NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

Smart Object Counter
for Highly Congested Scenes Using Deep Learning

Student: Kang Xinhui   Supervisor: Dr Loke Yuan Ren

Ground Truth: 59
Model output: 59.5

Ground Truth: 103
Model output: 106.9

Ground Truth: 55
Model output: 51.6

SLE

CTE

Land Transport Authority Live Traffic   Orchard Road Pedestrian   Microscopic cell
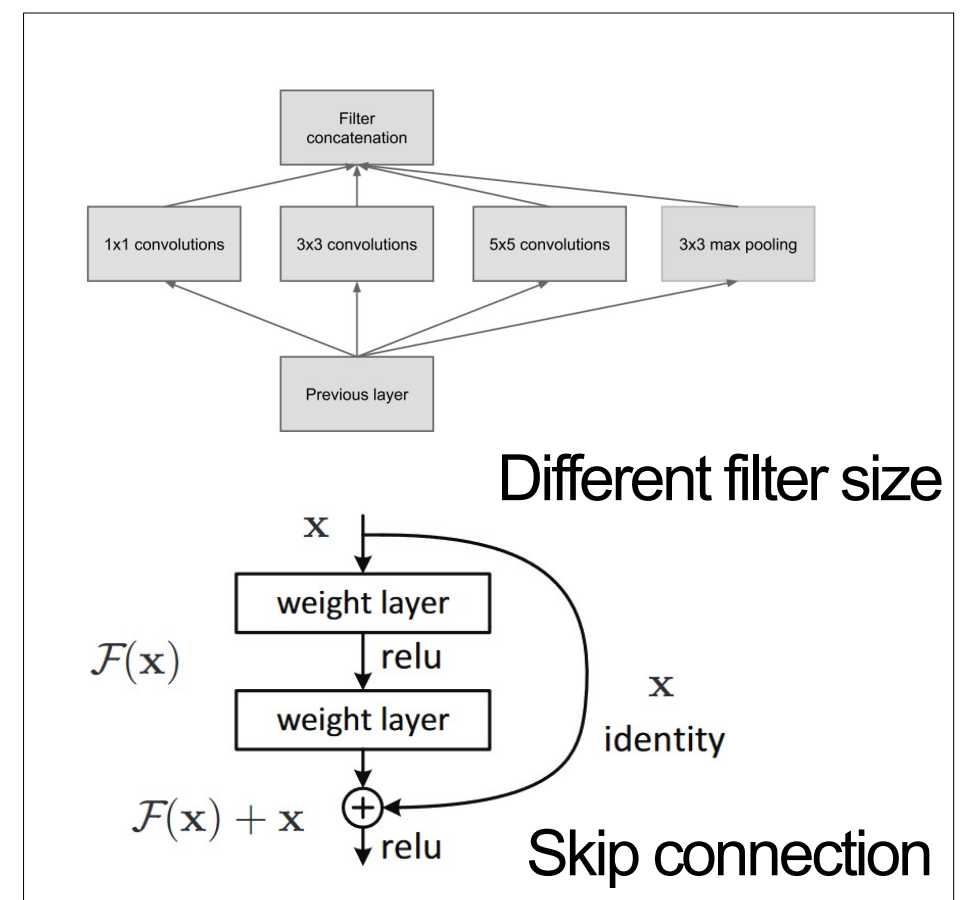
Project Objectives:

This project aims to deploy density-estimation-based approach on convoluted neural network to develop efficient, accurate and robust deep learning model. The effectiveness of Inception_v4 and Inception-Resnet-v1 building blocks are assessed and 3 new models are proposed. The **Inception** building blocks consists of different kernel sizes and can extract multi-scale information, and the skip connection in the **ResNet** design can alleviate the gradient varnishing issue. The combination of the two allows the model to be both wider and deeper, and hence able to recognise more complex features. This project proposed 3 model architectures, which are proven to be robust across different use cases in **crowd counting**, **vehicle counting** and **microscopic cell counting**.

Dataset	Best MAE	Best RMSE
ShanghaiTech A (Crowd)	57.6	99.2
ShanghaiTech B (Crowd)	7.8	13.8
TRANCOS (Vehicle)	1.5	2.0
Malaria Cell	2.7	3.9

Targeted Problems:

- Overlap and occlusion
- Uneven distribution
- Object scale variation
- Viewpoint variation
- Scene geometry variation
- Background segmentation

Filter concatenation

1x1 convolutions   3x3 convolutions   5x5 convolutions   3x3 max pooling

Previous layer

Different filter size

$\mathbf{x}$

weight layer
relu
weight layer

$\mathcal{F}(\mathbf{x})$

$\mathbf{x}$
identity

$\mathcal{F}(\mathbf{x}) + \mathbf{x}$ relu

Skip connection

Proposed Architectures

Reduced to 3 in the shallow version
10   5   2
      20   10

2 max pooling layers removed

Input

Add padding=1

Stem of Inception-Resnet

Inception-Resnet-A   Inception-Resnet-B   Inception-Resnet-C   Un-sampling
Reduction A   Reduction B

Add padding=1

Stem of Inception-v4

Filter concat
3x3 Conv (192 V)   MaxPool (stride=2 V)
Filter concat
3x3 Conv (96 V)
3x3 Conv (96 V)   1x7 Conv (64)
1x1 Conv (64)   7x1 Conv (64)
1x1 Conv (64)

Add padding = 1

Filter concat
3x3 MaxPool (stride 2 V)   3x3 Conv (96 stride 2 V)
3x3 Conv (64)
3x3 Conv (32 V)
3x3 Conv (32 stride 2 V)
Input (299x299x3)

Input

4   7   3

Inception-A   Inception-B   Inception-C   Un-sampling

Output