# Google team's neural network approach works on street numbers

10 January 2014, by Nancy Owano



Dif?cult but correctly transcribed examples from the internal street numbers dataset. Some of the challenges in this dataset include diagonal or vertical layouts, incorrectly applied blurring from license plate detection pipelines, shadows and other occlusions. Credit: arXiv:1312.6082 [cs.CV]

(Phys.org) —A Google team has worked out a neural network approach to transcribe house numbers from Street View images, reading those house numbers and matching them to their geolocation. Google Street View has the user advantage of allowing the user to advance to street level to see the area of interest in detail. Google's accomplishment in automation is impressive both in the scope of the task involved and the way in which it was done. Consider that Google's Street View cameras have recorded massive numbers of panoramic images carrying massive numbers of house numbers. "We can for example transcribe all the views we have of street numbers in France in less than an hour using our Google infrastructure," said the researchers, who have authored the paper, "Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks." Ian J. Goodfellow, Yaroslav Bulatov,

Julian Ibarz, Sacha Arnoud, Vinay Shet are the authors.

The paper was submitted to arXiv and was explored in a report earlier this week in *MIT Technology Review*, which examines their research. The team used a neural network that contains 11 levels of neurons trained to spot numbers in images. The researchers describe the network as "a deep convolutional neural network that operates directly on the image pixels." They said they used the DistBelief implementation of deep [neural networks](#) to train large, distributed neural networks on high-quality images. "We find that the performance of this approach increases with the depth of the convolutional network, with the best performance occurring in the deepest architecture we trained, with eleven hidden layers."

At specific operating thresholds, the performance of the proposed system, they said, is comparable to that of human operators. "To date, our system has helped us extract close to 100 million physical street numbers from Street View imagery worldwide."

As *MIT Technology Review* [points out](#), the very task of matching any building number to its location is not always easy. There are places in the world where buildings are not numbered in clear patterns and *Wired* made the point that some house numbers carry [styles](#) and character arrangements that make identification difficult.

Nonetheless, Goodfellow and team forged ahead, unleashing the network, designed with a number of built-in assumptions to ease the effort, including fixed length: The team assumed that the numbers showing up in any image were at least one third the width of the resulting frame. "In this work we assume that the street numbers have already been roughly localized, so that the input image contains only one street number, and the street number itself is usually at least one third as wide as the image

itself." They also assumed that a number would not exceed five digits. "One special property of the street number transcription problem is that the sequences are of bounded length. Very few street numbers contain more than five digits, so we can use models that assume the sequence length n is at most some constant N, with N = 5 for this work."

The authors believe the Street View experience with a neural network could apply to other excursions in technology research. "This approach of using a single neural network as an entire end-to-end system could be applicable to other problems, such as general text transcription or speech recognition."

Goodfellow's research work at the Université de Montréal has been in machine learning and computer vision.

The authors have also submitted the paper to the ICLR 2014.

**More information:** Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks, arXiv:1312.6082 [cs.CV] arxiv.org/abs/1312.6082

Recognizing arbitrary multi-character text in unconstrained natural photographs is a hard problem. In this paper, we address an equally hard sub-problem in this domain viz. recognizing arbitrary multi-digit numbers from Street View imagery. Traditional approaches to solve this problem typically separate out the localization, segmentation, and recognition steps. In this paper we propose a unified approach that integrates these three steps via the use of a deep convolutional neural network that operates directly on the image pixels. We employ the DistBelief implementation of deep neural networks in order to train large, distributed neural networks on high quality images. We find that the performance of this approach increases with the depth of the convolutional network, with the best performance occurring in the deepest architecture we trained, with eleven hidden layers. We evaluate this approach on the publicly available SVHN dataset and achieve over 96% accuracy in recognizing complete street numbers. We show that on a per-digit recognition task, we improve upon the state-of-the-art and achieve 97.84% accuracy. We also evaluate this approach on an even more challenging dataset generated from Street View imagery containing several tens of millions of street number annotations and achieve over 90% accuracy. Our evaluations further indicate that at specific operating thresholds, the performance of the proposed system is comparable to that of human operators. To date, our system has helped us extract close to 100 million physical street numbers from Street View imagery worldwide.