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
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
The Residual Module (RM) consists of either only 2- or 3-layers [10] based four residual units (RU).
Concatenation-based crowd regression head (v4) concatenates the lower-resolutions with the highest-level channels using the bilinear upsampling.
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

<table>
<thead>
<tr>
<th>Method</th>
<th>ShanghaiTech Dataset</th>
<th>UCF-QNRF Dataset</th>
<th>AHU-Crowd Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>CFF [30]</td>
<td>65.2</td>
<td>109.4</td>
<td>93.8</td>
</tr>
<tr>
<td>RRSP [33]</td>
<td>63.1</td>
<td>96.2</td>
<td>-</td>
</tr>
<tr>
<td>CAN [21]</td>
<td>62.3</td>
<td>100.0</td>
<td>107</td>
</tr>
<tr>
<td>TEDNet [17]</td>
<td>64.2</td>
<td>109.1</td>
<td>113</td>
</tr>
<tr>
<td>L2SM [38]</td>
<td>64.2</td>
<td>98.4</td>
<td>104.7</td>
</tr>
<tr>
<td>BL [21]</td>
<td>62.8</td>
<td>101.8</td>
<td>88.7</td>
</tr>
<tr>
<td>ZoomCount [27]</td>
<td>66.6</td>
<td>94.5</td>
<td>128</td>
</tr>
<tr>
<td>PRM-Based [28]</td>
<td>67.8</td>
<td>86.2</td>
<td>94.5</td>
</tr>
<tr>
<td>v5 (ours)</td>
<td>67.1</td>
<td>81.0</td>
<td>96.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPM [8]</td>
<td>395.4</td>
<td>-</td>
</tr>
<tr>
<td>BOW-SVM [7]</td>
<td>218.8</td>
<td>-</td>
</tr>
<tr>
<td>Ridge Regression [6]</td>
<td>207.4</td>
<td>-</td>
</tr>
<tr>
<td>Hu et. al. [14]</td>
<td>137</td>
<td>-</td>
</tr>
<tr>
<td>DSRM [41]</td>
<td>81</td>
<td>129</td>
</tr>
<tr>
<td>ZoomCount [27]</td>
<td>88.2</td>
<td>126.1</td>
</tr>
<tr>
<td>PRM-Based [28]</td>
<td>74.9</td>
<td>111</td>
</tr>
<tr>
<td>v5 (ours)</td>
<td>60.2</td>
<td>91.7</td>
</tr>
</tbody>
</table>
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 😊