Embedded Large–Scale Handwritten Chinese Character Recognition

Y. Chherawala, H. Dolfing, R.S. Dixon, J. Bellegarda
Apple - ICASSP 2020

Chinese handwriting recognition

On-device deployment requirements

- Character inventory selection
- Robustness to writing style
- High accuracy, real time recognition

Agenda

Dataset design
Recognition model
Model training
Results
Conclusion

Dataset design

Chinese characters

- Over 100,000 unique Chinese characters in dictionaries
- Large number of existing character encodings (GB2312-80, Big5, Big5E, CNS 11643-92, HKSCS-2008, GB18030-2005)
- Ideal set of characters varies from user to user

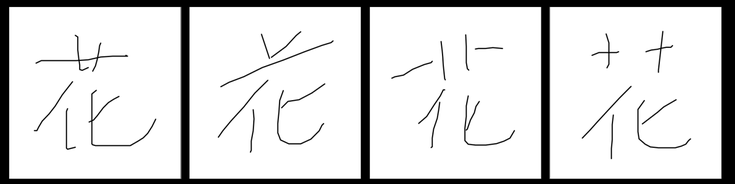
Character inventory selection

- Selection of characters suitable for daily correspondence
  - Hànzì part of GB18030-2005, HKSCS-2008, Big5E
  - Core ASCII set
  - Visual symbols and emojis
- Total of ~30K characters

Writing style coverage

- Large-scale collection across Greater China
- Diverse writer population
- 30K unique characters
- Over 10 million handwriting samples

花 (flower - U+82B1)



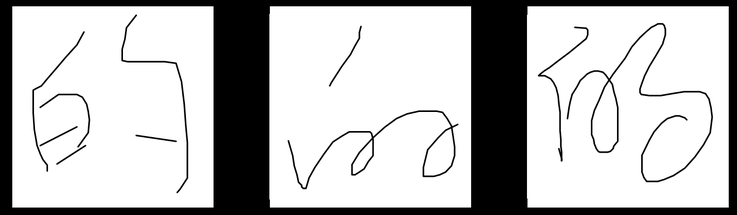
Printed radical variations

Cursive radical variations

Unconstrained variations

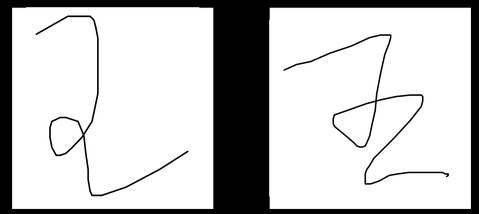
Writing style variations

的 (U+7684)



以 (U+4EE5)

王 (U+738B) and 五 (U+4E94)



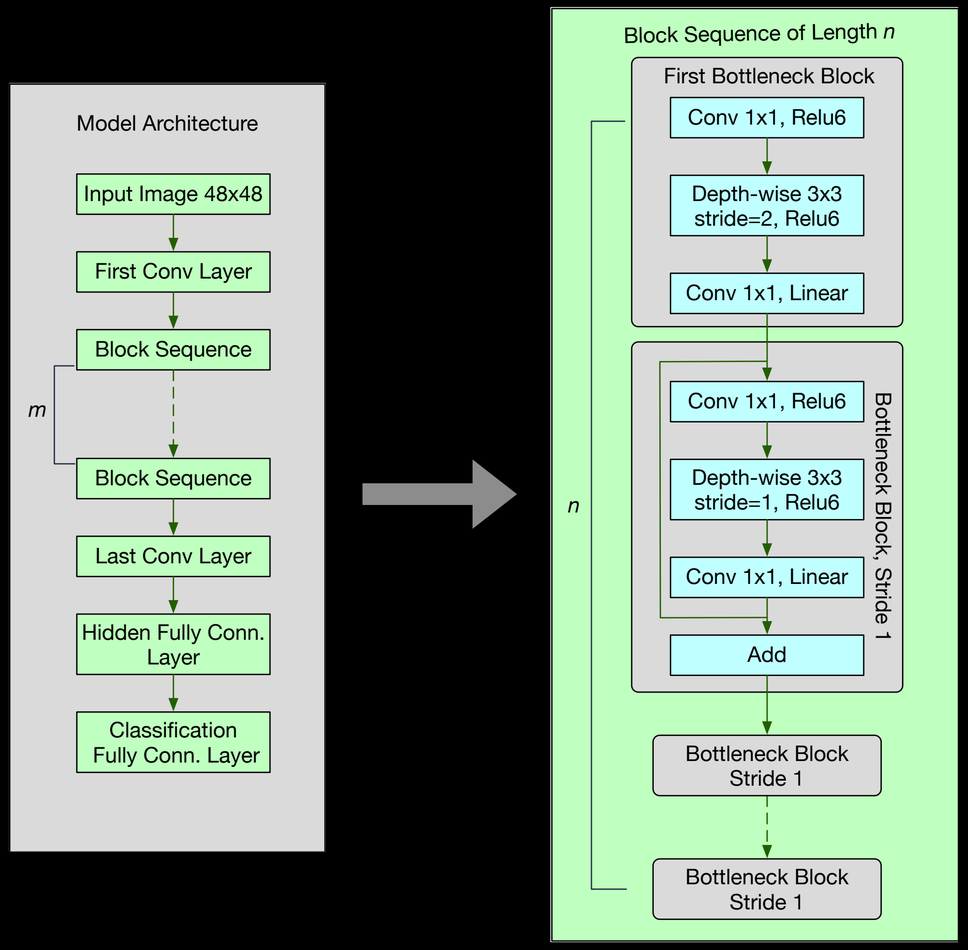
Similar shapes

Recognition model

MobileNet v2 variant

- Independence to writing stroke-order
- High accuracy, real time recognition
- Suitable for embedded devices

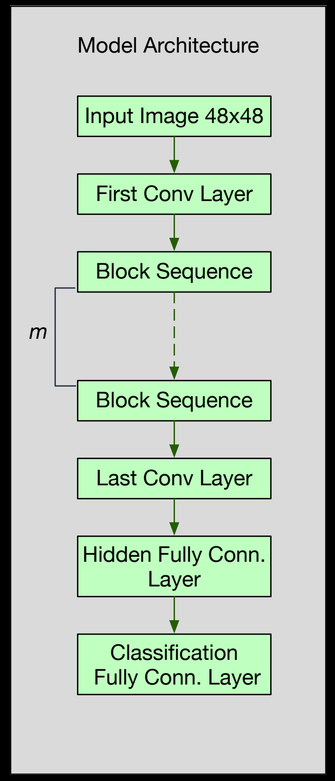
MobileNet v2 variant



- Spatial subsampling directly controlled by the number of block sequences
- Large number of output classes, limit parameters explosion through the last layers size

Model training

Architecture optimization



Hyper-parameters	Values
Expansion factor *t*	[2, 4, 6, 8]
Nb. block sequence	[3, 4, 5]
Max sequence length	[2, 4, 6, 8, 10]
First conv. channels	[16, 24, 32, 48]
Last conv. channels	[128, 256]
Fully conn. layer size	[0, 128, 256]

- Randomly sample 50 model configurations for each experiment
- Accuracy delta of 0.6% between the 10 best configurations

Architecture optimization

- Different configurations lead to similar top accuracy
- Small last dense layer provides model with reasonable disk footprint

Configuration 1 (19M disk footprint)

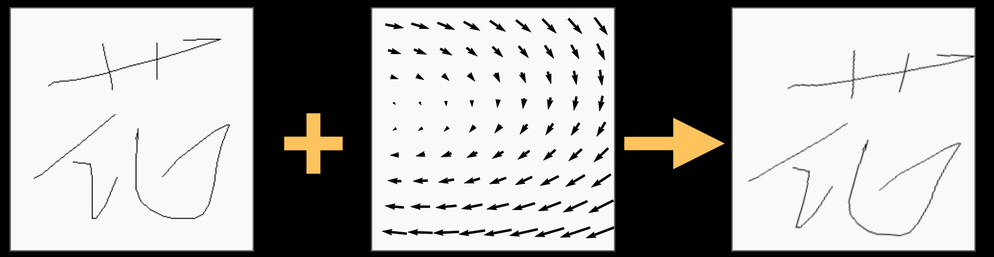
Input	Layer	*t*	*c*	*n*
48x48x1	conv2D	-	16	-
48x48x16	bottleneck	8	64	4
24x24x64	bottleneck	8	64	4
12x12x64	bottleneck	8	64	4
6x6x64	bottleneck	8	96	2
3x3x96	conv2D	-	128	-
3x3x128	dense	-	128	-
128	dense	-	30K	-

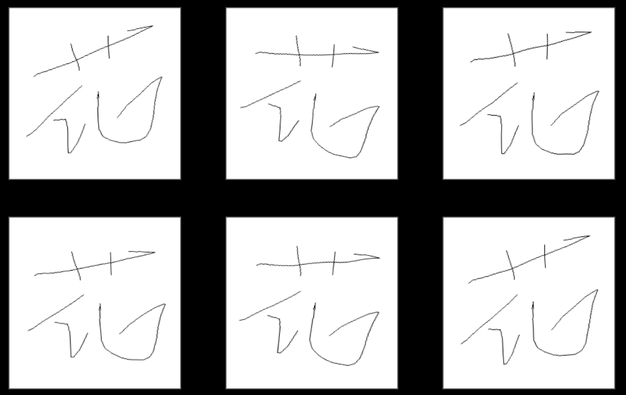
Configuration 2 (17M disk footprint)

Input	Layer	*t*	*c*	*n*
48x48x1	conv2D	-	24	-
48x48x16	bottleneck	4	32	7
24x24x64	bottleneck	4	64	8
12x12x64	bottleneck	4	96	4
3x3x96	conv2D	-	128	-
3x3x128	global avg. pool	-	128	-
128	dense	-	30K	-

Data augmentation

- Runtime synthetic data augmentation
- Virtually infinite training variations
- Elastic deformations
- Affine transformations





Results

Scaling up to 30K

- Accuracy on CASIA database test set

Character inventory	Training data	Top 1	Top 4	Top 10	Model size
3,755 (Hànzì-1)	CASIA	95.1%	98.9%	99.5%	11MB
3,755 (Hànzì-1)	CASIA + In-House	96.8%	99.4%	99.7%	13MB
30K	CASIA + In-House	96.6%	99.3%	99.6%	19MB

Scaling up to 30K

- Accuracy on the in-house database test set

Model	Character inventory	Training data	Top 1	Top 4	Top 10	Model size
mobileNetV2	30K	CASIA + In-House	97.2%	99.6%	99.8%	19MB
LeNet	30K	CASIA + In-House	92.6%	98.4%	99.2%	15MB

Conclusion

Conclusion

- Unique challenge of Chinese character recognition
- Data collection conditions
- Real time performance with small disk footprint for mobile devices
- Coverage of larger inventory is within reach
  - 75,000 CJK characters present in the Unicode character set
  - Learning curve extrapolation indicates that a Top 1 accuracy of 96% should be possible for a coverage of 100,000 characters with sufficient data

An open invitation to open minds.

bit.ly/apple-icassp2020