

Cucumber classification

CUCUMBER

By 성창규, 최홍범 and 오케스

What is the problem?

Different shape, color, size, price....

Categorize

Makoto Koike



https://artfarmllc.files.wordpress.com/2012/11/artfarm-cucumbers.jpg

Categorizing

- Categorization requires
 - Manual work
 - Knowledge

- Automatic sorters
 - Limited performance
 - o Cost



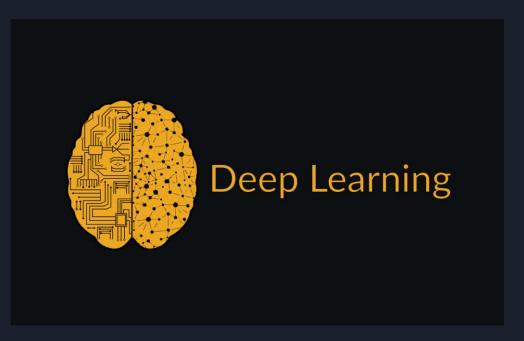
https://www.christies.com/img/LotImages/2012/CKS/2012_CKS_05602_0041_000(pieter_brueghel_ii_summer_the_harvesters).jpg

FOCUS ON THE RIGHT THINGS

How Makoto solved the problem

Automate the sorting

AlphaGo



https://cdn-images-1.medium.com/max/2000/1*u9L UJbV0Qfg1PZQkHna2g.png

What kinds of data are used?

• Makoto used about 7000 images of 80 x 80 pixel pictures of cucumber sorted by his mother.

• By using these pictures of cucumber, Makoto created image classification model of cucumber.

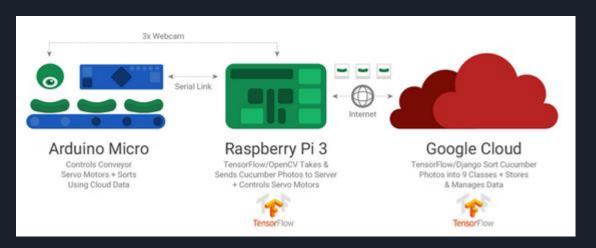




2L	-	-	-		-			-		_
L		_	-	-	-	-	_	^	_	-
М	-	-	-	-	_	~	_	_	_	-
S	-	-	-	-	-	-	-	-	-	-
28	~		-	-	-	-	-	-	-	-
BL	-	1		~	-	-		~	~	_
ВМ	-	-	-	~	-	~	-	-	^	~
BS	-	-	-	-	-	-	~	^	-	
С	^	-	-	-	^	-	^	-	^	-

Which techniques are used?

Matkoto's automated sorting system.



Makoto used CNN architecture to train the cucumber sorting model!

CNN

RED Channel



Green Channel



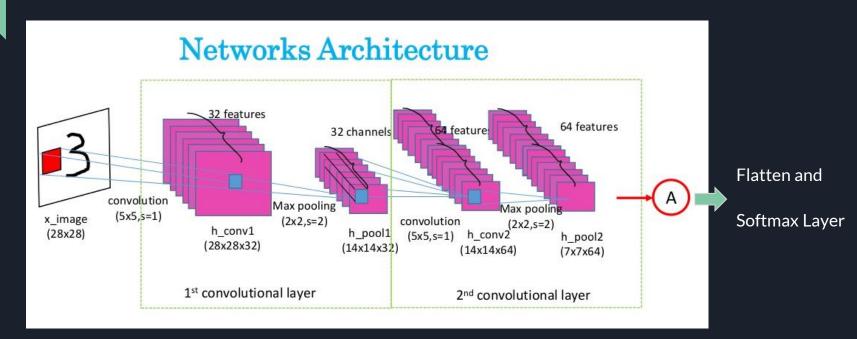
Blue Channel





이미지 출처: https://en.wikipedia.org/wiki/Channel_(digital_image)

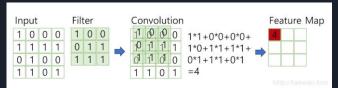
Network Architecture of CNN

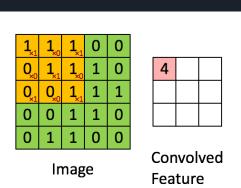


Feature Extraction

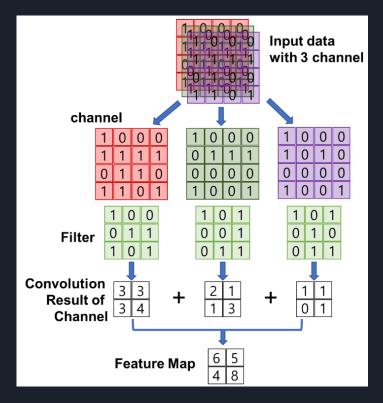
Convolution Layer

The process of performing convolution multiplication on input data with one filter

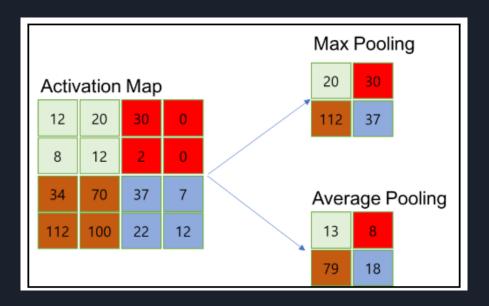




Procedure for convolution multiplication by applying several filters to multi-channel input data



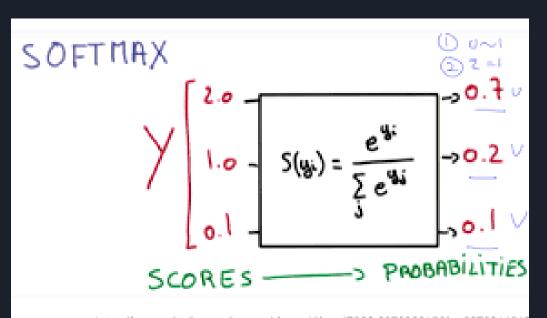
Pooling Layer



Pooling layer reduce the size of feature map or emphasize specific data on feature map

Flatten and Softmax Layer

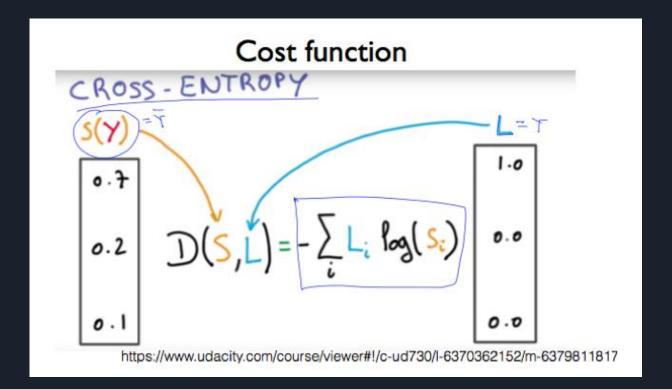
- Flatten layer changes the data type of CNN into Fully Connected Neural Network.
- There is no parameter in the Flatten layer, only the shape change of the input data is performed



 Softmax layer changes the scores from Convolutional through Flatten layer into probabilities to each class

https://www.udacity.com/course/viewer#1/c-ud730/1-6370362152/m-6379811817

Cost Function



What kinds of implication could be derived from the results of data analysis?

90% of accuracy for the trained set 60%~70% of accuracy for the real world test

The use case indicates the possibility of deploying CNN architecture in the real world.

Makoto's thoughts: "60%~70% is not enough for the classification in the real agricultural business"

There is a hurdle to deploy CNN algorithms for the cucumber classification!

How much benefit is obtained from the analysis?

1. Automating the classifying procedure

Makoto's farm has started this categorizing on *July* and is used in real use case since then. This saves *8 hours a day* of categorizing labor.

2. CNN algorithms does not require feature engineering processes

One of the main problem that each farm has, is that *classification rule differs* by person. CNN *learning by itself* for training.

Challenges and improvement

Difference between **test datasets** and **real use case** → Training more, to **avoid overfitting** problem!

Hardware restrictions → Google *Cloud ML* can be the breakthrough!

References

- https://cloud.google.com/blog/products/gcp/how-a-japanese-cucumber-farmer-is-using-deep-learning-and-tensorflow
- https://www.slideshare.net/ssuser06e0c5/explanation-on-tensorflow-example-deepmnist-for-expert
- http://deeplearning.stanford.edu/wiki/index.php/Feature_extraction_using_convolution
- http://taewan.kim/post/cnn/