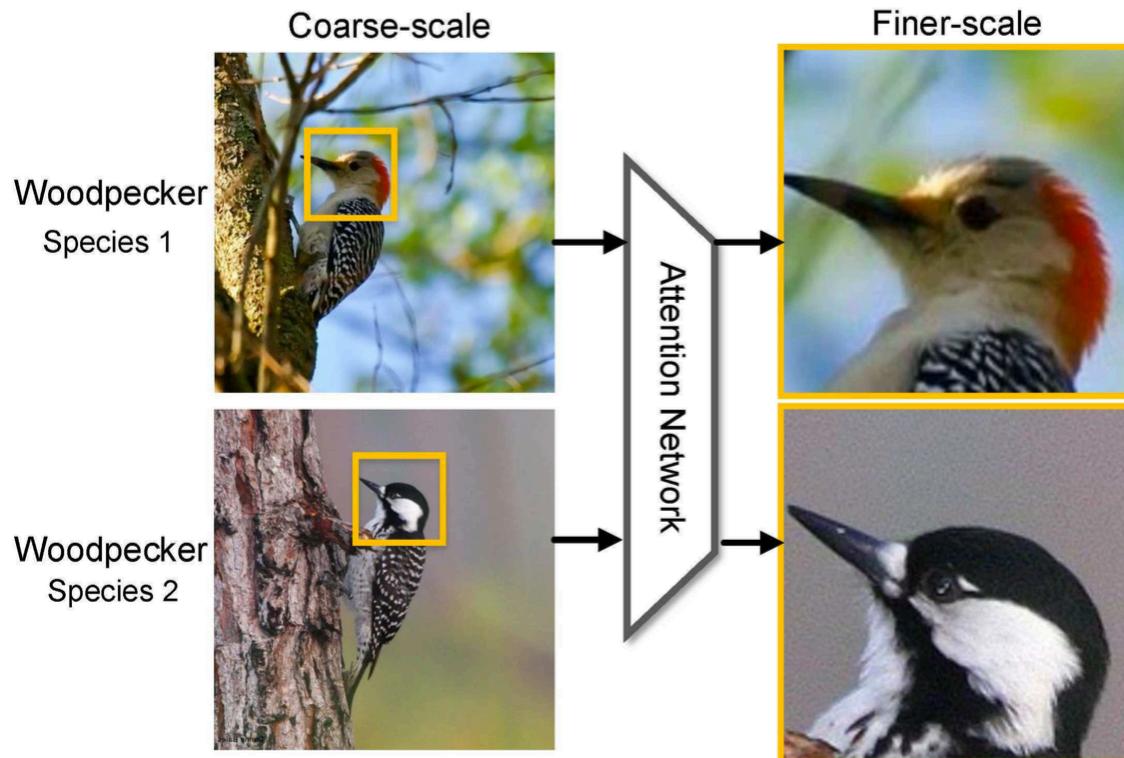


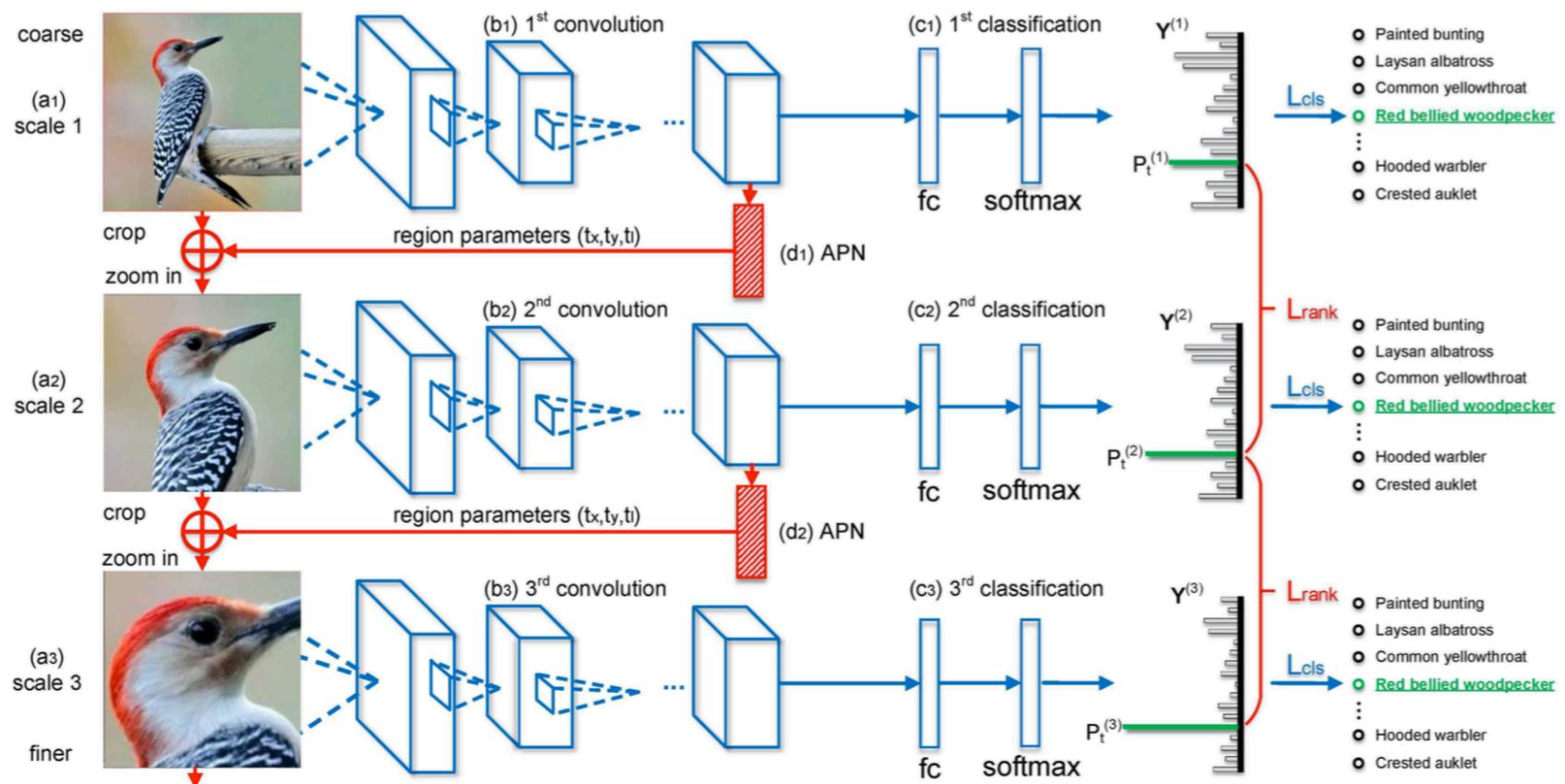
Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-grained Image Recognition

Jinna Cui
2017.7.22

Coarse and Finer



RA-CNN



APN: Attention Proposal Network

$$\circ p(X) = f(W_c * X)$$

X : input image

W_c : overall parameters

$f(\quad)$: fully-connected layers

APN: Attention Proposal Network

Localization & Amplification:

$$X^{att} = X \odot M(t_x, t_y, t_l)$$

t_x, t_y : the square's center coordinates in terms of x and y axis

$M(\cdot)$: attention mask

Classification & Ranking

$$L(X) = \sum_{s=1}^3 \{L_{cls}(Y^{(s)}, Y^*)\} + \sum_{s=1}^2 \{L_{rank}(p_t^{(s)}, p_t^{(s+1)})\}$$

$Y^{(s)}$: the predicted label vector from a specific scale

Y^* : the ground truth label vector

L_{cls} : classification loss

L_{rank} : ranking loss

L_{rank}

$$L_{rank} \left(p_t^{(s)}, p_t^{(s+1)} \right) = \max\{0, p_t^{(s)} - p_t^{(s+1)} + margin\}$$

Enforce: $p_t^{(s+1)} > p_t^{(s)} + margin$

Multi-scale Joint Representation

$$\{F_1, F_2 \dots F_N\}$$

F_i : donates the feature descriptor at a specific scale

Results



Results

Approach	Train Anno.	Accuracy
DeepLAC	✓	80.3
Part-RCNN	✓	81.6
FCAN	✓	84.3
B-CNN	✓	85.1
VGG-19		77.8
DVAN		79
B-CNN		84.1
RA-CNN(scale 2)		82.4
RA-CNN(scale 3)		81.2
RA-CNN(scale 1+2)		84.7
RA-CNN(scale 1+2+3)		85.3

Results

Stanford Dogs

Chihuahua



Japanese spaniel



Blenheim spaniel



Siberian husky



Brabancon griffon



Boston bull



Stanford Cars

Acura



Audi



Chrysler



Dodge



Chevrolet



Smart



Thank you!

Q&A