Coarse to Fine: Progressive and Multi-Task Learning for Salient Object Detection
Donggoo Kang, Sangwoo Park, and Joonki Paik
Department of image, Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University

1. Multi-Task Learning

- Encoder learns abstract representations that explanatory factors of the data and compress the important features as layers get deeper.
- Decoder reconstructs the feature extracted from the encoder back to the original image size.

2. Progressive Learning

- Progressive learning scheme progressively grows decoder in the training phase. In other words, it starts from easier low-resolution layers, and adds new higher-resolution layers.
- This can weakly localize candidates in row-resolution and improve stability in high-resolution layers.

Proposed Method

1. Overall Architecture

- It consists of one weight shared encoder for feature embedding and two task specific decoders of the same structure
- Learning progresses gradually from phase 1 to final phase.

2. Progressive Decoding Block

- Concatenate feature map of encoder to decoding layer to preserve encoder information. After learning the current phase, remove the red arrow layer and add another phase block. Layers that inside the blue dotted line are transfer layers to the next phase.

3. Cost Function

\[ L_i(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} (y_i \log (\hat{y}_i) + (1 - y_i) \log (1 - \hat{y}_i)), \] (1)

where \( y \) is the pixel value of ground truth image and \( \hat{y} \) is the pixel value of the network output. \( i \in \{0,1, ..., N\} \). \( N \) is the number of pixel of input image. The final loss of each training phase is the sum of cross-entropy loss of two branches

\[ L_i^{\text{final}} = L_i(S_i C_i E_1 D_1) + L_i(C_i E_1 D_2), \] (2)

Where \( S \) and \( C \) represent the saliency map and contour ground truth, respectively. \( C_i E_1 D_1 \) and \( C_i E_1 D_2 \) represent output of each branch. \( i \in \{0, ..., M\} \). \( M \) is the number of training phase.

Conclusion

- We propose a novel method that combines a progressive learning scheme with a multi-task learning scheme.
- Compared with existing state-of-the-art methods that attempted to capture the context of an object through modification of the convolution block, the proposed method better captured the context of the object only by the manipulation of the learning scheme.