



LEARNING SALIENT OBJECTS IN A SCENE USING SUPERPIXEL-AUGMENTED CONVOLUTIONAL NEURAL NETWORK

PROBLEM OVERVIEW



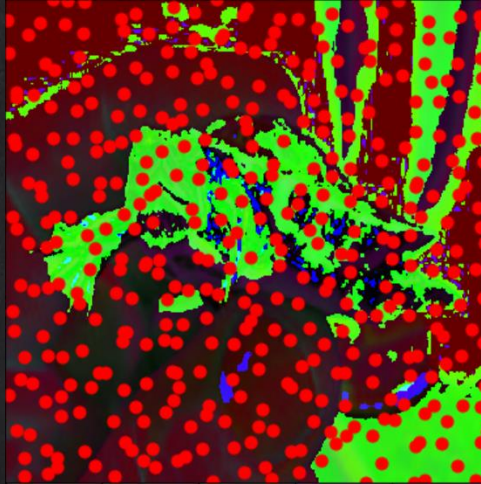
GIVEN A NATURAL SCENE IMAGE, OBJECTIVE IS TO DETECT “SALIENT” REGIONS.

METHOD OVERVIEW



1. SUPERPIXELS ARE GENERATED USING SLIC SEGMENTATION ALGORITHM
2. SALIENCY DETECTION SPECIFIC HAND-CRAFTED REPRESENTATIONS ARE OBTAINED
3. PER SUPERPIXEL BINARY CLASSIFICATION IS DONE USING SHALLOW CNN

SLIC SUPERPIXEL GENERATION



$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2}$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

$$D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2}$$

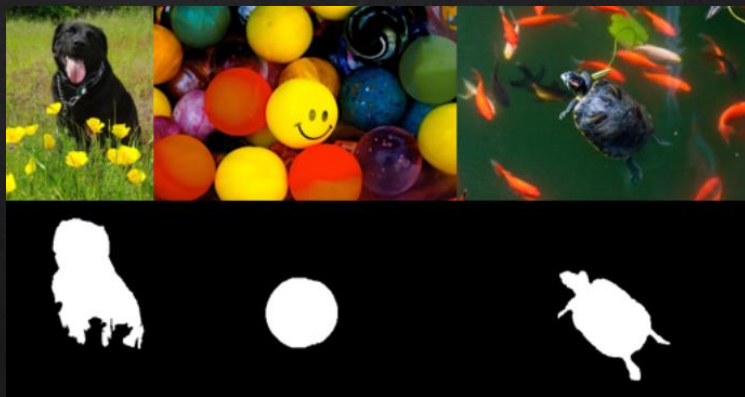
SUPERPIXEL REPRESENTATIONS

$$Q = \begin{bmatrix} q_{11}^c & \dots & q_{1j}^c & \dots & q_{1M}^c \\ \vdots & \ddots & & \ddots & \\ q_{x1}^c & \dots & q_{xj}^c & \dots & q_{xM}^c \\ \vdots & \ddots & & \ddots & \\ q_{N1}^c & \dots & q_{Nj}^c & \dots & q_{NM}^c \end{bmatrix}$$

$$q_{xj}^c = t(r_j) \cdot |C(r_x) - C(r_j)| \cdot w(P(r_x), P(r_j))$$

$$Q' = \begin{bmatrix} q_{11}^d & \dots & q_{1j}^d & \dots & q_{1M}^d \\ \vdots & \ddots & & \ddots & \\ q_{x1}^d & \dots & q_{xj}^d & \dots & q_{xM}^d \\ \vdots & \ddots & & \ddots & \\ q_{N1}^d & \dots & q_{Nj}^d & \dots & q_{NM}^d \end{bmatrix}$$

$$q_{xj}^d = t(r_j) \cdot |P(r_x) - P(r_j)| \cdot w(C(r_x), C(r_j))$$



CLASSIFICATION USING SHALLOW CNN

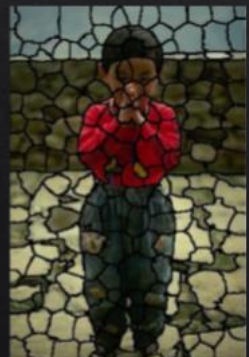
Type	Size/Stride	Number of channels
input	M	3
Conv1D, tanh	10×1	5
MaxPool	$2 \times 1/2 \times 1$	-
Conv1D, tanh	20×1	10
Maxpool	$2 \times 1/2 \times 1$	-
Conv1D, tanh	20×1	20
Maxpool	$2 \times 1/2 \times 1$	-
Global-avg-pool	-	-
Fully-con, tanh	20	-
Dropout (0.3)	-	-
Fully-con, Softmax	2	-

Table 1: CNN architecture used in this project. Note: size of Q matrix is $M \times M$.

TOTAL NUMBER OF PARAMETERS: 5,233

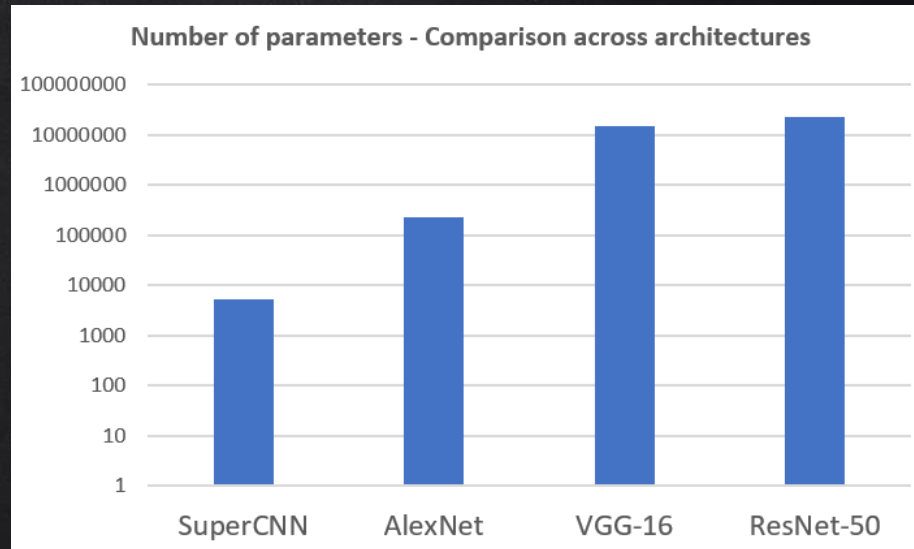
TOTAL NUMBER OF PARAMETERS IN CONV LAYERS OF VGG-16: 14,714,688

RESULTS



CONCLUSIONS

1. HAND-CRAFTED FEATURES CAPTURING GLOBAL SPATIAL CONTEXT AT SUPERPIXEL LEVEL ALONG WITH CNN CLASSIFIER SHOWS DECENT PERFORMANCE FOR SALIENCY DETECTION IN IMAGES.
2. SUPERPIXEL-CNN BASED APPROACH DOES NOT REQUIRE LARGE TRAINING DATA AND NUMBER OF PARAMETERS IN CNN ARE IN FEW THOUSANDS COMPARED TO 10s OF MILLIONS OF PARAMETERS IN STANDARD CNNs ON RAW IMAGE PIXELS.
3. LOWER NUMBER OF PARAMETERS REDUCES THE COMPUTATIONAL COST SIGNIFICANTLY AT BOTH TRAINING AND TEST TIME.





THANKS!

Shashank Tripathi
Yash Patel

shashanktripathi123@github
yash0307@github

Code is available here: <https://github.com/yash0307/SuperCNN>