# **Directed Variational Cross-encoder Network for Few-shot Multi-image Co-segmentation**

Sayan Banerjee, S Divakar Bhat, Subhasis Chaudhuri, Rajbabu Velmurugan Indian Institute of Technology Bombay

> ICPR 2020 Milan, Italy

# Directed Variational Cross-encoder Network for Few-shot Multi-image Co-segmentation

#### **Input Set of Images**

coseg set

quide set





#### **Traditional Supervised Co-segmentation** Approach



#### **Proposed Few-shot Co-segmentation**

**Approach** 



# **Directed Variational Cross-encoder Network for Few-shot Multi-image Co-segmentation**



Given a guide set  $\mathcal{G} = \{\mathcal{P}^g \cup \mathcal{N}^g\},\$ containing a small number of images let's say *n* with common foreground  $\{\mathcal{P}^g\}$ , and random number of *outliers*  $\{\mathcal{N}^g\}$ ,

from a large corpus of images named as *co-seg* set  $\mathcal{C} = \{\mathcal{P}^c \cup \mathcal{N}^c\},\$ 

- determine images which has foreground specified by the guide set.
- and extracts that foreground from each image.

\*\*Please note that red boxes are outliers

# **Guided Co-segmentation of Multiple Images using Few-shot Learning**

The dataset for such application let's say  $\mathcal{D}_{target}$  which has very less training samples

# direct training leads to **overfitting**

### **Class Agnostic Episodic Training Scheme:**

Find any larger dataset  $\mathcal{D}_{base}$  developed for co-segmentation even if  $\mathcal{D}_{target} \cap \mathcal{D}_{base} = \phi$ Randomly sample a episode consists of a guide set  $\mathcal{G} = \{\mathcal{P}^g \cup \mathcal{N}^g\} = \{(x_1^g, y_1^g), \dots, (x_k^g, y_k^g)\}$  containing  $n \text{ positive samples } \{\mathcal{P}^g\}$  and  $k\text{-}n \text{ outliers } \{\mathcal{N}^g\}$  and a co-seg set  $\mathcal{C} = \{(x_1^c, y_1^c), \dots, (x_m^c, y_m^c)\}$ Calculate common object prototype from the set  $\mathcal{G}$  $\forall x_i^c \in \mathcal{C}$  calculate individual feature similarity with the common object prototype and predict mask  $\hat{y}_i^c$ 

Calculate Loss, 
$$-\sum_{j=1}^{m} \sum_{(a,b)} y_j^c \log P(\hat{y}_j^c(a,b)) + \lambda \times reg$$
 and backpropagate

5

 $\prec$ 

## **Directed Variational Cross-encoder Network**



### **Directed Variational Cross-encoder Network**



### **DVICE: Directed Variational Inference Cross Encoder**



# **Experimental Results and Ablation Studies**

co-seg set

**(b)** 

with variational

inference and

attention

variational

inference and attention

without

8

