Smart Inference for Multidigit CNN-based Barcode Decoding

Thao Do et al.
Korea Advanced Institute of Science and Technology
thaodo@kaist.ac.kr

Abstract

Barcodes are ubiquitous and have been used in most daily activities for decades. However, most traditional barcode scanners require well-founded barcode under standard conditions. While wilder conditioned barcodes such as underexposed, occluded, wrinkled are commonly captured in reality, traditional scanners show weakness of recognizing. This work aims to solve the detecting and decoding problem using deep convolutional neural network with the possibility of running on portable devices. We proposed a special modification of inference based on the attribute of having self-validation (checksum) at in the prediction phase of a trained model. The experiments' results demonstrated the SI effectiveness with the highest accuracy of 95.85% which outperformed existing decoders on the evaluation set. Finally, we minimized the best model by knowledge distillation to a shallow model which is shown to have high accuracy (90.85%) with a good inference speed of 34.2 ms per image on a real edge device.

Introduction

Linear 1D barcodes appeared in the 1970s and are now become ubiquitous on almost all consumer products and for logistics due to its ease of identification. For the barcode, the readers (or the scanners) are categorized into 3 types: laser-based, LED-based, and camera-based. Camera-based readers have some advantages over laser/LED-based solutions. The first advantage is built on the fact that numerous smartphones with high-quality cameras integrated are already in use. With Internet connection, useful mobile applications were born by online retrieval of product information and giving out ingredients information.

However, most current techniques (static image processing and pattern matching) being used in camera-based readers have flaws that limit their usability. The main problem with them is the need for well-framed flatbed-scanned-style input than normal captured. Wilder but common-captured conditions such as underexposed, occluded, blurry or curved, non-horizontal position.

Main Contributions

1. We proposed Smart Inference - 3 algorithms leveraging the feature of having checksum and test-time augmentation built on top of trained deep CNN models which considerably boost model accuracies and reduce false prediction;
2. We made a challenging 2500-sample-cropped EAN13(UPC-A, ISBN13 are its subsets) barcode dataset from real captured images on various (included harsh) conditions on numerous products - this dataset is publicly available.
3. Lastly, we applied knowledge distillation technique to have a lightweight model from the best model which is suitable on handheld devices, the experimental result confirmed the possibility by a good speed on a real edge board.

Methods

The base approach we use in this work is to train a probabilistic model of decoding barcode sequences given barcode images as [1]. Let $\mathbf{D}$ represent the barcode sequence and X represent the input barcode image. The goal is to learn a model of $P(\mathbf{D}|X)$ by maximizing $\log P(\mathbf{D}|X)$ on the training set. $\mathbf{D}$ is modelled as a collection of n random variables $D_1, ..., D_n$ representing n digits of the decoded sequence. To simplify, assume that the value of the each digits are independent from each other, so that the probability of a sequence $d = d_1, ..., d_n$ is given by $P(\mathbf{D} = \mathbf{d}|X) = \prod_{i=1}^{n} P(D_i = d_i|X)$. Each of the digits is discrete and has 10 possible values (0 to 9). This means each digit could be represented with a softmax classifier that receives as input features extracted from X by a CNN. This type of model is originally proposed by [2], so we call it Multidigit CNN.

However, in the inference phase, instead of normal inference, we propose a modification named Smart Inference (SI).

Smart Inference

Normally after getting logits from the model given barcode images, we apply softmax function to get the probabilities of each value (0 to 9) for all n digits; then, we pick the value with the highest probability. By this way, we finally have n-digit sequences from values having the highest probabilities. *However, the value with the highest probability is not always the correct value*. Instead, the correct value maybe the value with the second-highest or third-highest probability. Besides, since most 1D barcodes have a characteristic of checksum satisfaction.

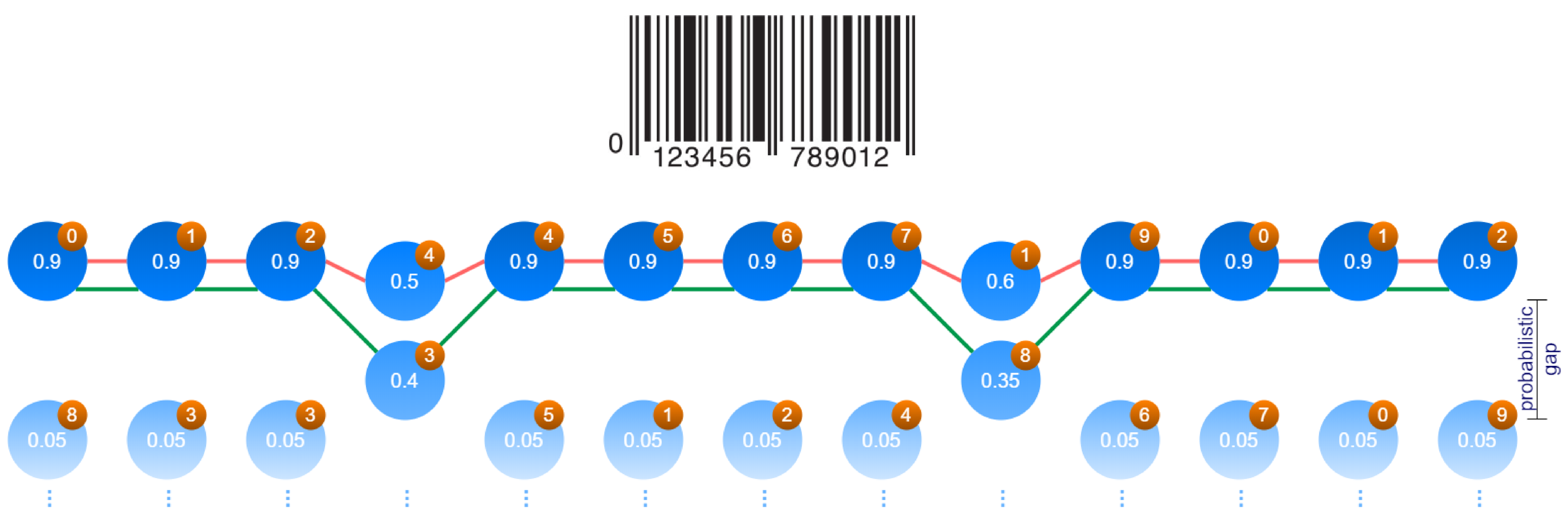

Figure 1: A Smart Inference simplified example

Results

Below is the best performance experiment using Smart Inference (SI) and test-time augmentation. With SI, model used ResNet50 jumps from 93.35% to 95.8%, MobileNetV2 from 83.45% to 90.6% and so on. Note that compared with industrial tools, best of them are just 93.1% from Dynamsoft but it's very slow $\sim$ 900 $ms/images$.

	Model	*nonMPA*	*max=1*	*max=2*	*max=3*	*max=4*
Accuracy	ResNet50	0.9335	**0.958**	0.956	0.951	0.946
	MobiletNetV2	0.8345	**0.906**	0.8975	0.8855	0.8705
	DenseNet169	0.849	**0.9155**	0.9075	0.8915	0.877
	ResNet34	0.892	**0.9375**	0.9315	0.9235	0.916
	Non-residual	0.808	**0.89**	0.879	0.861	0.8445
# of errors	ResNet50	133	**56**	73	95	108
	MobiletNetV2	331	**121**	182	225	259
	DenseNet169	302	**113**	172	216	246
	ResNet34	216	**95**	129	152	168
	Non-residual	384	**145**	212	274	311

Table 1: Using MPA & Augmentation performances

Finally, our experimental tests on the Jetson Nano board show that MobileNetV2, ResNet34 achieved average speeds of *34.2*, *45.6* milliseconds per images respectively. This speed is equivalent to a smooth frame-per-second experience with the robustness of the model.

References

[1] Fredrik Fridborn. Reading barcodes with neural networks, 2017.

[2] Ian J Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, and Vinay Shet. Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*, 2013.