LEARNING MULTI-ATTENTION CONVOLUTIONAL NEURAL NETWORK FOR FINE-GRAINED IMAGE RECOGNITION

HELIANG ZHENG, JIANLONG FU, TAO MEI, JIEBO LUO

(ICCV, 2017)

7107053120 許庭瑄

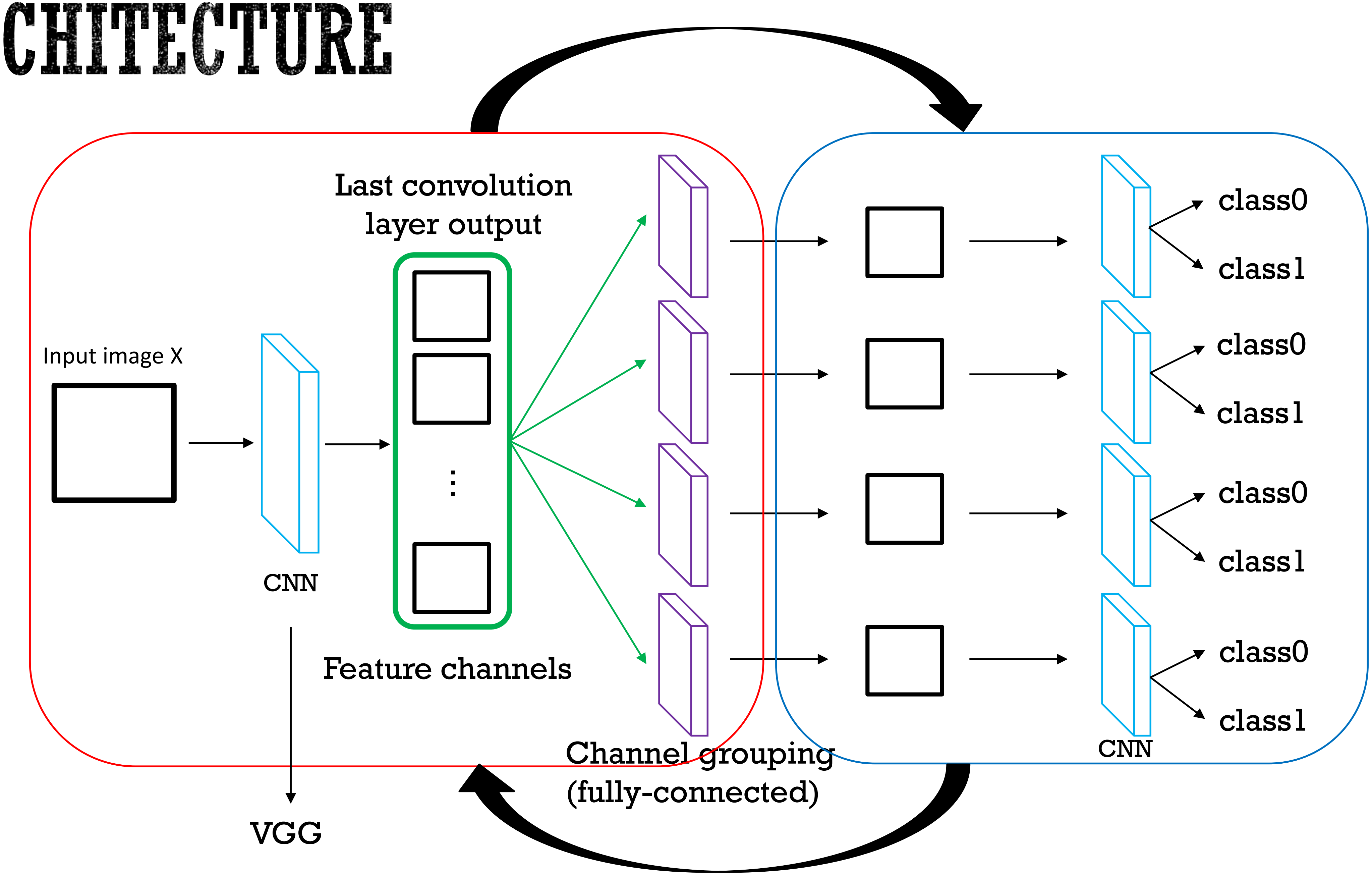
INTRODUCTION

- ~~200 species of birds~~

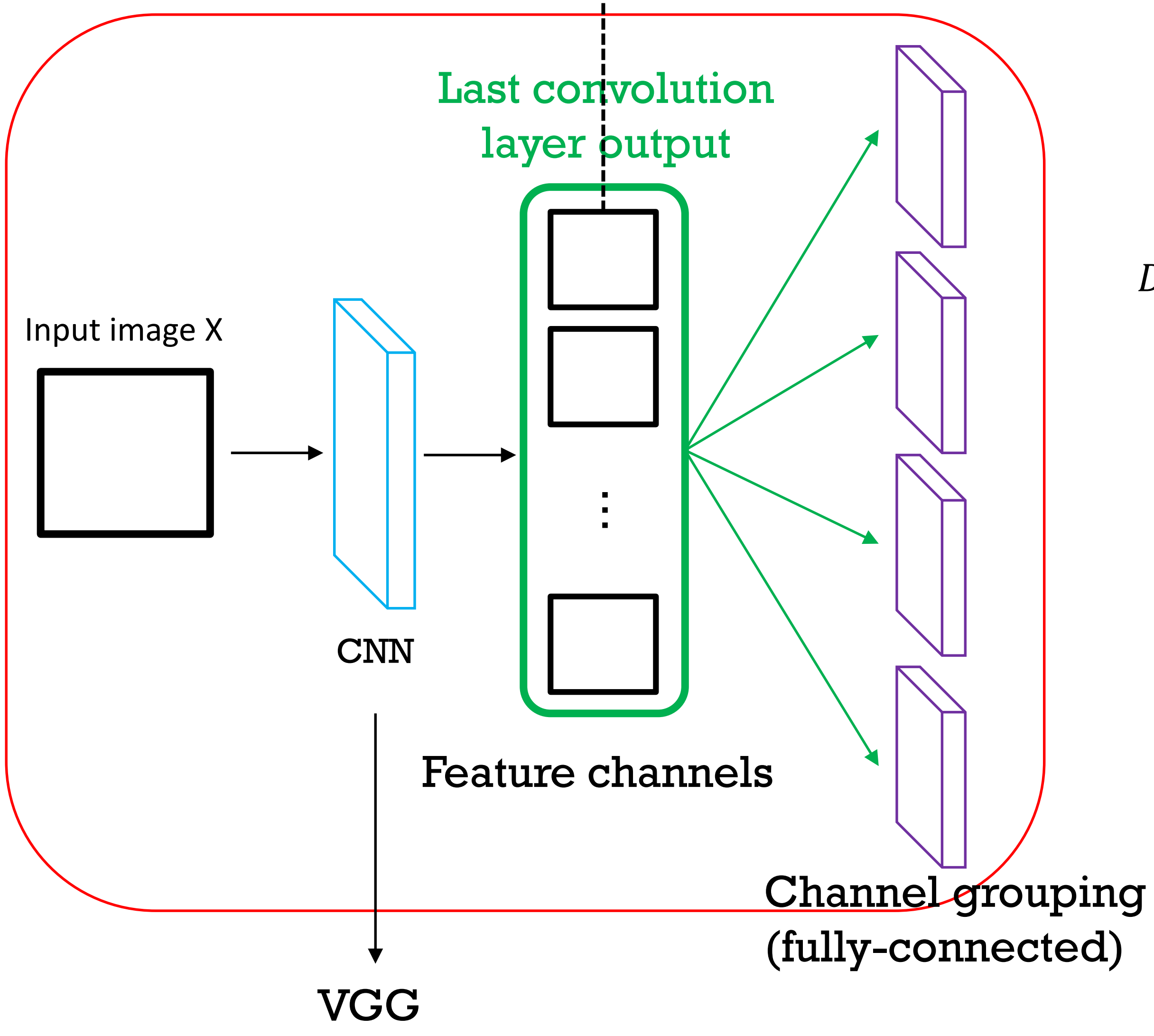
  → Two species of woodpecker

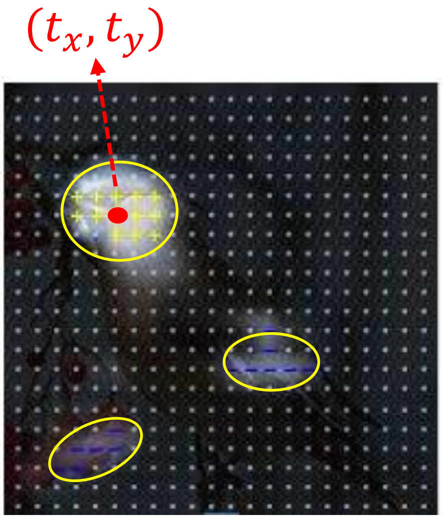
- Objective : Classification

ARCHITECTURE

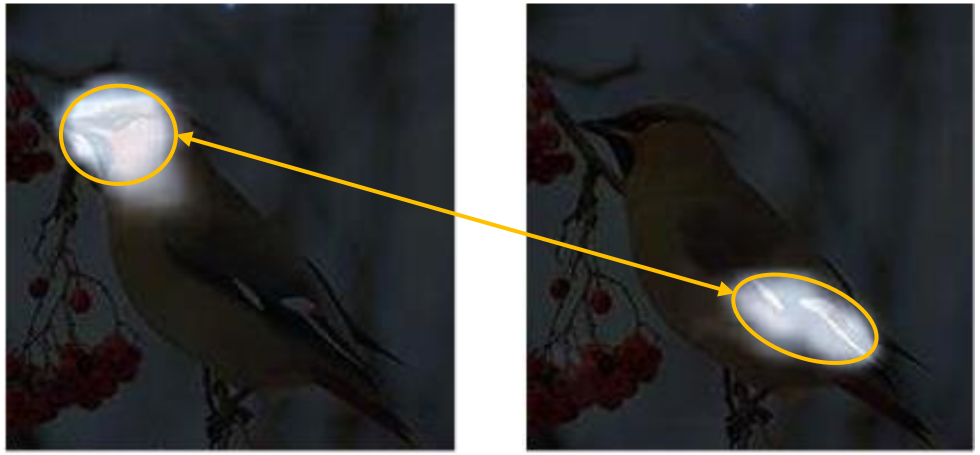


ARCHITECTURE





$$Dis(M_i(X))=\sum_{(x,y)\in M_i(X)} m_i(x,y)\,[||x-t_x||^2+||y-t_y||^2]$$



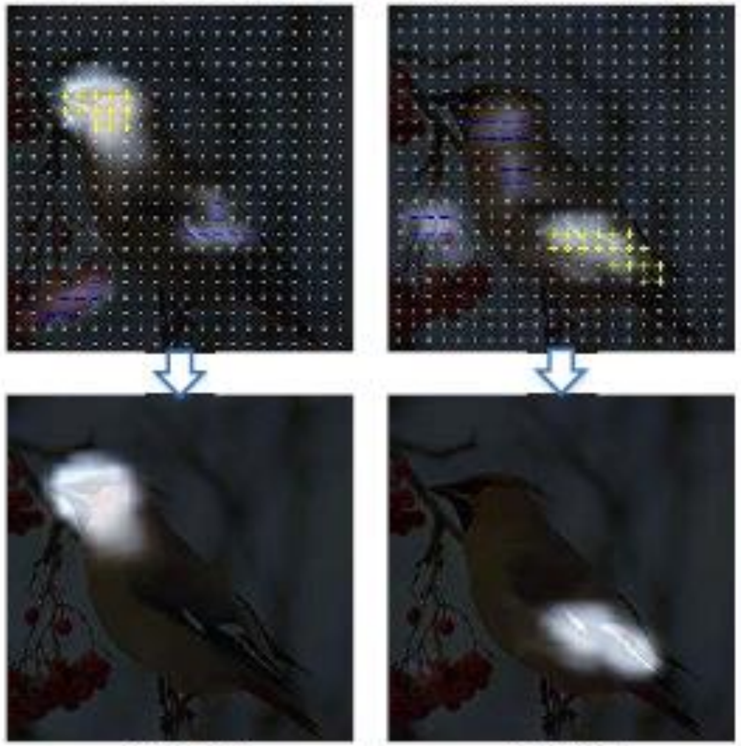
$$Div(M_i(X)) = \sum_{(x,y)\in M_i(X)} m_i(x,y)[\max_{k\neq i} m_k(x,y) - mrg]$$

OBJECTIVE FUNCTION

- $L(X) = \boxed{\sum_{i=1}^{4}\left[L_{cls}\left(Y^{(i)}, Y^{*}\right)\right]} + L_{cng}(M_i(X))$

4 parts classification loss(cross-entropy)

- $L_{cng}(M_i(X)) = Dis(M_i(X)) + \lambda Div(M_i(X))$
- $Dis(M_i(X))=\sum_{(x,y)\in M_i(X)} m_i(x,y)\ [||x-t_x||^2 + ||y-t_y||^2]$
- $Div(M_i(X)) = \sum_{(x,y)\in M_i(X)} m_i(x,y)[\max_{k\neq i} m_k(x,y) - mrg]$



(a) head

(b) wing

$Y^{(i)}$ ：predict label vector

Y^* ：ground truth label vector

L_{cng} ：channel grouping loss

$m_i(x,y)$ ：the value of $M_i(X)$ at (x, y)

mrg ：a margin

REFERENCE

[1] Very deep convolutional networks for large-scale image recognition.

[2] Caltech-UCSD Birds 200.

[3] Bird species categorization using pose normalized deep convolutional nets.

[4] The Application of Two-level Attention Models in Deep Convolutional Neural Network for Fine-grained Image Classification.