**Introduction**

- **Goal**
  - Developing a framework to retrieve aerial images with rotational variations
  - Merger group convolution with attention mechanism and metric learning

- **Motivation**
  - Retrieving rotated aerial images is highly complex
  - Contains small objects and buildings with variations
  - Robust retrieval framework for rotated aerial images in demand

- **Challenges**
  - Viewpoint changes from aircraft with an onboard camera
  - Large variations in rotation, angle, and scale
  - Difficult to extract features from or compare similarities to each other
  - Heavy computation cost due to large size and complexity of aerial images

- **Related Works**
  - Group equivariant convolutional networks [Cohen et al., 2016]
  - Extract features from rotated filters
  - Convolutional block attention module [Woo et al., 2018]
  - Focuses on critical regions given an image
  - Deep metric learning using triplet network [Hoffer et al., 2015]
  - Considers distance between three tuples

**Methods**

- **Group convolutional neural network**
  - Utilizing rotated filters to pretrain the network for classification task
  - Similar number of parameters compared to CNN
  - Input image is convoluted with different rotated filters
  - Fine-tuning network with attentive G-CNN and metric learning

- **Deep metric learning**
  - Transforming convoluted features maps into features in embedding space
  - Integrating triplet loss function to train data tuples
  - Anchor image is the target ground truth image
  - Positive image denotes the same location image but with time variation
  - Negative image is a completely different region and time image
  - Minimizing the relation distance between the anchor and positive tuples
  - Maximizing the distance between the anchor and negative tuples

- **Channel attention module**
  - Emphasizing the important feature maps among layers
  - Refining feature maps with spatial transformation information after passing G-CNN
  - Considering inter-channel relations
  - Focusing on critical regions given an input image
  - Improving retrieval performance compared to the baseline G-CNN

**Experiments**

- **Quantitative results**
  - Evaluation metric: Recall@n
  - Recall@n is the percentage of correctly retrieved queries within top n retrieved database images

- **Class activation mapping results**
  - Large variations in rotation, angle, and scale
  - Difficult to extract features from or compare similarities to each other
  - Heavy computation cost due to large size and complexity of aerial images