Fine-Grained Object Classification via Self-Supervised Pose Alignment

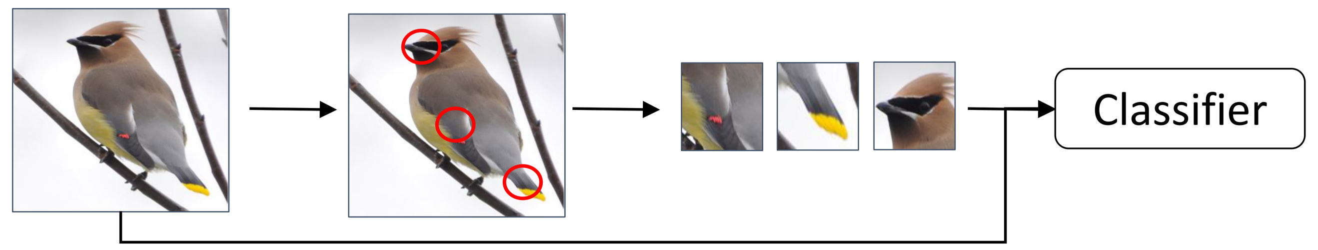
Xuhui Yang[1], Yaowei Wang[1], Ke Chen[2], Yong Xu[2], Yonghong Tian[1]
[1]Peng Cheng Laboratory, [2]South China University of Technology
yangxh@pcl.ac.cn, wangyw@pcl.ac.cn, chenk@scut.edu.cn,yxu@scut.edu.cn, tianyh@pcl.ac.cn

Fine-grained Image Classification

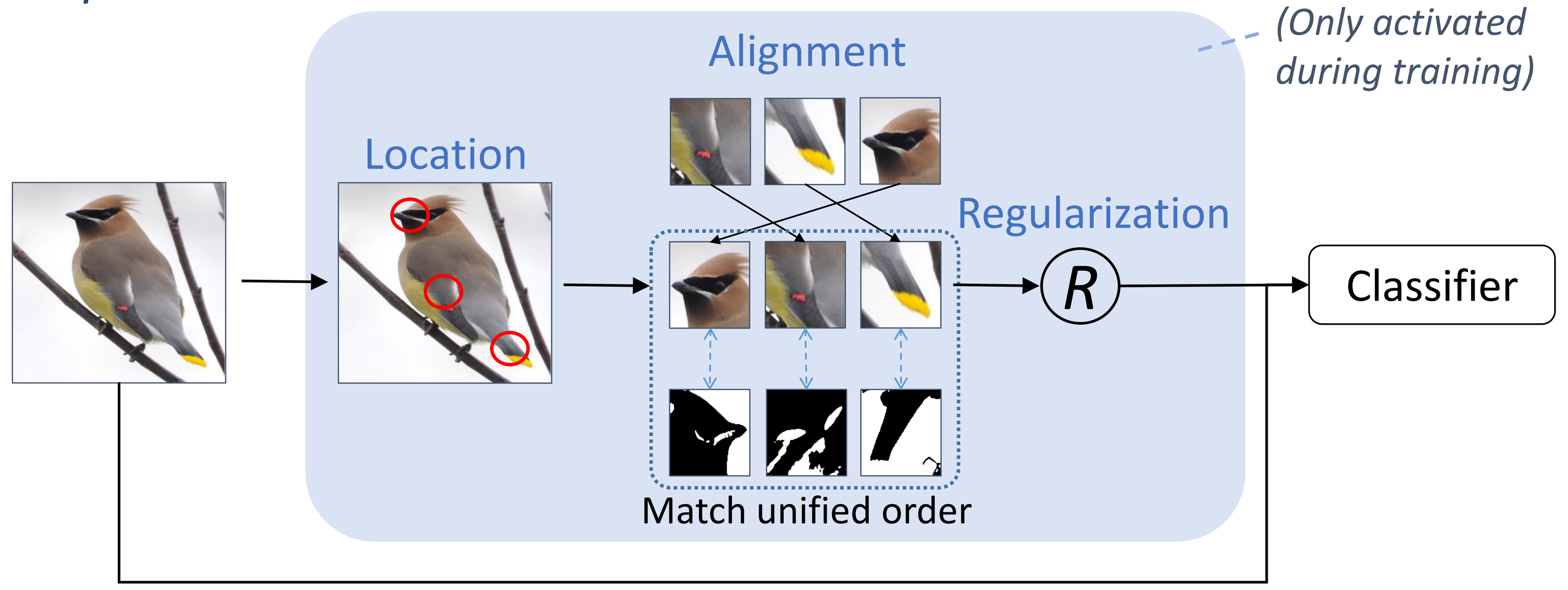
- Semantic patterns in image-based object classification are determined by visual appearance and shape of object classes.
- A good representation in fine-grained classification should not only be sensitive to the subtle detail changes that usually anchored on specific parts but also be invariant to the deformations of object parts and the changes of viewing angles.

Introduction

Conventional part-based methods:



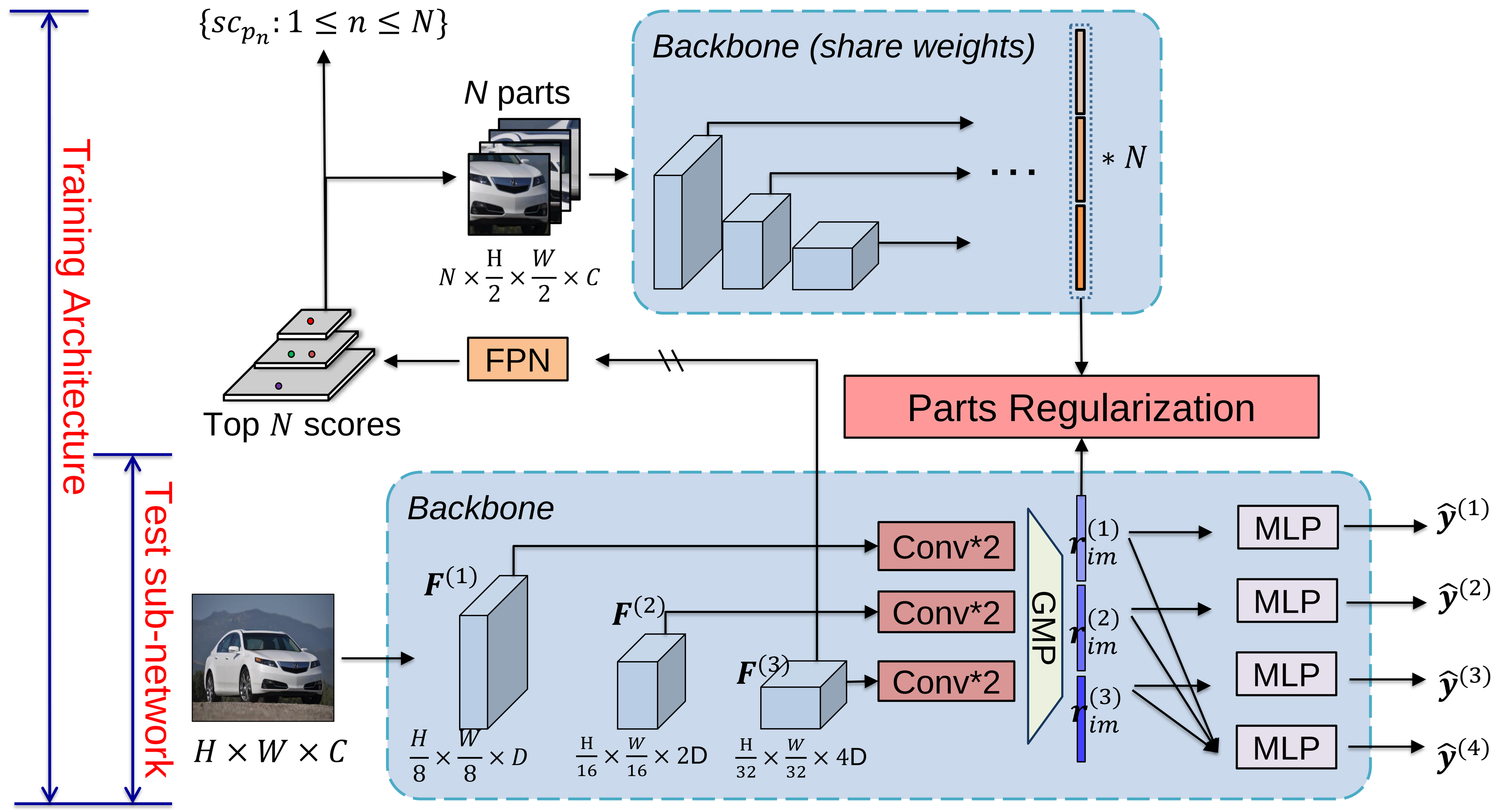
Proposed method:



- ◆ Localize discriminative parts via weakly supervised detection to capture visual details.
- ◆ Align discriminative parts via self-supervised graph matching to eliminate pose variations.
- ◆ Regularize the image representation via feature regularization to speed inference.

Method

Architecture:

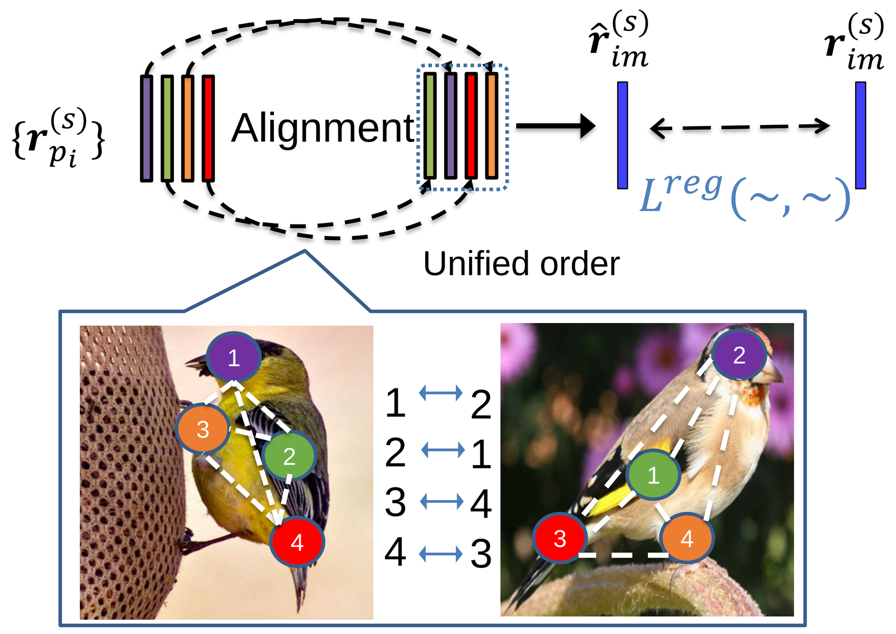


Parts detection:

1. $\{L_{p_n}^{cls}\}$: classification losses of parts given image label using shared backbone.
2. $\{sc_{p_n}\}$: confidence scores of top N discriminative parts.
3. $L^{rank}(\sim,\sim)$: ranking loss to encourage $\{L_{p_n}^{cls}\}$ and $\{sc_{p_n}\}$ to rank in reverse order.

Parts alignment and regularization:

See the figure on the right.



Match correlation matrix

Training loss:

Cross entropy loss of parts classification

KL divergence of distributions

$$L = L_{im}^{cls} + L_{parts}^{cls} + L^{rank} + \beta L^{reg}$$

Cross entropy loss of image classification

Ranking loss

Final prediction:

$$\hat{y}^{(final)} = \sum_{s=1}^{4} \hat{y}^{(s)}$$

Experiments

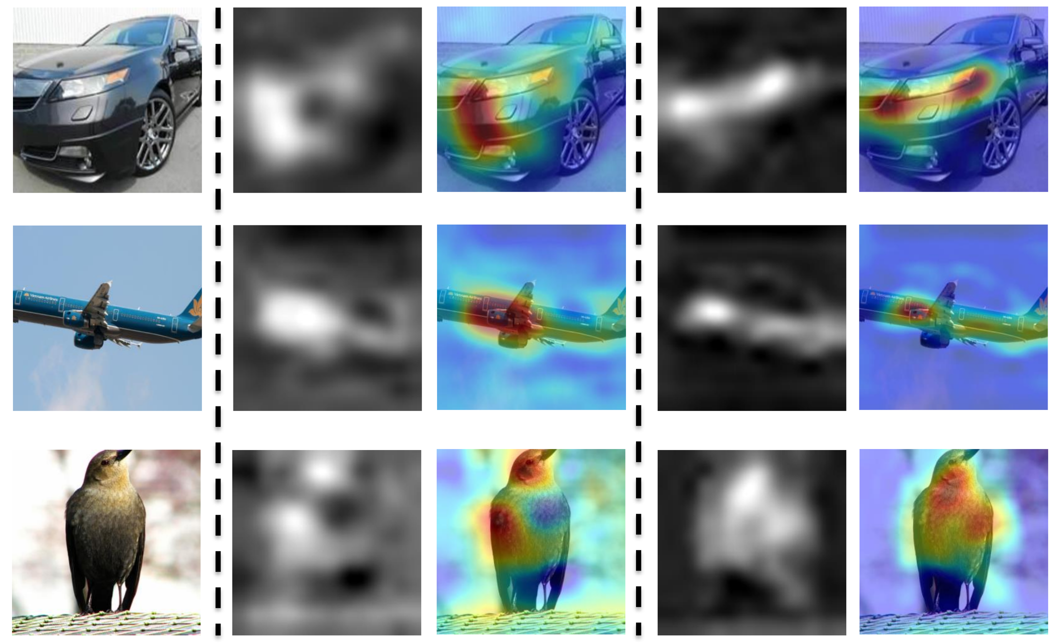
SOTA performance on benchmarks:

Method	Backbone	Accuracy (%)		
		CUB	CAR	AIR
B-CNN	VGG	84.1	91.3	84.1
MACNN	VGG19	86.5	92.8	89.9
MAMC	ResNet50	86.3	93.0	
NTS-Net	ResNet50	87.5	93.9	91.4
DCL	ResNet50	87.8	94.5	93.0
TASN	ResNet50	87.9	93.8	-
Cross-X	ResNet50	87.7	94.6	92.6
S3N	ResNet50	88.5	94.7	92.8
LIO	ResNet50	88.0	94.5	92.7
DF-GMM	ResNet50	88.8	94.8	93.8
PMG	ResNet50	89.6	95.1	93.4
API-Net	ResNet101	88.6	94.9	93.4
API-Net	DenseNet161	90.0	95.3	93.9
Our P2P-Net	ResNet34	89.5	94.9	92.6
Our P2P-Net	ResNet50	**90.2**	**95.4**	**94.2**

Computation complexity (values listed as train/test if they are different):

Model	Params (M)	FLOPs (G)	Time (sec)
ResNet50	23.92	16.44	0.064/0.034
NTS-Net	29.03	41.91	0.126/0.069
PMG	45.13	37.47	0.270/0.043
API-Net	46.06/42.91	31.53/31.52	0.104/0.054
P2P-Net	64.09/44.63	75.43/37.47	0.136/0.041

Class activation maps:



Input　Baseline　P2P-Net

Contributions

- ✓ Graph matching algorithm for alignment to mitigate object pose variations.
- ✓ Simple regularization scheme to accelerate inference without loss of accuracy.