ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Rotation Invariant Aerial Image Retrieval with Group Convolutional Metric Learning

Hyunseung Chung, Woo-Jeoung Nam, Seong-Whan Lee
Pattern Recognition & Machine Learning Laboratory

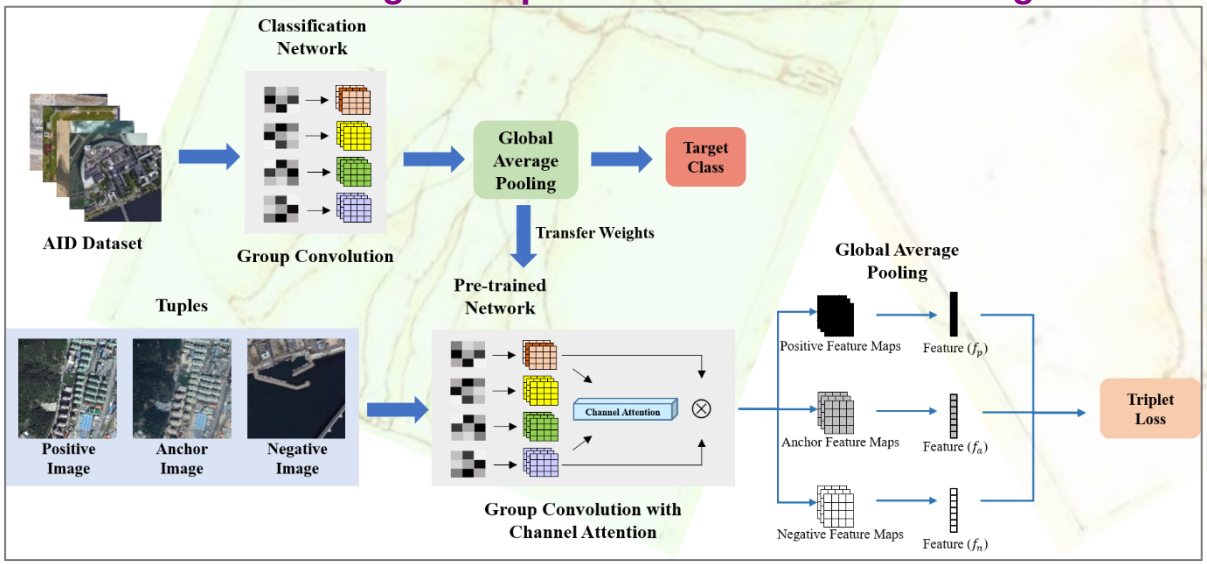
Introduction

- **Goal**
  - Developing a framework to retrieve aerial images with rotational variations
- **Motivation**
  - Rotated aerial image retrieval is highly complex task
    - Contains small objects and buildings with variations in scale, shooting range, and height
  - Robust retrieval network for rotated aerial images in demand
    - Variations in camera viewpoint depending on aircraft

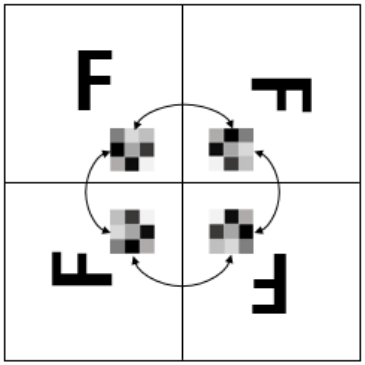
Example images from AID, NWPU-RESISC45, and Google Earth South Korea dataset

Methods

- **Network modules**
  - **Group convolutional neural network (G-CNN)**
    - **Utilize rotated filters to pretrain the network for classification task**
      - **Aerial Image Dataset (AID)**
    - **Fine-tune the pre-trained network with attentive G-CNN and metric learning**
  - **Channel attention module**
    - **Emphasize the important features among G-CNN layers**
  - **Deep metric learning**
    - **Integrate triplet loss function for training with anchor, positive, and negative**



Overall architecture of proposed method



Group convolution filters

Experimental Results (1/3)

- **Performance evaluation on rotated Google Earth South Korea dataset**
  - **Recall@ n**
    - **The percentage of correctly retrieved query up to different n values**

$$Recall\ @\ n = \frac{1}{k}\sum_{i=1}^{k}\frac{|Ret_i^n \cap Rel_i|}{|Rel_i|}$$

- Ret^n : Top-n retrieved items
- Rel : Total relevant items
- k : Number of query items

(†) indicates additional rotation training

(*) indicates added channel attention

All methods are pre-trained in AID dataset

Methods	*Recall @ n (%)*			
	$n = 1$	$n = 5$	$n = 10$	$n = 100$
R-MAC descriptor†	6.5	14.5	24.8	64.0
NetVlad†	7.4	17.4	25.1	68.2
Contrastive loss†	8.1	16.8	24.5	65.0
Triplet loss†	6.9	18.0	24.7	65.6
LDCNN†	8.9	18.4	24.7	66.6
G-CNN (p4m) + Cont. loss	17.8	32.4	38.9	72.0
G-CNN (p4m) + Cont. *	18.5	32.8	39.8	72.0
G-CNN (p4) + Triplet loss	20.1	36.0	43.2	76.9
G-CNN (p4) + Triplet. *	21.2	36.4	43.9	77.2
G-CNN (p4m) + Triplet loss	23.4	44.0	51.7	84.6
G-CNN (p4m) + Triplet loss *	**24.5**	**46.9**	**52.8**	**86.3**

Experimental Results (2/3)

- **Performance evaluation on rotated NWPU-RESISC45 dataset**

(†) indicates additional rotation training

(*) indicates added channel attention

All methods are pre-trained in AID dataset

Methods	*Recall @ n (%)*			
	$n = 1$	$n = 2$	$n = 4$	$n = 8$
R-MAC descriptor†	23.1	41.4	62.4	83.3
Contrastive loss†	26.6	46.1	66.7	86.4
NetVlad†	27.1	45.3	66.7	85.5
Triplet loss†	35.3	53.1	73.2	88.7
LDCNN†	36.0	54.5	74.0	88.8
G-CNN (p4m) + Cont. loss	39.9	56.6	74.9	88.3
G-CNN (p4m) + Cont. *	41.1	59.8	76.9	89.3
G-CNN (p4) + Triplet loss	43.1	60.8	77.6	89.5
G-CNN (p4) + Triplet. *	43.4	61.3	77.7	89.2
G-CNN (p4m) + Triplet loss	44.8	62.1	77.8	89.5
G-CNN (p4m) + Triplet loss *	**45.7**	**64.3**	**80.2**	89.5

Experimental Results (3/3)

- Examples of retrieval results in the rotated Google Earth South Korea dataset