Discovery of Image Pixels highly Contributing to CNN regression

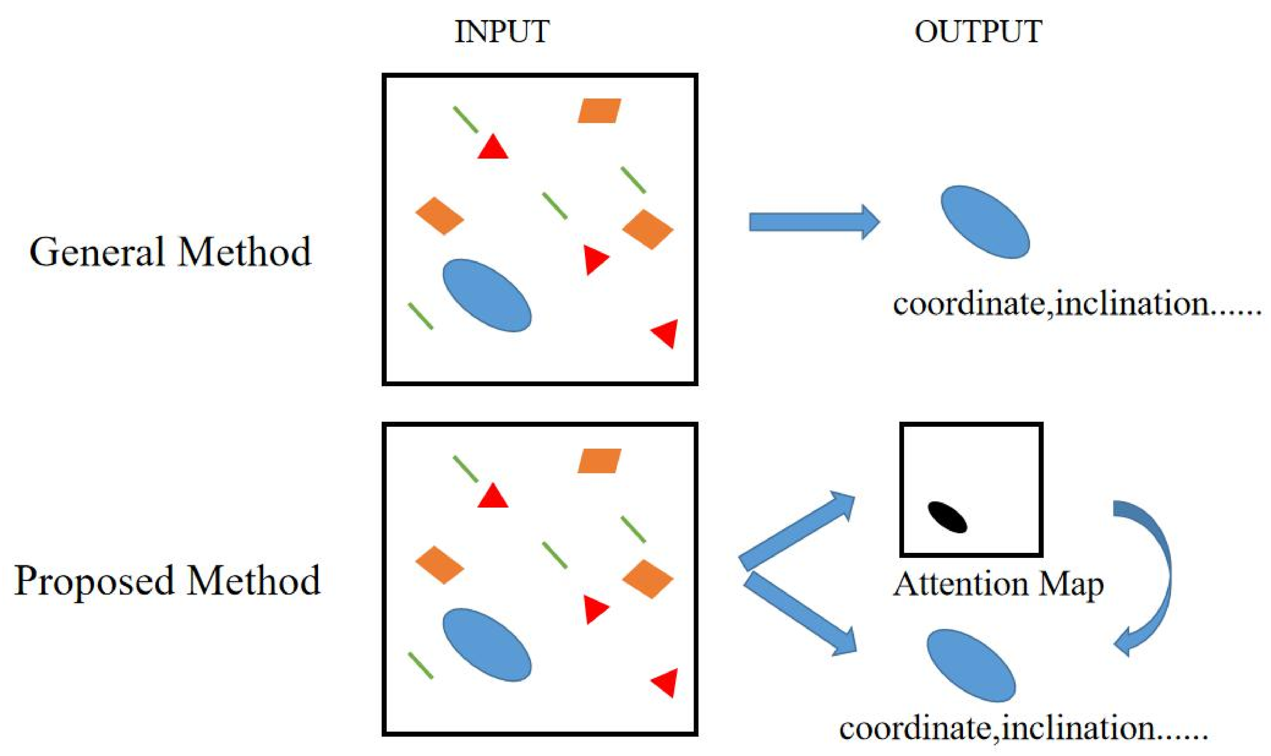
Yitian Li, Nobutaka Shimada

Ritsumei University,Shiga,Japan

Introduction

When we want to output some information from an image, we only ***focus on a part of the image or a specific object***.But,there is always some noise elements in the image,such as other object or the background,they can affect the output result.

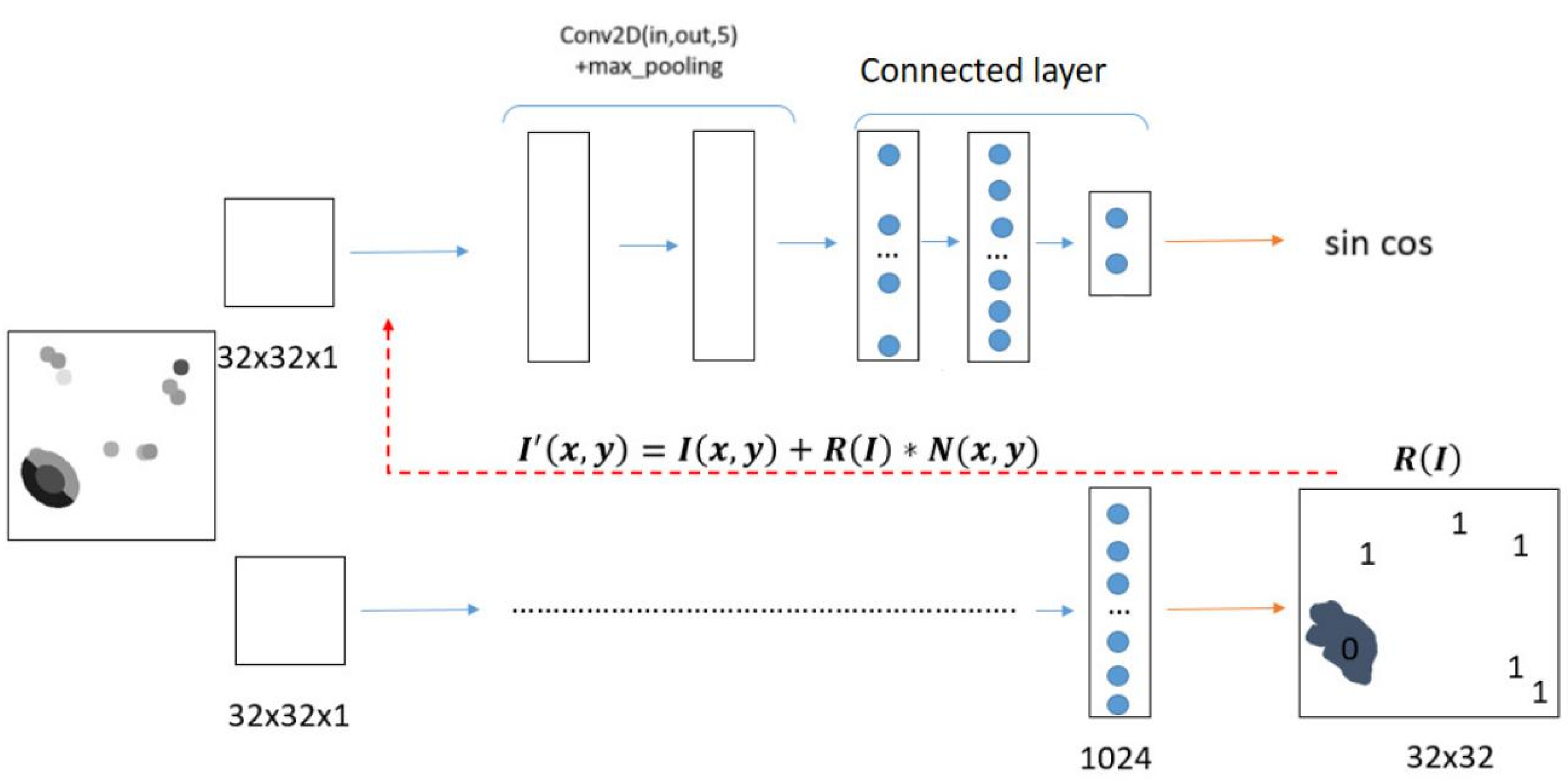
This study focuses on such problem and aims to ***visualize the pixel region*** that contributes to the CNN regression by ***generating an attention map*** that can represent where the object area exists in the image.



Methods

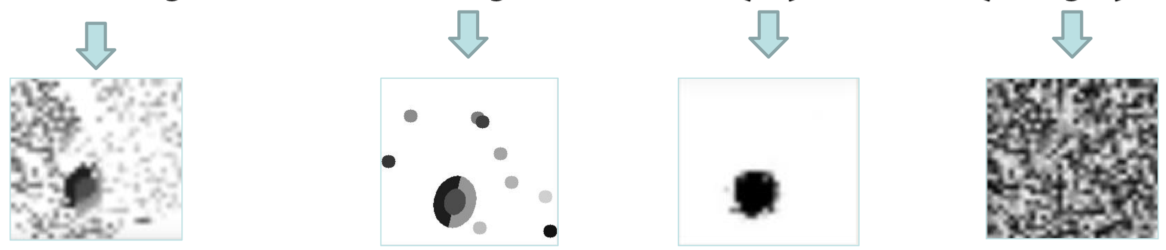
In this study, we will construct a network, which has two parts.

- Above part: normal supervised learning CNN network
  Input: Image I (ellipse+noise) (size 32*32)
  Output: corresponding ***Parameters*** (inclination of the ellipse)
- Below part:unsupervised CNN learning network
  Input: Image I (ellipse+noise) (size 32*32)
  Output:***Attention Map*** R(I) (mask: ellipse)



How to make I' ?

$$I'(x,y) = I(x,y) + R(I) * N(x,y)$$



LOSS FUNCTION

$$\mathrm{L}(i) = a\,|F(I_i) - P(I_i)|^2 + b\,|F(I_i') - P(I_i)|^2 + c\,|Area(R(I_i)) - S_R|^2$$

$$L_{total} = \sum_i L(i)$$

$$Area(R(I)) = \frac{\int R(I)dxdy}{M*N}$$

The ratio $1\text{-}S_R$ of the object area is limited using the integral value of the entire weighted image R(I). The area of the ellipse is about 0.05 and the intergral valueof S_R is about 0.95.

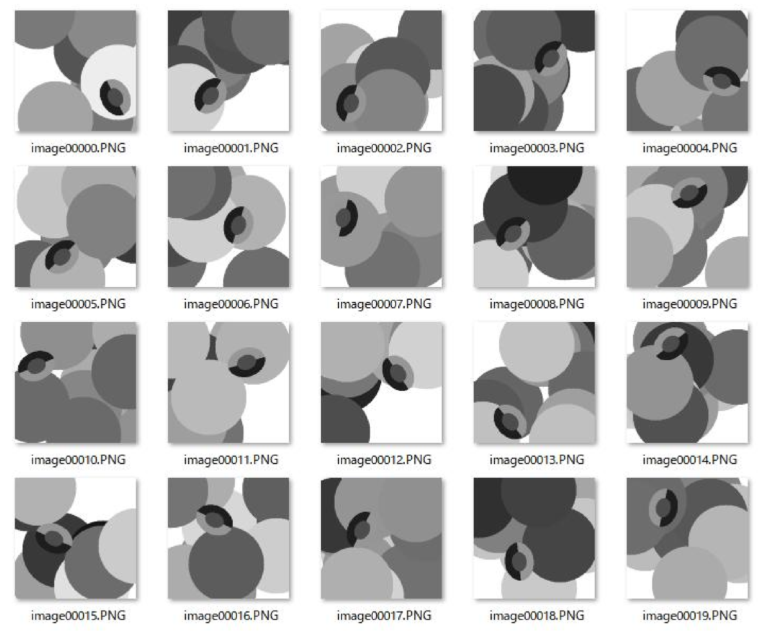
The total number of pixels of images are M*N(32 × 32 = 1024).

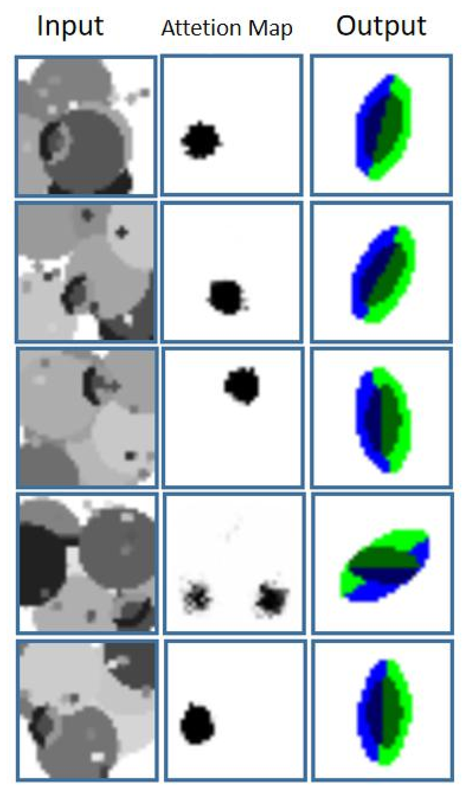
Results

In this section, we will test our network in three experiments.

A. Dataset1(ellipse+cirlcle noise)

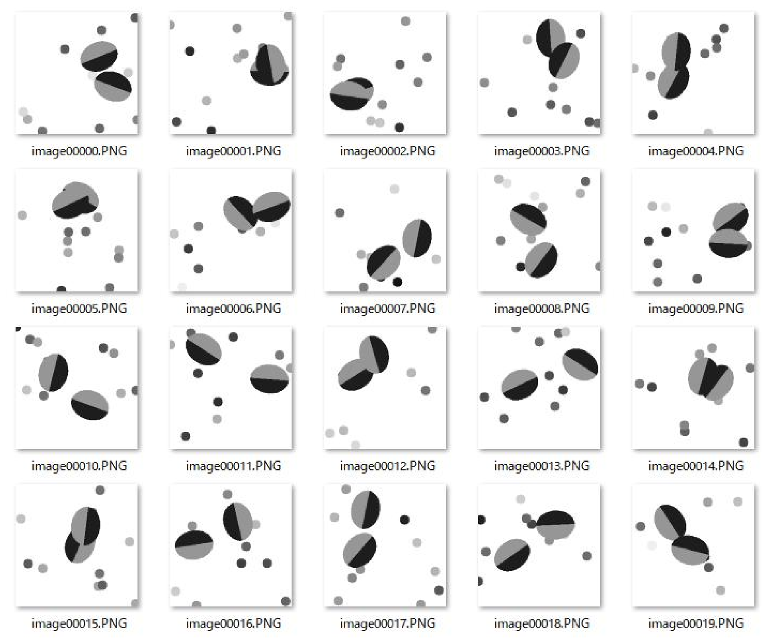
Purpose:extract the region where the ellipse exists;get the inclination of the ellipse.
Training: 8000 images
Test : 3000 images (average loss is 13[degree])
Attetion Map: grayscale image with a value range of 0-255
Dark region:ellipse region
Output: large ellipse - the supervisory signal
small ellipse - estimated inclination

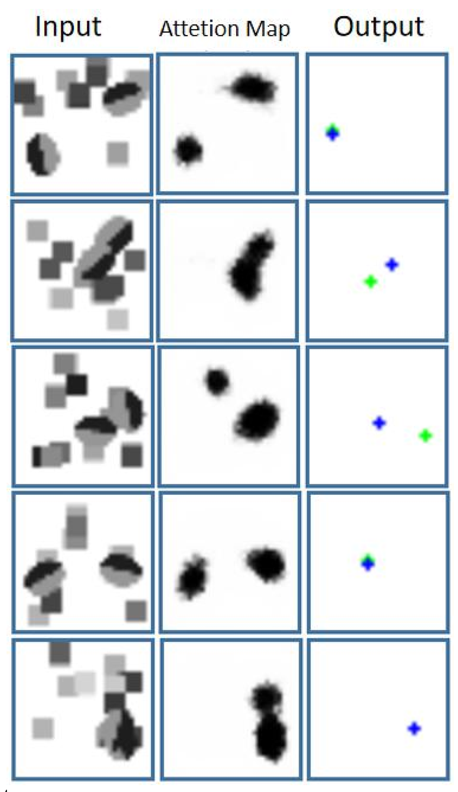




B. Dataset2(two ellipses+point noise)

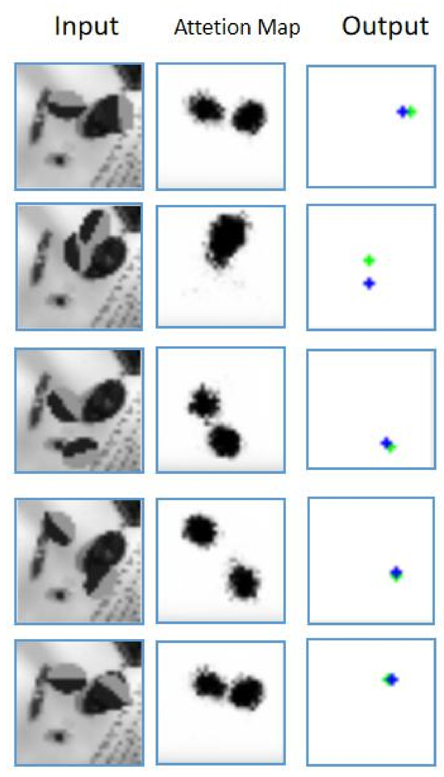
Purpose:extract the region where the ellipse exists
get the intersection coordinate of major axis of two ellipses
Training: 8000 images
Test : 3000 images (average loss is 2.5[pixel])
Attetion Map: grayscale image with a value range of 0-255
Dark region:ellipse region
Output: green mark - the supervisory signal
blue mark - estimated inclination





C. Dataset3(two ellipses+real object noise)

Purpose:extract the region where the ellipse exists
get the intersection coordinate of major axis of two ellipses
Training: 3000 images
Test : 1000 images (average loss is 3.28[pixel])
Attetion Map: grayscale image with a value range of 0-255
Dark region:ellipse region
Output: green mark - the supervisory signal
blue mark - estimated inclination



Conclusions

We proposed a method by using the generated attention map to represent where the object area exists in the image.

For future work, we will focus on the practial application of this network in real life, rather than using ellipse simulation.