Multi-Resolution Fusion and Multi-scale Input Priors Based Crowd Counting

Usman Sajid, Wenchi Ma, and Guanghui Wang
{usajid, wenchima, ghwang}@ku.edu

University of Kansas, KS, USA

CHALLENGES INVOLVED

Crowd Counting

1. Large **variation** in crowd density across images
   - Large perspective
   - Severe occlusion

2. **Cluttered** "crowd-like" background regions in images
   - SOTAs find it hard to recognize such background patterns.

CHALLENGES INVOLVED

Crowd Counting

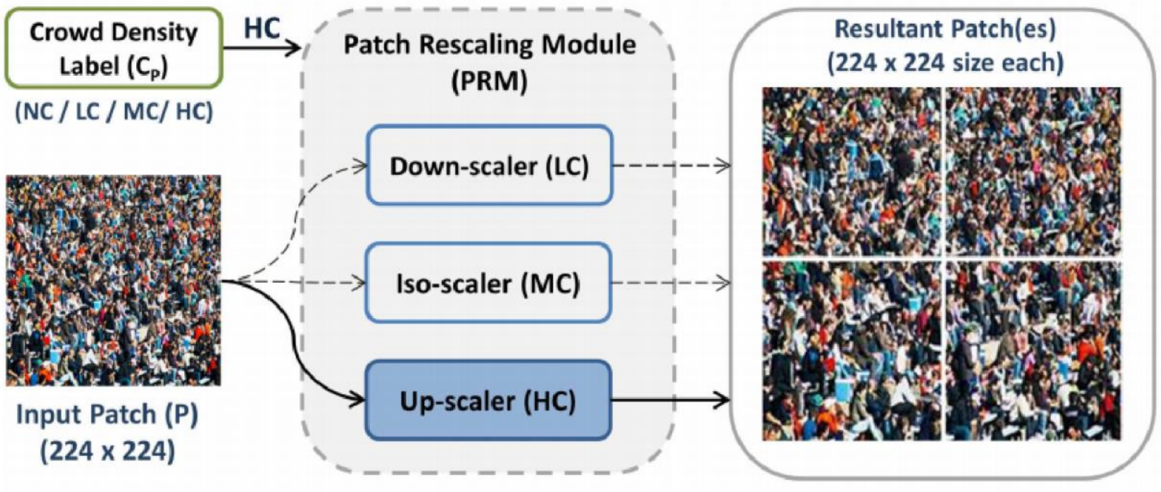
1. Recently, Sajid et al. [27], [28] observed that suitable rescaling (down-, no-, or up-scaling) of the input image or patch, according to its crowd density level (low-, medium-, or high-crowd), gives more effective results as compared to the multi-column or multi-regressor based methods. Based on this observation, they also designed a **patch rescaling module (PRM)** [28] that **rescales** the input image accordingly based on its **crowd-density class label**.

2. It has key shortcomings:

   - **Requires** the crowd-density classification label of the original input patch

   - **Selects** only one of three available rescaling operations (down-, no-, or up-scaling) for any given input patch

3. This **limits** the overall effectiveness of the PRM module and only utilize the deployed observation partially.



The PRM Module [28]

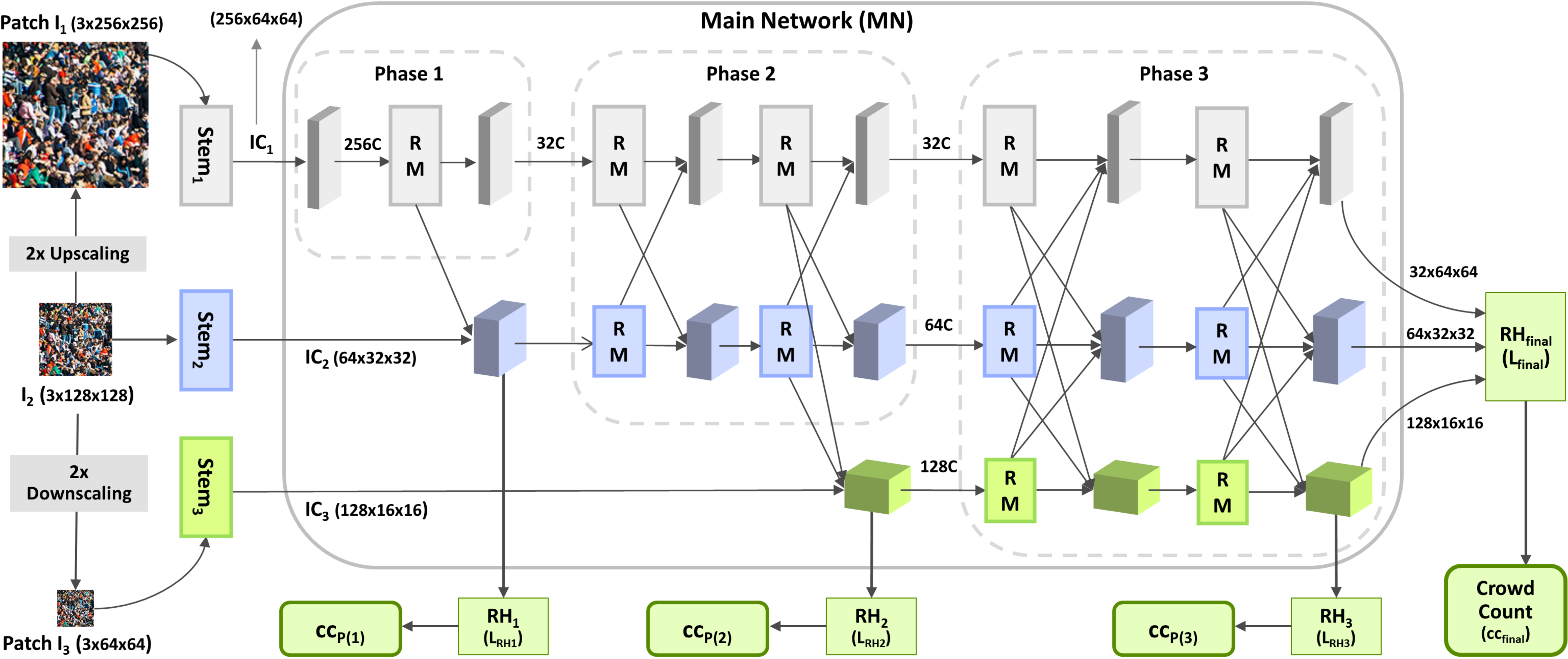
OUR OBJECTIVES

Crowd Counting

1. **Better generalization ability:** Design a multi-column crowd counting method with better generalization ability towards huge crowd variations.

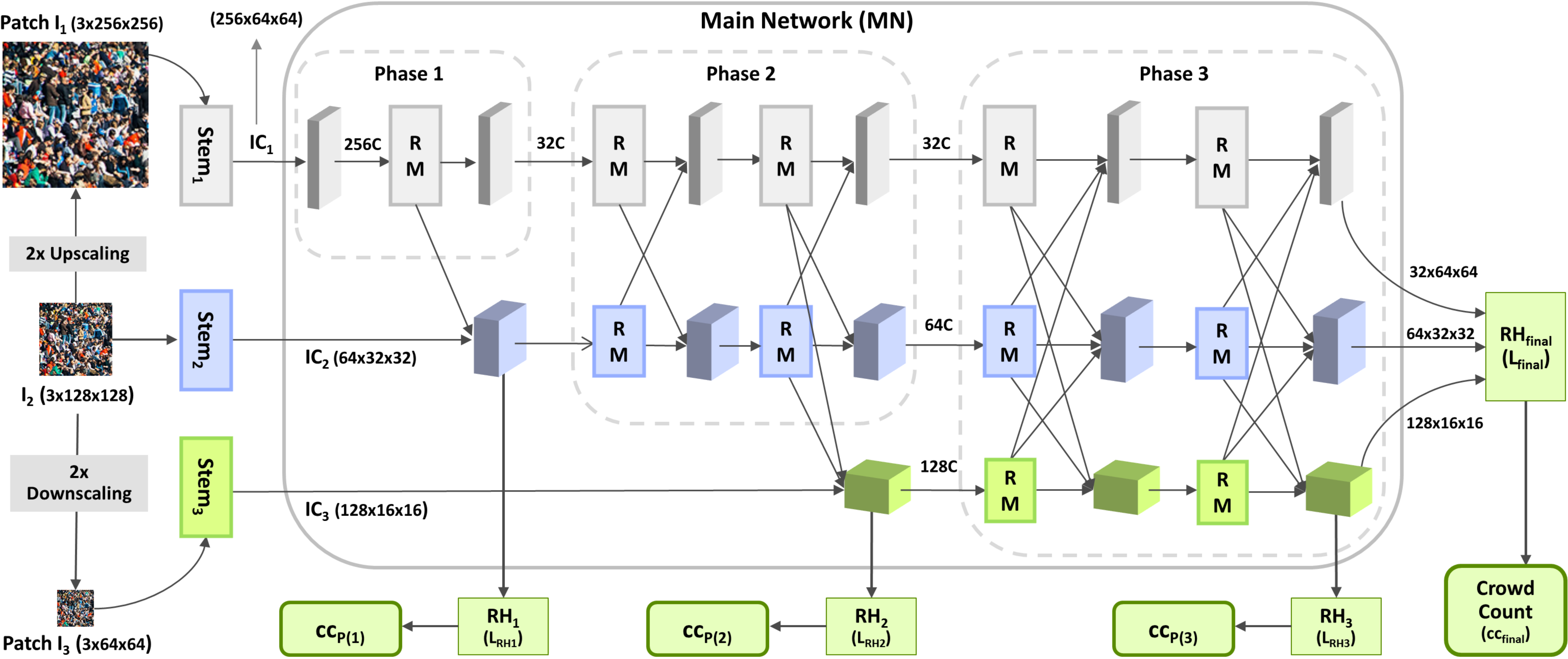
2. **Effective input priors:** Utilize the input patch rescaling based effective observation [27], [28] (as discussed above) without performing any expensive and compromising crowd-density classification process, and also use all three crowd-density levels (low-, medium, and high-crowd) in a more effective manner than the PRM module [28].

Multi-Resolution Fusion and Multi-scale Input Priors Based Crowd Counting



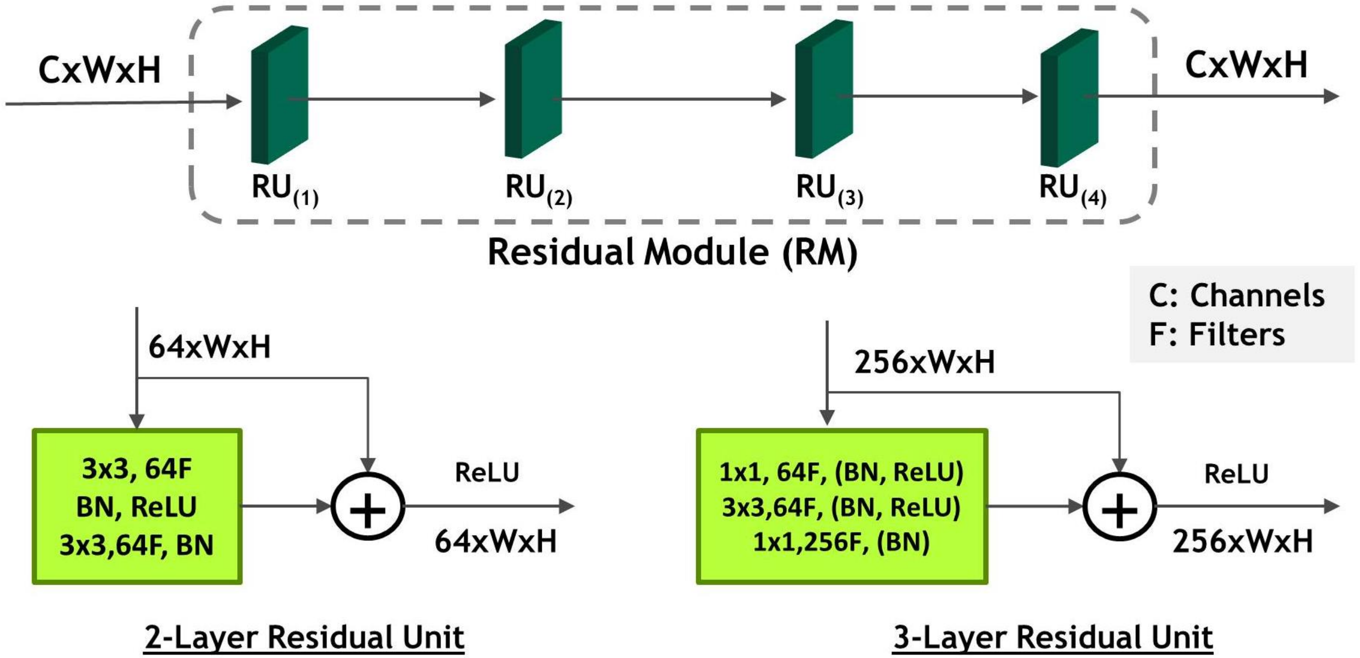
The Proposed Architecture

Multi-Resolution Fusion and Multi-scale Input Priors Based Crowd Counting



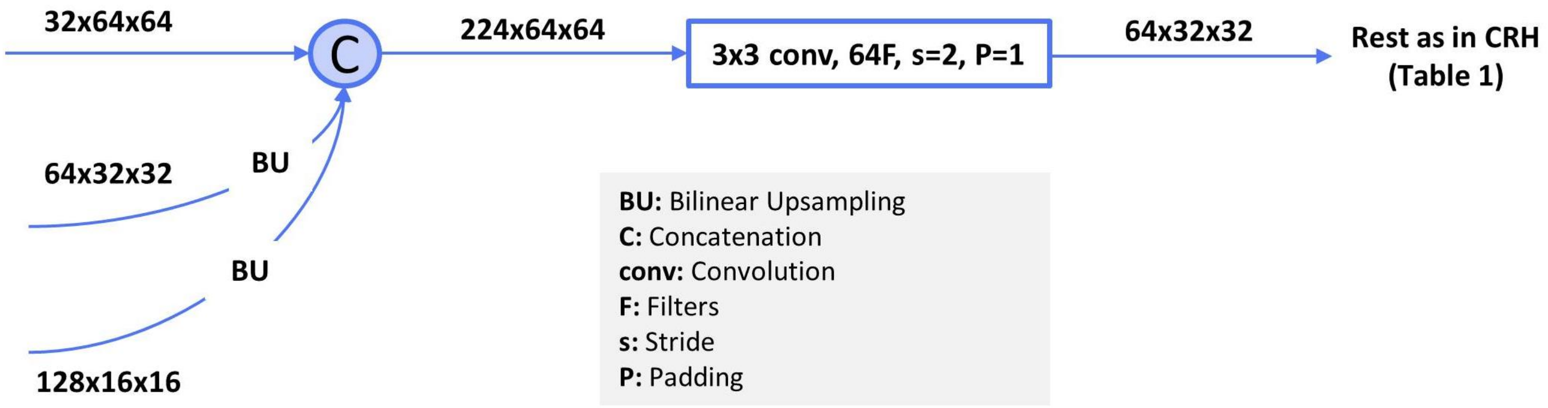
The Proposed Architecture

Residual Module (RM)



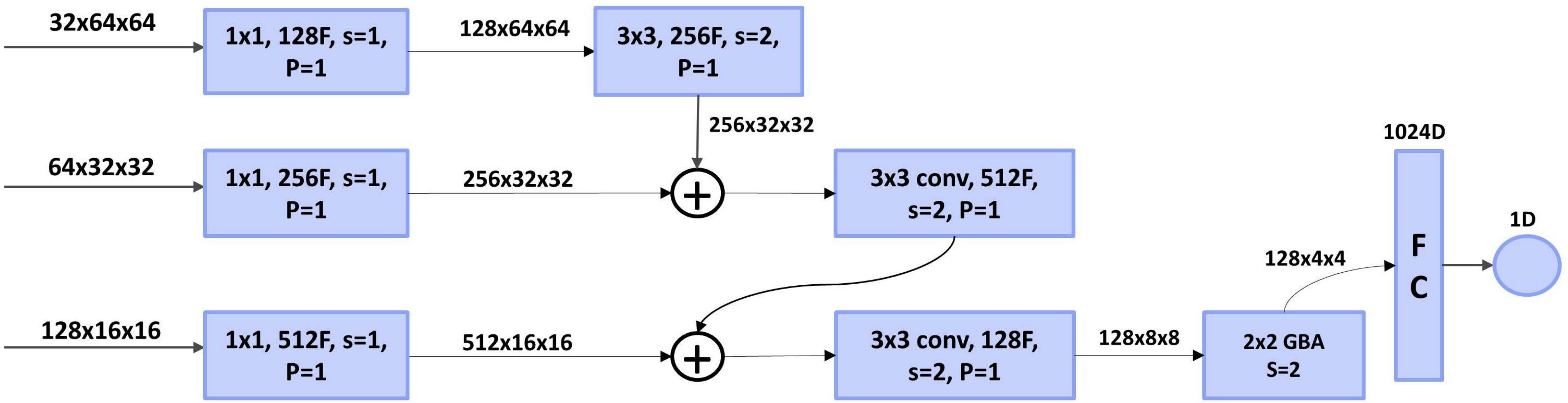
The Residual Module (RM) consists of either only 2- or 3-layers [10] based four residual units (RU).

Crowd Regression Head (RH_{final})



Concatenation-based crowd regression head (v4) concatenates the lower-resolutions with the highest-level channels using the bilinear upsampling.

Crowd Regression Head (RH_{final})



The summation-based head (v5) adds the higher-level channels into the lowest-resolution feature maps.

QUALITATIVE ANALYSIS

Crowd Counting with Our Approach

Actual Count=597
Our estimate=595
PRM=431
Density Map=301

Actual Count:1929
Our estimate:1920
PRM=1395
Density Map=623

Actual Count=3653
Our estimate=3639
PRM=2792
Density Map=2792

Actual Count=1070
Our estimate=1072
PRM=1011
Density Map=722

QUANTITATIVE ANALYSIS

Comparison with recent state-of-the-art methods

	ShanghaiTech Dataset		UCF-QNRF Dataset	
Method	MAE	RMSE	MAE	RMSE
CFF [30]	65.2	109.4	93.8	146.5
RRSP [33]	63.1	96.2	-	-
CAN [21]	**62.3**	100.0	107	183
TEDNet [17]	64.2	109.1	113	188
L2SM[38]	64.2	98.4	104.7	173.6
BL [21]	62.8	101.8	**88.7**	154.8
ZoomCount [27]	66.6	94.5	128	201
PRM-Based [28]	67.8	86.2	94.5	141.9
v5 (ours)	67.1	**81.0**	96.9	**130.1**

	AHU-Crowd Dataset	
Method	MAE	RMSE
DPM [8]	395.4	-
BOW-SVM [7]	218.8	-
Ridge Regression [6]	207.4	-
Hu et. al. [14]	137	-
DSRM [41]	81	129
ZoomCount [27]	88.2	126.1
PRM-Based [28]	74.9	111
v5 (ours)	**60.2**	**91.7**

MAJOR CONTRIBUTIONS

Crowd Counting

1. Designed a new multi-resolution feature-level fusion based end-to-end crowd counting approach for still images that effectively deals with significant variations of crowd-density, lighting conditions, and large perspective

2. Proposed an alternative patch rescaling module by more effectively using the input priors

3. Outperformed the state-of-the-art methods, including the PRM based schemes, by a large margin with up to 10% improvements

THANK YOU ☺