Counting in Dense Crowds using Deep Learning

Logan Lebanoff, Haroon Idrees

University of Central Florida

UCF Center for Research in Computer Vision

Problem and Motivation:



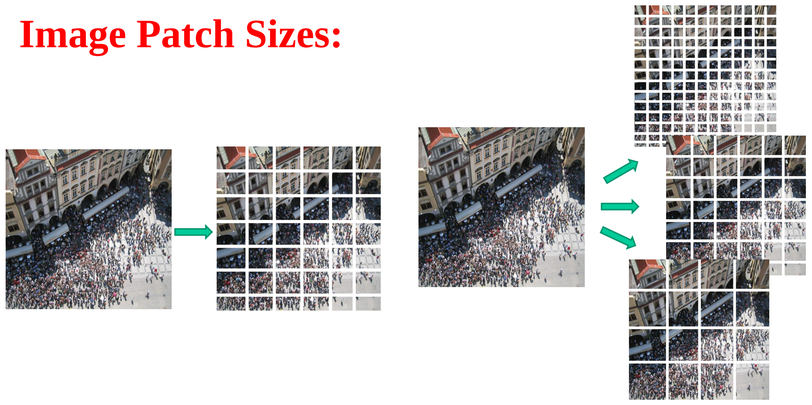
1940



469

- Use deep learning to count the number of people in images of dense crowds
- A first step in crowd analysis
- Helpful in development of transportation infrastructure
- Safety for concerts, political rallies, sports events, etc.

Image Patch Sizes:



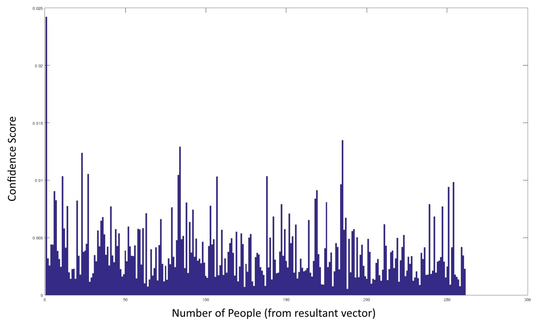
Loss Functions:

Classification Model:

Softmax Loss

$$y = -\sum_{ij}\left(x_{ijc} - \log\sum_{d=1}^{D} e^{x_{ijd}}\right)$$

$$\frac{dz}{dx_{ijd}} = -\frac{dz}{dy}\left(\delta_{d=c} - y_{ijc}\right)$$



Regression Model:

Euclidean Loss

$$y = \frac{1}{2}\sum(x - x_0)^2$$

$$\frac{dy}{dx} = \sum(x - x_0)$$

Difference-to-Sum Ratio Loss

$$y = \sum(x + 2x_0(ln(x + x_0))$$

$$\frac{dy}{dx} = \sum(\frac{x - x_0}{x + x_0})$$

- Regression model should compute the weights and error more accurately than a classification model
- When calculating error, the network knows the distance from the result to the correct value
- Euclidean loss only accounted for the absolute difference in error
- This new ratio loss function minimizes the normalized error, so that error for smaller counts is penalized the same as that for larger counts.

Results:



Groundtruth: **581** Result: **597**



Groundtruth: **2391** Result: **1946**



Groundtruth: **1961** Result: **1982**



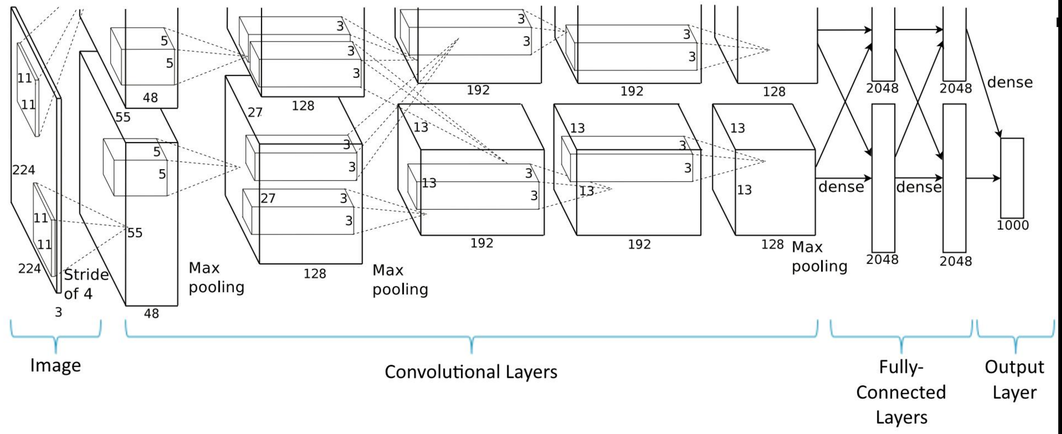
Groundtruth: **2550** Result: **2407**

	Absolute Difference	Normalized Absolute Difference
Previous method	324.42	30.91 %
Softmax Loss	393.56	35.43 %
Euclidean Loss	296.50	47.55 %
Difference-to-Sum Ratio Loss	261.73	30.09 %
Varying patch sizes	271.35	24.89 %

Method:







Conclusion:

- Deep learning is a viable method for counting in dense crowds and possibly for counting in other densely-packed scenarios.
- The regression loss function with difference-to-sum ratio improves results.
- Training on varying image patch sizes increases robustness to scale.
- Fine-tuning a deeper neural network may perform better, as counting the number of people in a crowd is a fairly complex problem.