

Northumbria University NEWCASTLE


WINSENSE 赢识科技

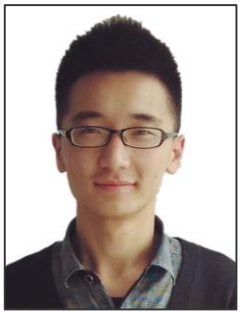
UG^2+ CHALLENGE

Selective Refinement Network for Dark Face Detection

Track 2.2

2019.06.16

Team Member



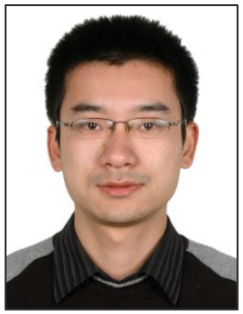
Shifeng Zhang

Ruizhe Liu

Cheng Chi

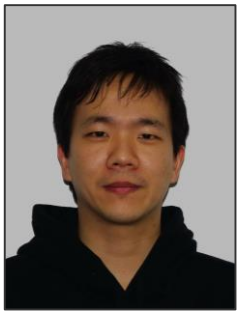
Zheming Zuo

Chubing Zhuang



Zhen Lei

Shizheng Wang



Dong Yi

Stan Z. Li

Outline

- Introduction
- Solution Pipeline
- Image Enhancement
- Face Detection
- Result

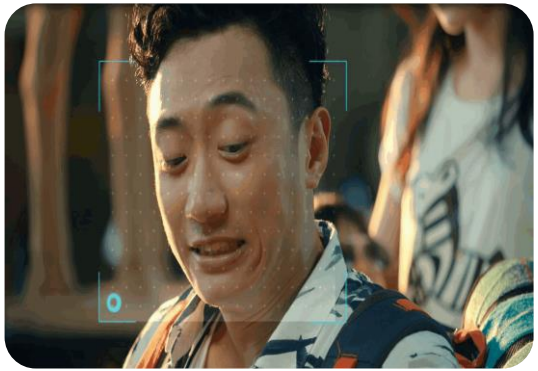
Outline

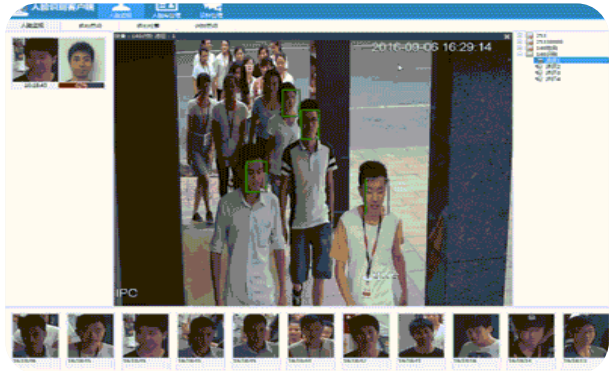
- **Introduction**
- Solution Pipeline
- Image Enhancement
- Face Detection
- Result

Introduction

- Face Detection: Determine whether there are faces on an image, and if so, give their location.

Introduction





Face Analysis

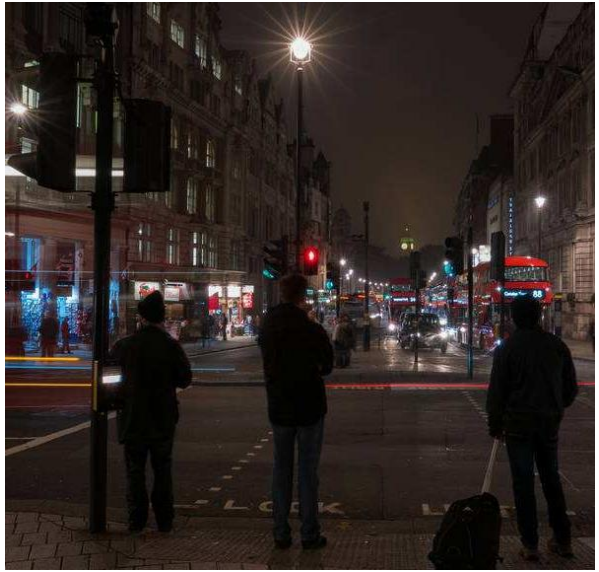
Video Surveillance

Face Unlock

Introduction

- One of challenges: detecting faces in poor visibility environments.

Haze

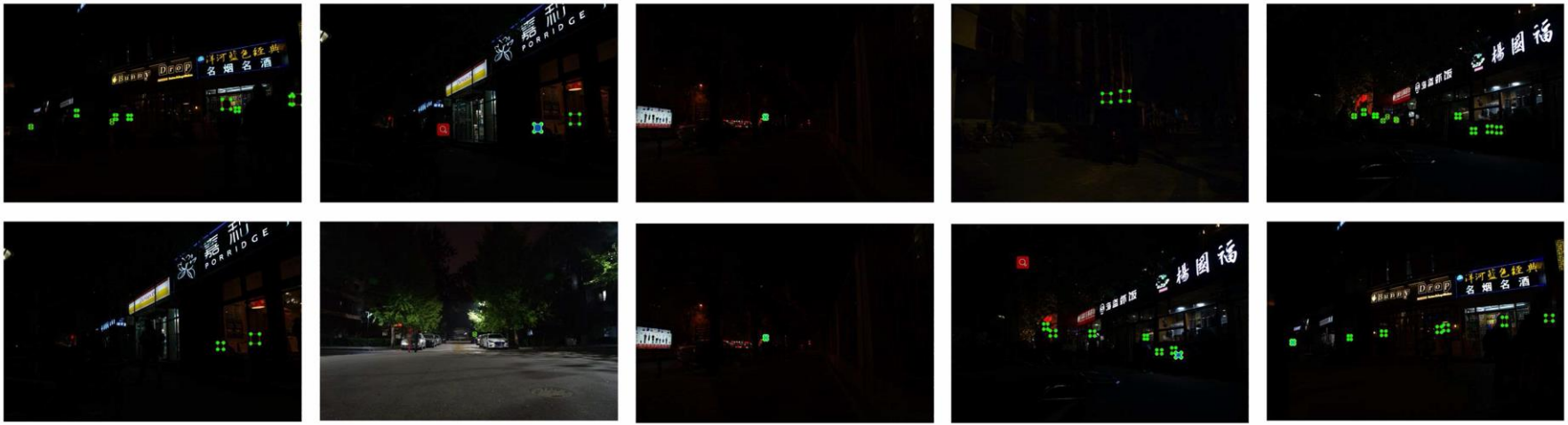


Low Light

Raindrop

Introduction

Track 2.2: (Semi-)Supervised Face Detection in Low Light Condition



- Captured during nighttime under low Light Conditions
- Training: 6,000 images with 43,849 faces
- Testing: 4,000 images with 32,571 faces

Outline

- Introduction
- **Solution Pipeline**
- Image Enhancement
- Face Detection
- Result

Solution Pipeline

1. Train **SRN-Res101**[1] on WIDER FACE dataset
2. Utilize **MSRCR**[2] to process DARK FACE dataset
3. Fine-tune pretrained **SRN-Res101** on processed DARK FACE dataset

[1] C. Chi, S. Zhang et al. Selective refinement network for high performance face detection. AAAI, 2019.
[2] S. Parthasarathy, P. Sankaran. An automated multi scale retinex with color restoration for image enhancement. NCC, 2012.

Outline

- Introduction
- Solution Pipeline
- **Image Enhancement**
- Face Detection
- Result

Image Enhancement

[3] X. Guo et al. Lime: Low-light image enhancement via illumination map estimation. TIP, 2017
[4] C. Wei et al. Deep Retinex Decomposition for Low-Light Enhancement. BMVC, 2018
[5] Z. Zuo et al. Saliency-informed spatio-temporal vector of locally aggregated descriptors and fisher vectors for visual action recognition. BMVC, 2018

Original

Lime[3]

RetinexNet[4]

Convolution

Saliency[5]

MSRCR

Image Enhancement

[3] X. Guo et al. Lime: Low-light image enhancement via illumination map estimation. TIP, 2017
[4] C. Wei et al. Deep Retinex Decomposition for Low-Light Enhancement. BMVC, 2018
[5] Z. Zuo et al. Saliency-informed spatio-temporal vector of locally aggregated descriptors and fisher vectors for visual action recognition. BMVC, 2018

Original

Lime[3]

RetinexNet[4]

Convolution

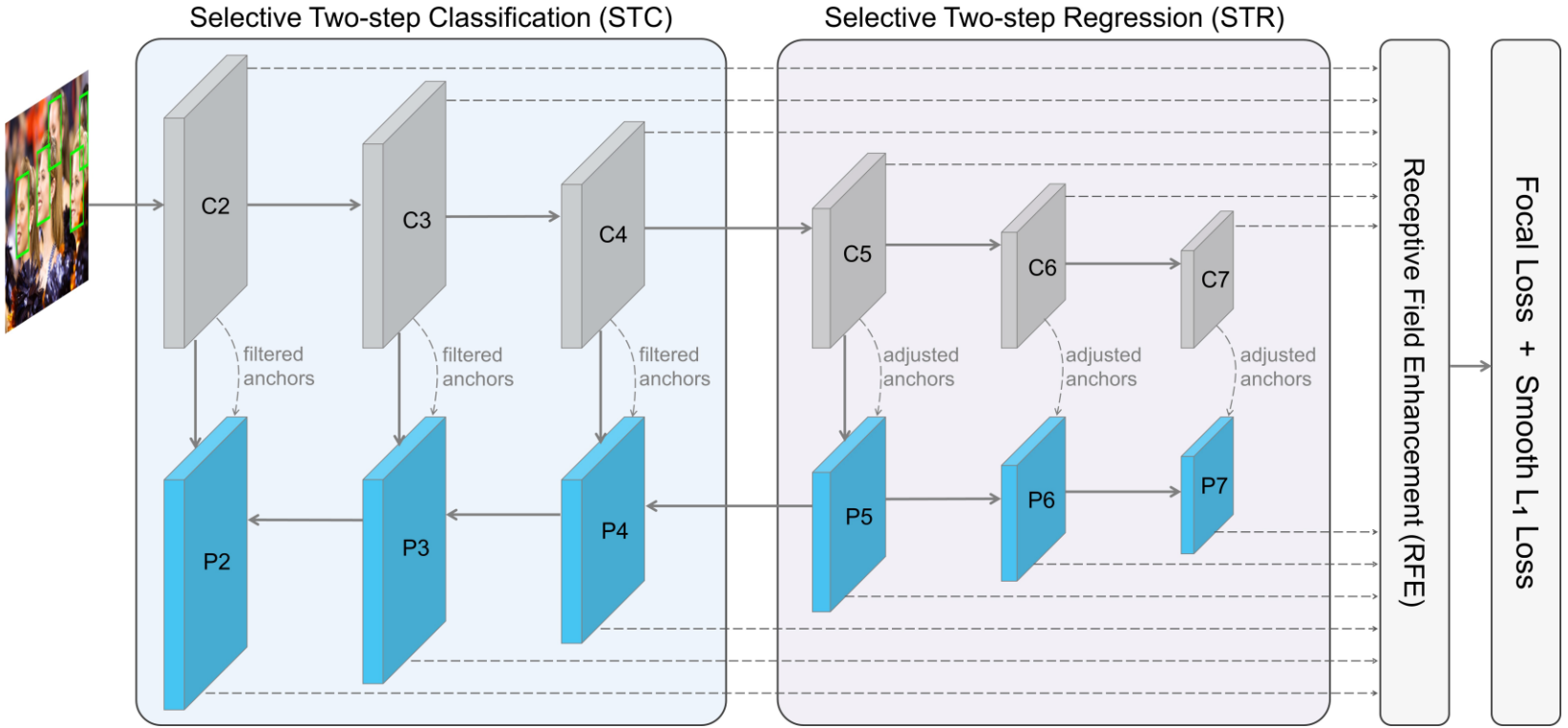
Saliency[5]

MSRCR

Outline

- Introduction
- Solution Pipeline
- Image Enhancement
- **Face Detection**
- Result

Face Detection



- Selective Refinement Network (SRN): STC、 STR、 RFE

1. Selective Two-Step Classification (STC)

- Need to tile plenty of small anchors to detect small faces
- Cause extreme class imbalance between positives and negatives
- The number of positive samples is only a few dozen or less
- Doing two-step classification is essential to reduce the false positives
- Performing two-step classification on all pyramid levels is unnecessary

STC	B	P2	P3	P4	P5	P6	P7
Easy	95.1	**95.2**	**95.2**	**95.2**	95.0	95.1	95.0
Medium	93.9	**94.2**	**94.3**	**94.1**	93.9	93.7	93.9
Hard	88.0	**88.9**	**88.7**	**88.5**	87.8	88.0	87.7

- Select P2, P3, and P4 to perform two-step classification

2. Selective Two-Step Regression (STR)

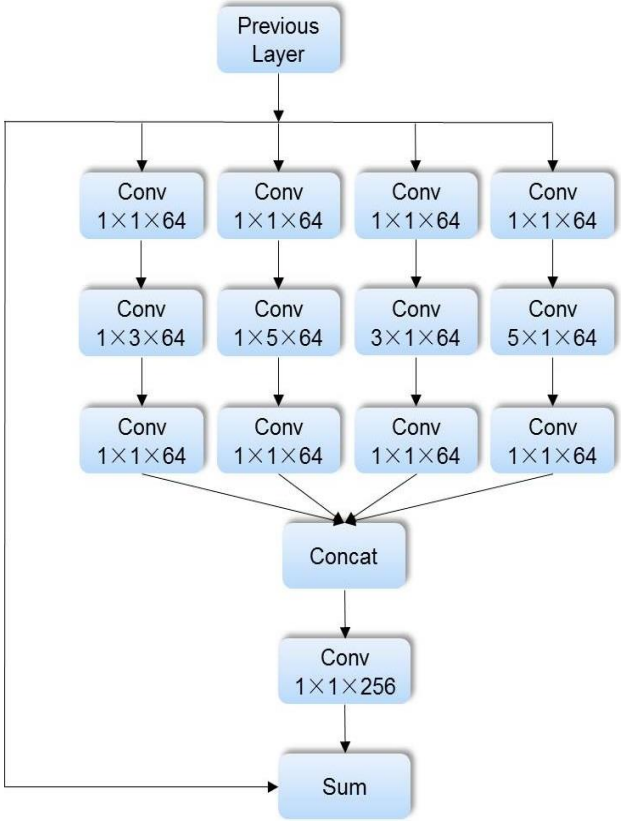
- Making the location of bounding box more accurate is a challenging issue
- Current one-stage methods rely on one-step regression
- It is inaccurate the in MS COCO evaluation metric
- Blindly adding multi-step regression is often counterproductive

STR	B	P2	P3	P4	P5	P6	P7
Easy	95.1	94.8	94.3	94.8	**95.4**	**95.7**	**95.6**
Medium	93.9	93.4	93.7	93.9	**94.2**	**94.4**	**94.6**
Hard	88.0	87.5	87.7	87.0	**88.2**	**88.2**	**88.4**

- Select P5, P6, and P6 to perform two-step regression

3. Receptive Field Enhancement (RFE)

- Current networks possess square receptive fields
- Mismatch between receptive fields and aspect ratio of faces affect the detection performance
- Propose RFE to diversify receptive fields before predicting classes and locations
- RFE replaces the middle two convolution layers in the class and box subnet of RetinaNet



Training Detail

- Backbone: ResNet-101 with 6-level FPN
- Loss: sigmoid focal loss + smooth L1 loss
- Data augmentation: color distortions, random cropping, random flipping
- Anchor design: two specific scales $(2, 2\sqrt{2})$ and one aspect ratio (1.25)
- SGD, 0.9 momentum, 0.0001 weight decay, batch size 32
- Learning rate to 0.01, 0.001 and 0.0001 for the 100, 20 and 10 epochs

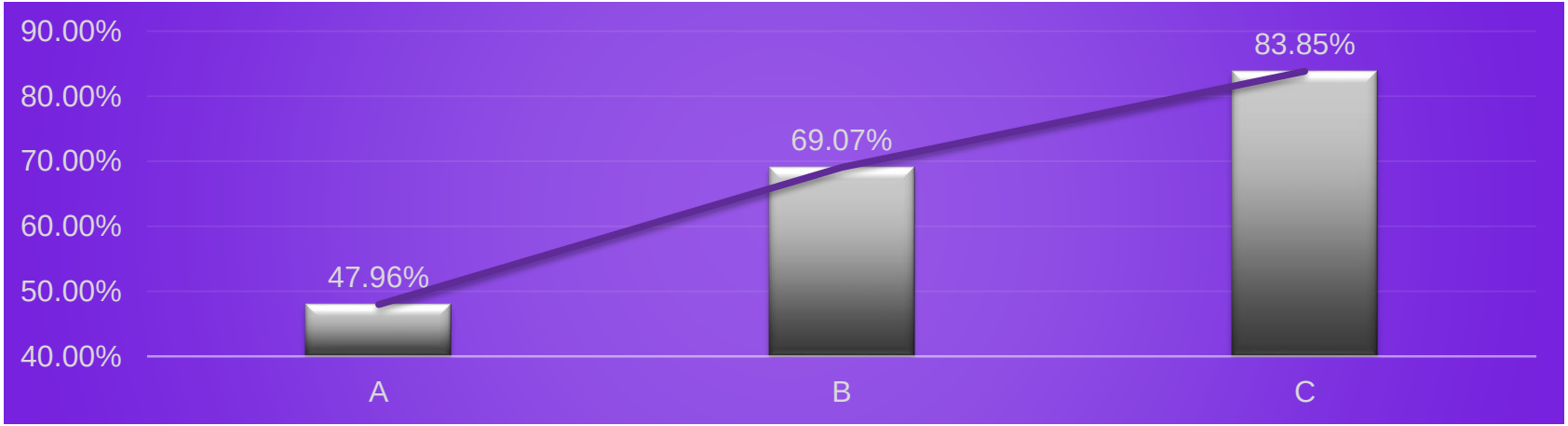
Code has been released publicly:

Outline

- Introduction
- Solution Pipeline
- Image Enhancement
- Face Detection
- **Result**

Result

Experiment	Training Dataset	Testing Dataset	AP
A	WIDER FACE	DARK FACE	47.96
B	WIDER FACE	Processed DARK FACE	69.07
C	WIDER FACE +Processed DARK FACE	Processed DARK FACE	83.85



Thank you!