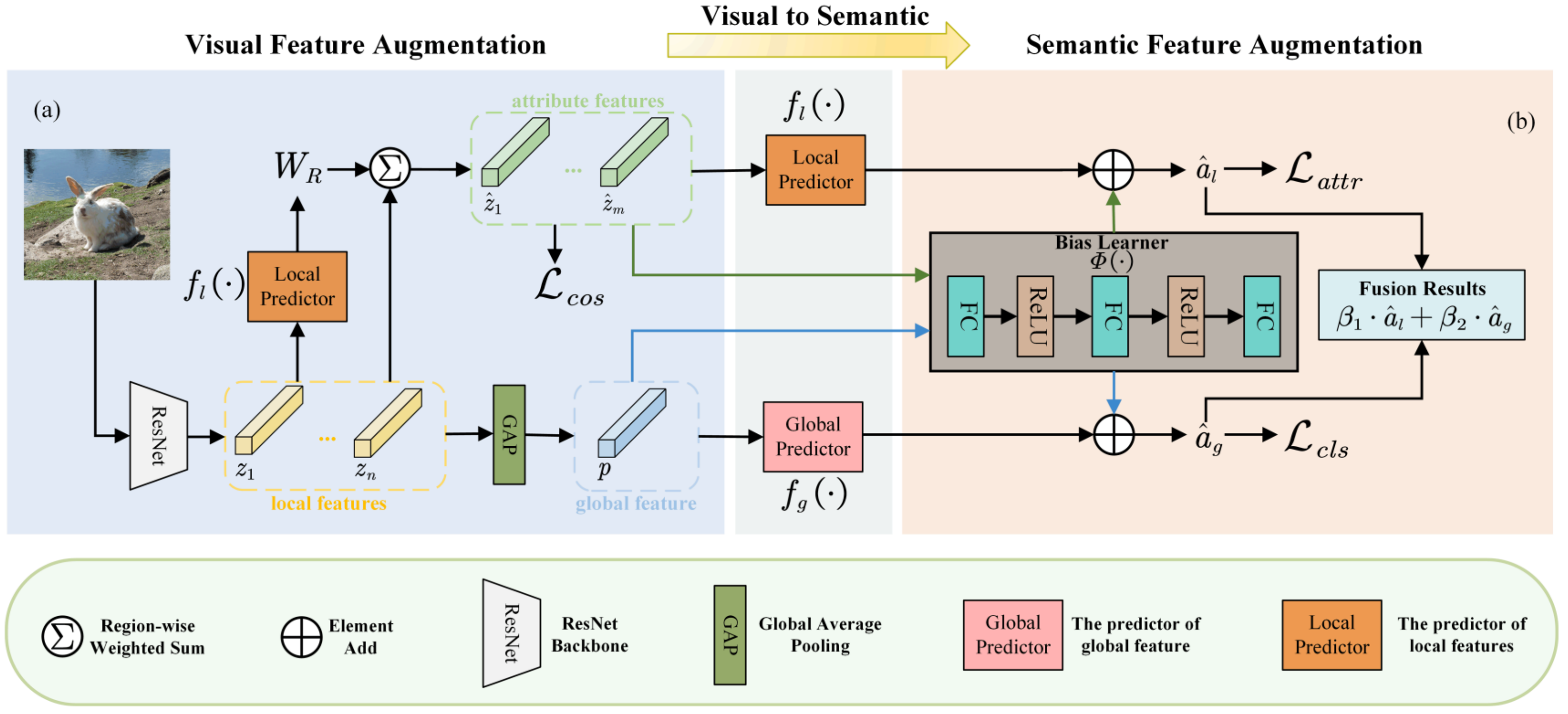
- The complex entanglement among different attributes' visual features in the embedding space.
- The diversity of attributes in images of the same category leads to inaccurate predictions.
- Using a single mapping function to map different granular visual features to semantic space degrades model performance.

Contribution

- We propose a **visual feature augmentation** that explicitly extracts attribute features and adopts a cosine similarity loss to disentangle them in the embedding space, enhancing the visual features.
- We propose a **semantic feature augmentation** containing a **bias learner**, which estimates an offset to alleviate the difference between the actual and predicted attributes, leading to improved class-level semantic features for each image.
- We employ **two mapping functions** to avoid inconsistencies in the mapping process of different granular visual features, thus reinforcing the semantic features corresponding to varying visual features.

Code **Paper**

Method



$$\mathcal{L}_{cos} = \left\| \hat{Z}^T \hat{Z} - I \right\|_2$$

where $I \in \mathbb{R}^M$ is the identify matrix and $\hat{Z} = [\hat{z}_1, \dots, \hat{z}_m]$

$$\mathcal{L}_{attr} = -\log \frac{\exp \langle \hat{a}_l, a_y \rangle}{\sum_{k=1}^{C_s} \exp \langle \hat{a}_l, a_k \rangle}$$

$$\mathcal{L}_{cls} = -\log \frac{\exp \langle \hat{a}_g, a_y \rangle}{\sum_{k=1}^{C_s} \exp \langle \hat{a}_g, a_k \rangle}$$

where a_y is the ground truth semantic feature

Results

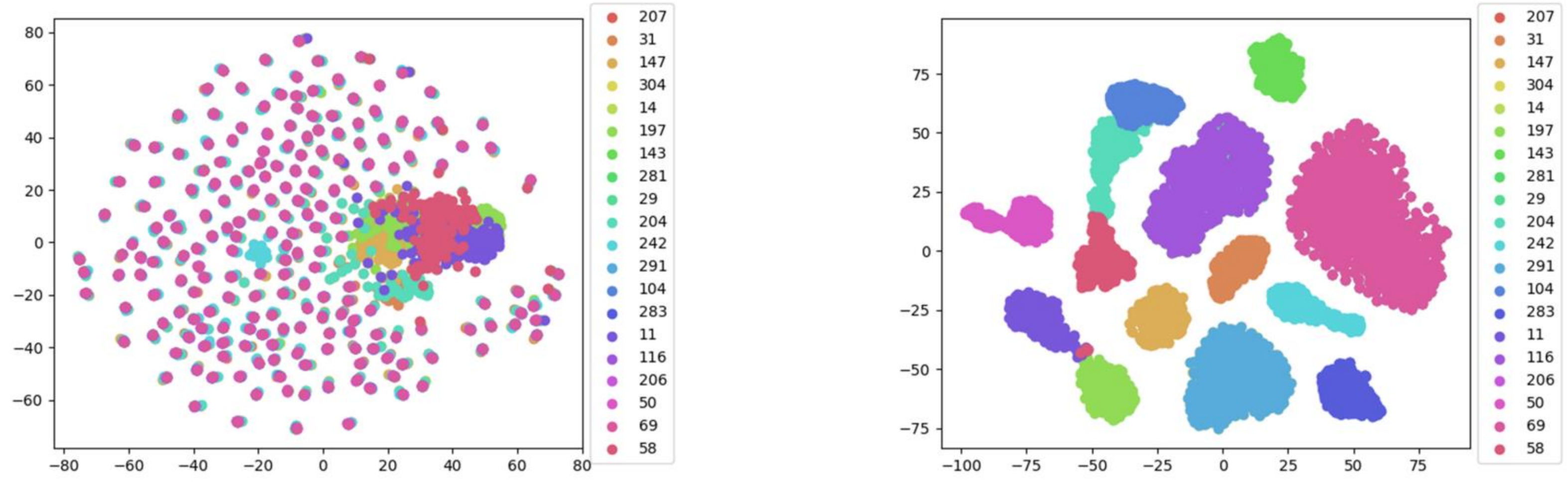
Comparison with State-of-the-Art Methods

Methods	CUB GZSL U	CUB GZSL S	CUB GZSL H	CUB CZSL acc	SUN GZSL U	SUN GZSL S	SUN GZSL H	SUN CZSL acc	AWA2 GZSL U	AWA2 GZSL S	AWA2 GZSL H	AWA2 CZSL acc
Generative Methods												
f-VAEGAN-D2 [40]	48.4	60.1	53.6	61.0	45.1	38.0	41.3	64.7	57.6	70.6	63.5	71.1
E-PGN [43]	52.0	61.1	56.2	72.4	–	–	–	–	52.6	83.5	64.6	73.4
Composer [15]	56.4	63.8	59.9	69.4	55.1	22.0	31.4	62.6	62.1	77.3	68.8	71.5
GCM-CF [44]	61.0	59.7	60.3	–	47.9	37.8	42.2	–	60.4	75.1	67.0	–
CE-GZSL [13]	63.9	66.8	65.3	77.5	48.8	38.6	43.1	63.3	63.1	78.6	70.0	70.4
FREE [5]	55.7	59.9	57.7	–	47.4	37.2	41.7	–	60.4	75.4	67.1	–
Non-generative Methods												
DAZLE [16]	56.7	59.6	58.1	–	52.3	24.3	33.2	–	60.3	75.3	67.1	–
APN [42]	65.3	69.3	67.2	72.0	41.9	34.0	37.6	61.6	57.1	72.4	63.9	68.4
GEM-ZSL [22]	64.8	77.1	70.4	77.8	38.1	35.7	36.9	62.8	64.8	77.5	70.6	67.3
SR2E [11]	61.6	70.6	65.8	–	43.1	36.8	39.7	–	58.0	80.7	67.5	–
MSDN [7]	65.3	69.3	67.2	72.0	52.2	34.2	41.3	65.8	62.0	74.5	67.7	70.1
TransZero [6]	69.3	68.3	68.8	76.8	52.6	33.4	40.8	65.6	61.3	82.3	70.2	70.1
ours	65.4	79.7	71.8	77.3	51.0	36.4	42.5	67.9	58.9	88.0	70.5	67.4

The best, second-best and third-best results are marked in Red, Blue and Green, respectively.

Experiments

T-SNE visualizations of attribute features



(a) without $\mathcal{L}_{cos}$ (b) with $\mathcal{L}_{cos}$

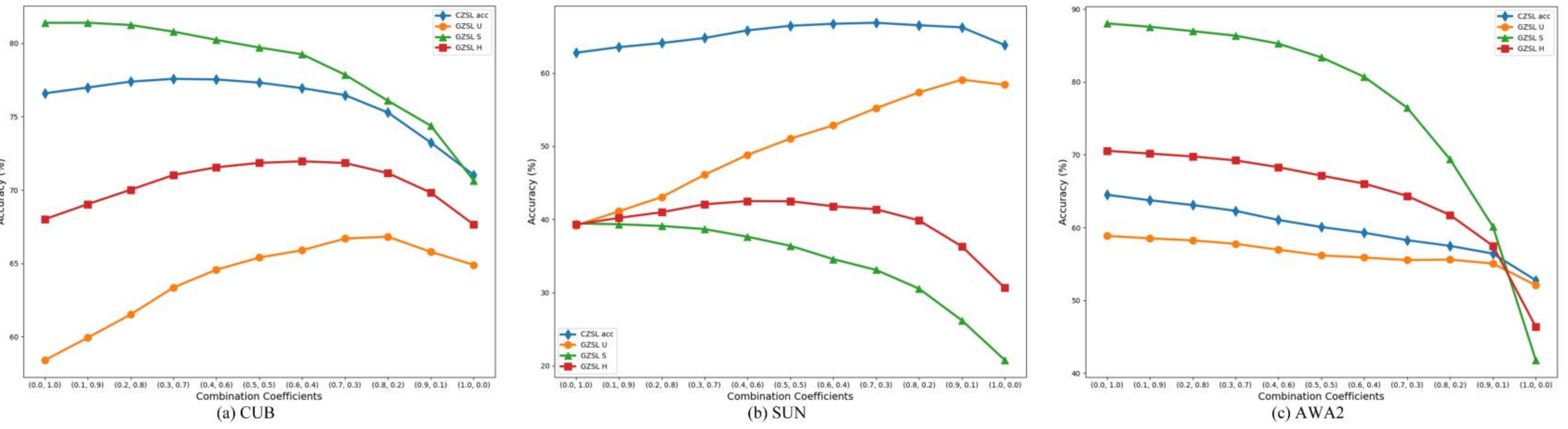
Effect of different modules

Methods	CUB GZSL U	CUB GZSL S	CUB GZSL H	CUB CZSL acc	SUN GZSL U	SUN GZSL S	SUN GZSL H	SUN CZSL acc	AWA2 GZSL U	AWA2 GZSL S	AWA2 GZSL H	AWA2 CZSL acc
single predictor	54.3	72.2	62.0	65.0	**43.3**	22.8	29.8	53.6	**56.8**	76.4	65.2	60.2
two predictors	60.8	80.3	69.2	76.0	39.7	32.2	35.6	58.5	55.5	87.3	67.8	62.7
two predictors + bias leaner	**63.2**	**81.2**	**71.1**	**76.7**	38.8	**33.2**	**35.7**	**59.3**	55.1	**88.9**	**68.0**	**63.4**

Effect of loss components

Methods	CUB GZSL U	CUB GZSL S	CUB GZSL H	CUB CZSL acc	SUN GZSL U	SUN GZSL S	SUN GZSL H	SUN CZSL acc	AWA2 GZSL U	AWA2 GZSL S	AWA2 GZSL H	AWA2 CZSL acc
$\mathcal{L}_{cls}$	54.3	72.2	62.0	65.0	43.3	22.8	29.8	53.6	56.8	76.4	65.2	60.2
$\mathcal{L}_{attr}$	51.8	69.4	59.3	61.5	45.3	26.5	33.5	62.0	50.6	88.4	64.4	63.7
$\mathcal{L}_{cls}$+$\mathcal{L}_{attr}$	60.8	**80.3**	69.2	76.0	39.7	32.2	35.6	58.5	55.5	87.3	67.8	62.7
$\mathcal{L}_{cls}$+$\mathcal{L}_{attr}$+$\mathcal{L}_{cos}$	**65.4**	79.7	**71.8**	**77.3**	**51.0**	**36.4**	**42.5**	**67.9**	**58.9**	**88.0**	**70.5**	**67.4**

Effect of coefficients (β1, β2)



- In fine-grained datasets, the CZSL accuracy and the harmonic mean increase first and then decrease.
- For the coarse-grained dataset, as the value of β1 gradually increases, the CZSL accuracy and the harmonic mean decrease and reach a minimum when β1 equals 1.