Fine-grained Few-shot Recognition by Deep Object Parsing

Ruizhao Zhu, Pengkai Zhu, Samarth Mishra, Venkatesh Saligrama
Boston University

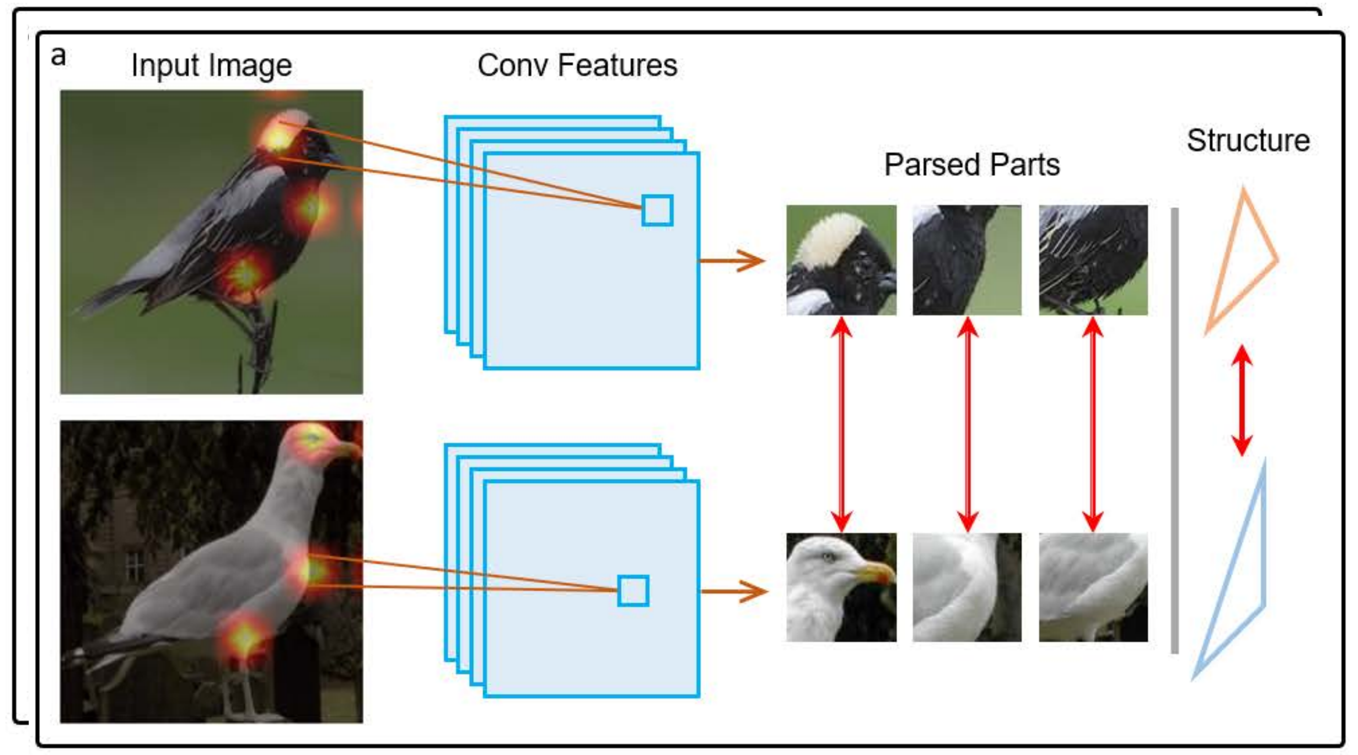
BMVC 2023

Motivation

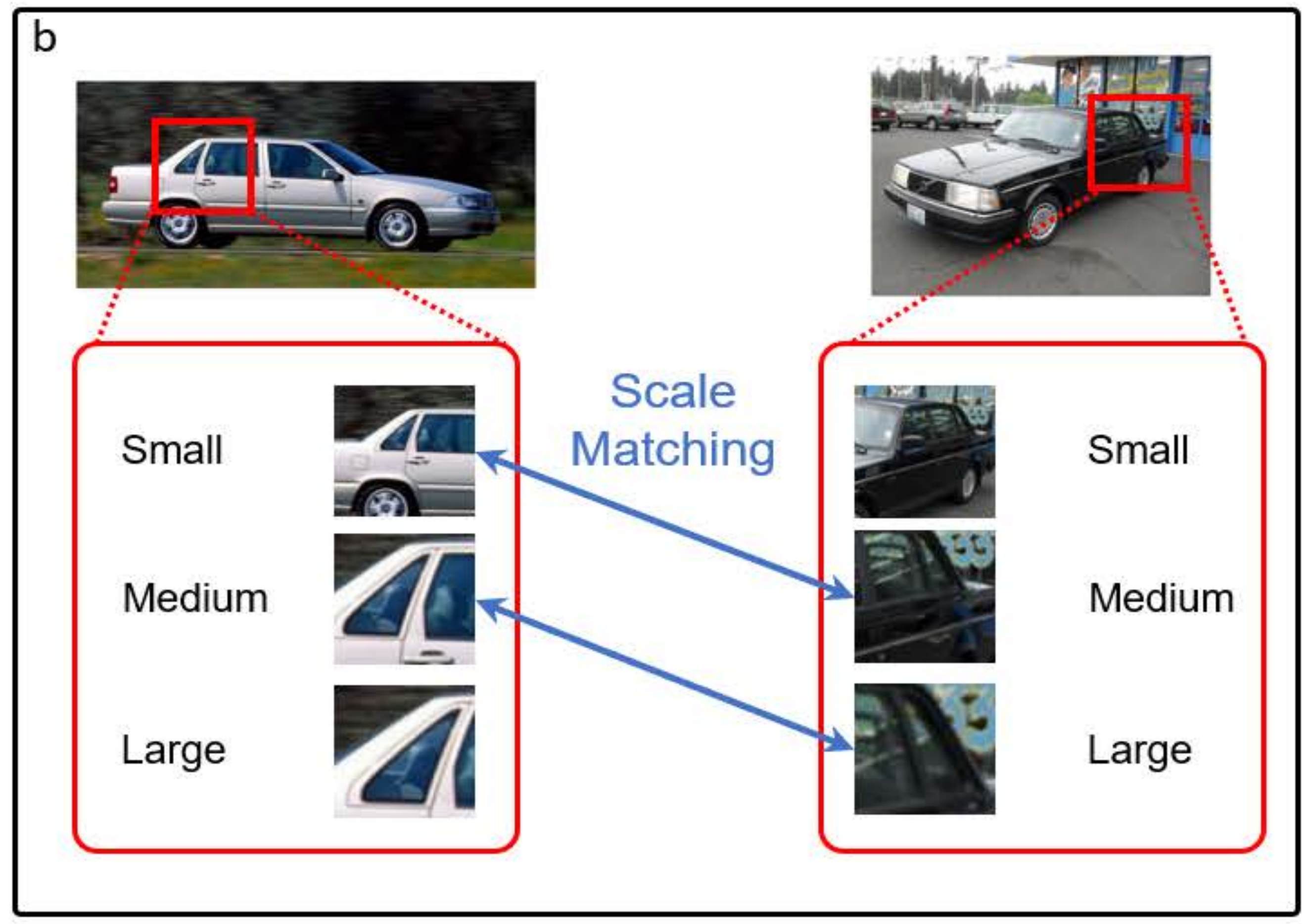
- Salient parts and the geometry of the parts are important to good recognition accuracy.
- Same part may be distorted or absent in the support samples due to the perspective and pose changes.

Our solution: Deep Object Parsing (DOP)

- DOP automatically learns salient parts and geometry.

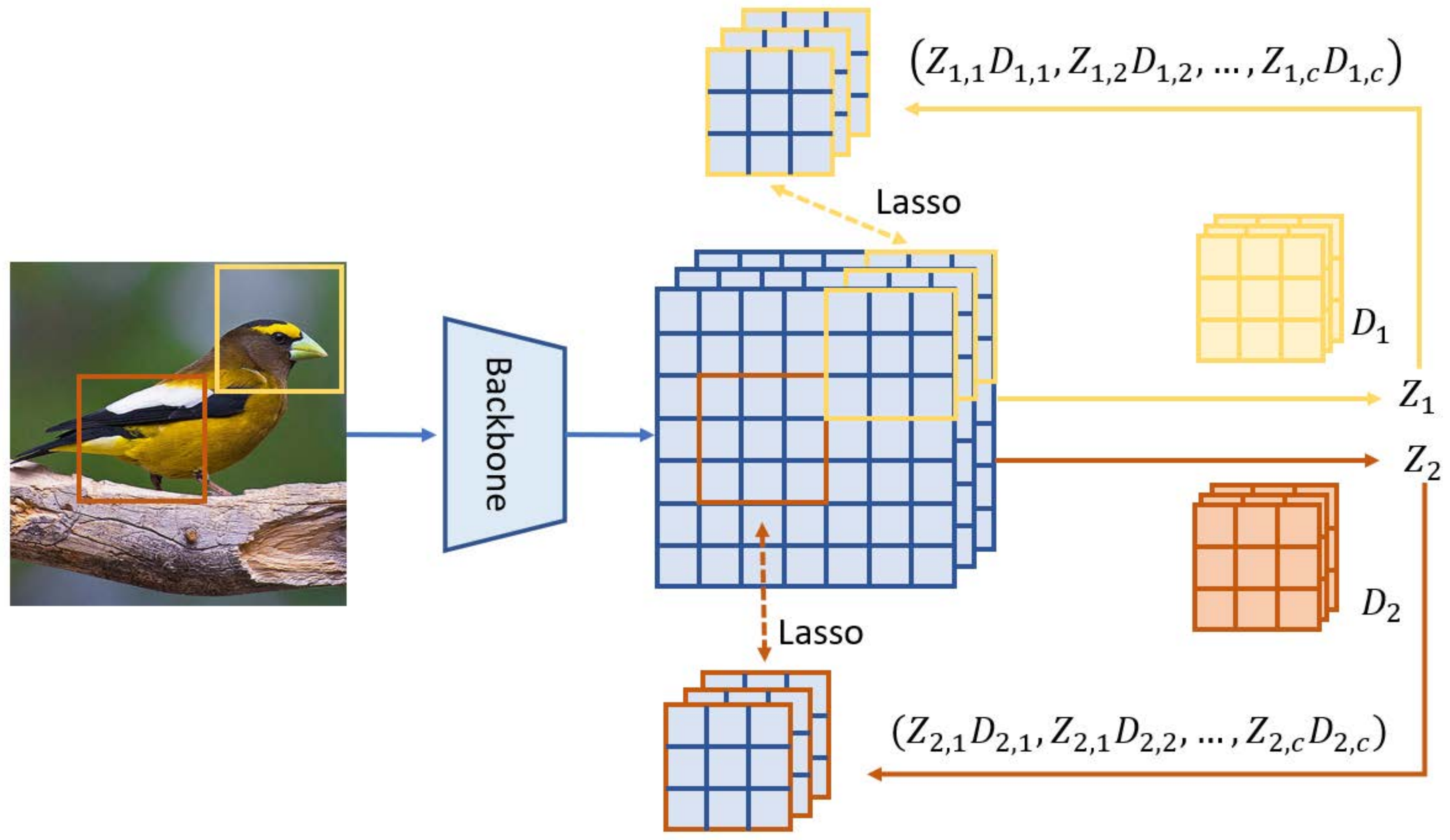


- DOP matches different scales of parts.



Method

- A feature $\phi \in \mathbb{R}^{G\times G\times C}$ from instance x is parsed into K parts and Each part p is an $s{\times}s$ mask $M(\mu)$ around μ. Part is expressed coefficient $Z_{p,c}(\mu)$ of channel-wise templates $D_{p,c}$



Algorithm 1: Object Parsing using DNNs

Input: image x
Parametric functions: convolutional backbone f, template collections $\mathcal{D}_p$
Get the convolutional feature $\phi = f(x)$
for $p \in [K]$ **do**
  Estimate $\mu_p = \text{argmax}\,\Pr_p(\mu \mid \phi; [\theta_{p,c}], [\lambda_c])$ by Eq. (4)
  Compute $z'_{p,c}(\mu) = \frac{(D_{p,c} * \delta_\mu) \odot \phi_c(\mu)}{\|D_{p,c}\|^2}$
  Thresholding: $z_{p,c}(\mu) = S_\varsigma(z'_{p,c}(\mu))$
end for
Output: Part locations $[\mu_p]_{p\in[K]}$ and template coefficients $[z_{p,c}]_{p\in[K],c\in[C]}$

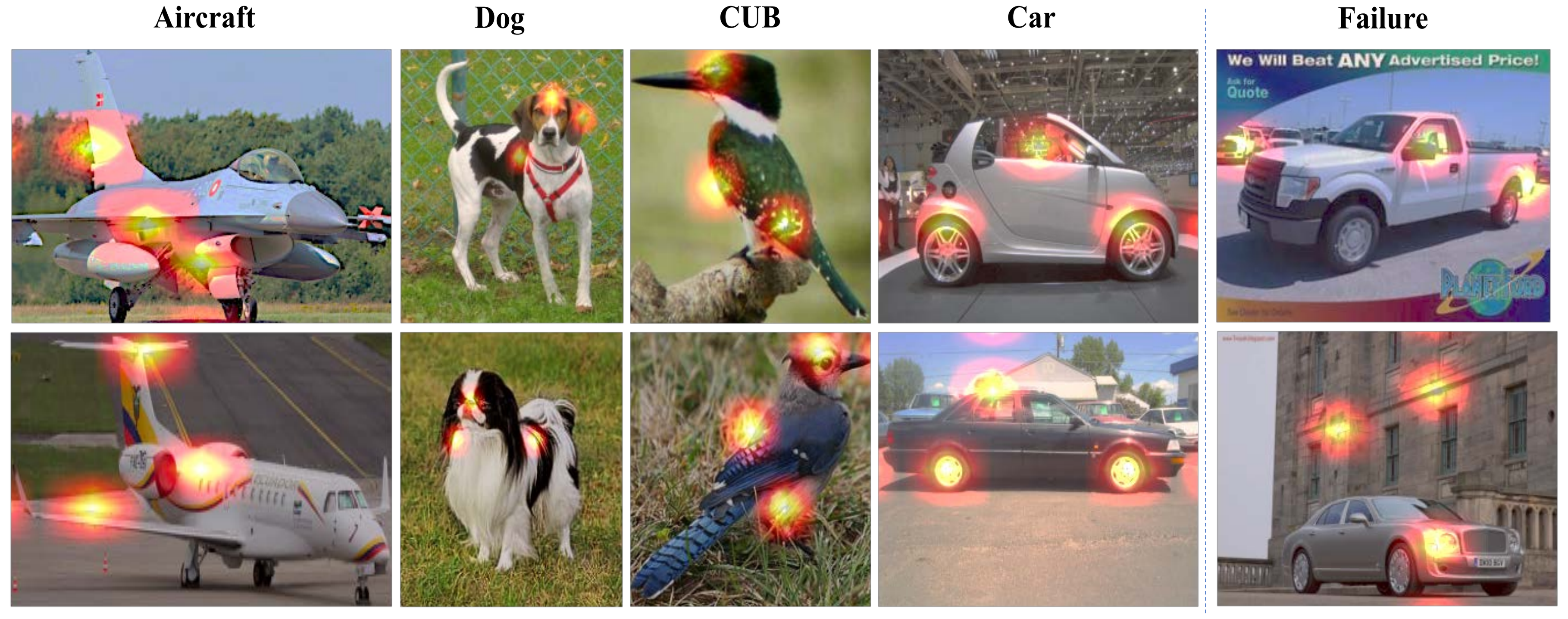
- Part expression as LASSO regression
- Part location estimation from optimal part expression
- Few-shot recognition based on distances of features and geometry

Experiments

Methods	Backbones	CUB 1-shot	CUB 5-shot	Dog 1-shot	Dog 5-shot
VFD†[51]	ResNet12	79.12±0.83	91.11±0.24	**70.60±0.91**	**85.74±0.53**
TOAN[19]	ResNet12	67.17±0.81	82.09±0.56	51.83±0.80	69.83±0.66
FRN[49]	ResNet12	83.16±0.19	92.59±0.23	62.07±0.22	83.18±0.14
TDM[23]	ResNet12	83.36±0.22	92.08 ±0.13	57.64±0.22	75.77±00.16
HelixFormer[55]	ResNet12	81.66±0.30	91.83±0.17	65.92±0.49	80.65±0.36
DOP	ResNet18	**82.62±0.65**	**92.61±0.38**	**70.56±0.75**	**84.75±0.41**
DOP	ResNet12	**83.39±0.82**	**93.01±0.43**	70.10±0.79	85.12±0.55

Methods	Backbones	Car 1-shot	Car 5-shot	Aircraft 1-shot	Aircraft 5-shot
CTX[8]	ResNet12	55.66±0.22	73.78±0.16	65.53±0.22	79.31±0.13
TOAN[19]	ResNet12	76.62±0.70	89.57±0.40	-	-
FRN[49]	ResNet12	55.49±0.21	74.54±0.16	69.58±0.22	82.98±0.14
TDM[23]	ResNet12	68.36±0.22	86.14 ±0.13	70.89±0.22	84.54±0.16
HelixFormer[55]	ResNet12	79.40±0.43	92.26±0.15	74.01±0.54	83.11±0.41
DOP	ResNet18	**81.41±0.71**	**93.48±0.38**	**83.26±0.24**	**92.41±0.45**
DOP	ResNet12	**81.83±0.78**	**93.84±0.45**	84.50±0.25	93.35±0.48

Visualization of exemplar parts location learned by DOPM .



Visualization of re-weighting.