IMPROVING RECOGNITION OF THAI HANDWRITTEN CHARACTER WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

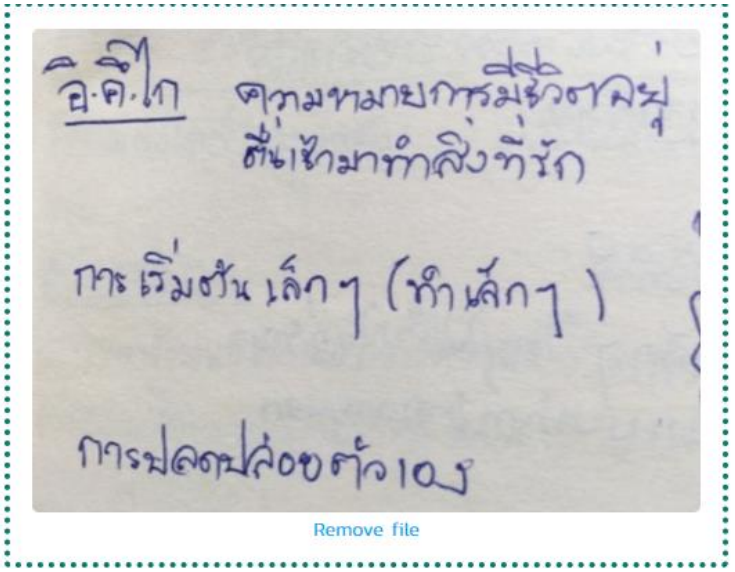
Sarayut Gonwirat
and
Olarik Surinta

Multi-agent Intelligence Simulation Laboratory
Department of Information Technology
Faculty of Informatics
Mahasarakham University, Thailand

Handwritten Character Recognition

- **Automatic Reading System**, how can machines understand the context in document.

- **Various Applications**:
  - Historical Document Analysis
  - Text Image Retrieval
  - Traffic-sign Recognition
  - Signature Verification



Previous Works

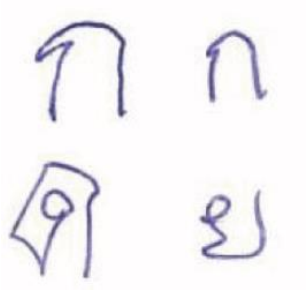
Different Language Characters

English

Chinese

Bangla

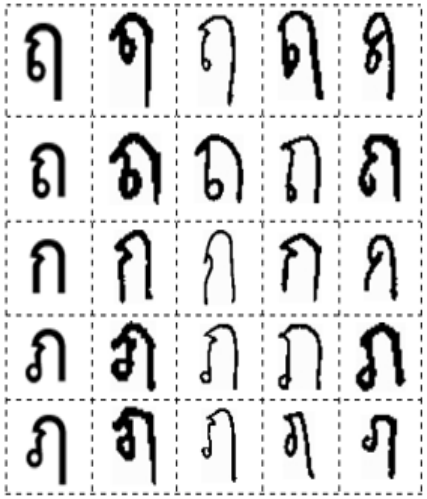
Hangul

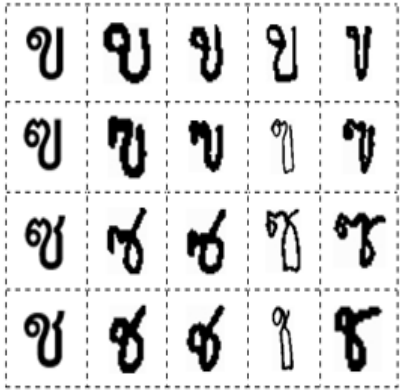


Thai

Difficulties :

- Different head and line-trail.
- Stroke and curve.
- Personal writing style.



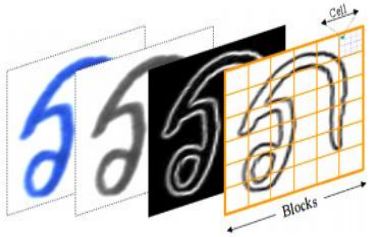


Feature Extraction

Handwritten Recognition Process

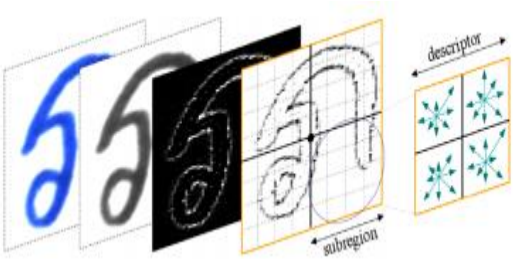
Images → Preprocessing → Segmentation → **Feature Extraction (Local Descriptors)** → Recognition (Machine Learning)

Local Descriptors



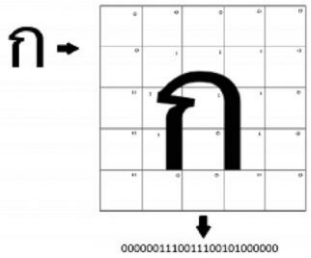
HoG

Histogram of Orientated Gradients



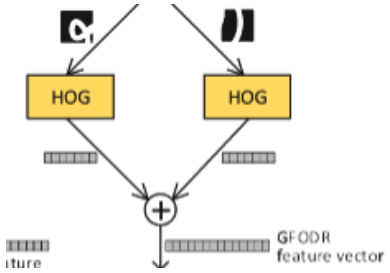
SIFT

Scale Invariant Feature Transform



LBP

Local Binary Pattern

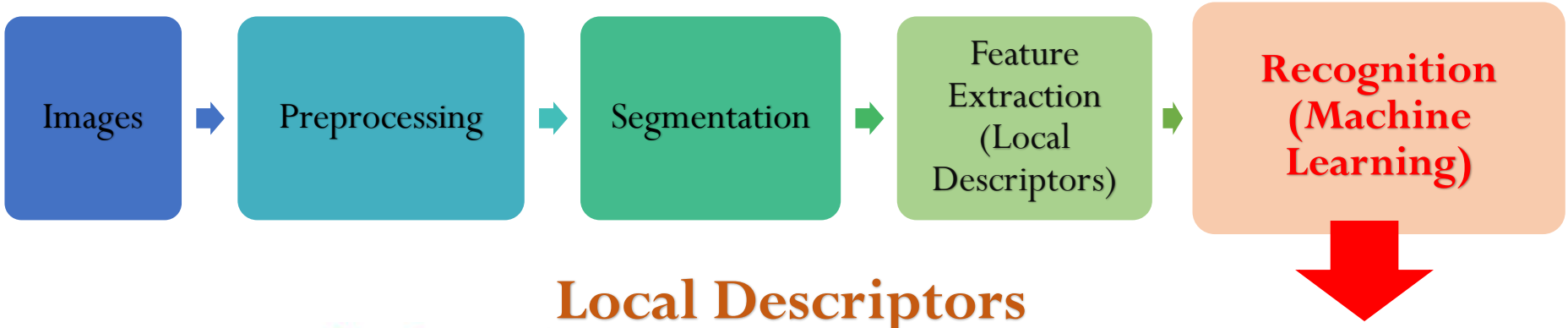


HoG-GFoDR

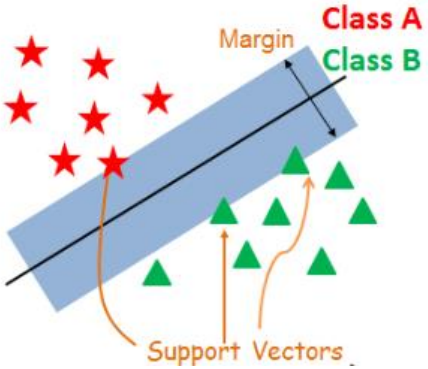
Gradient Features of Discriminative Regions

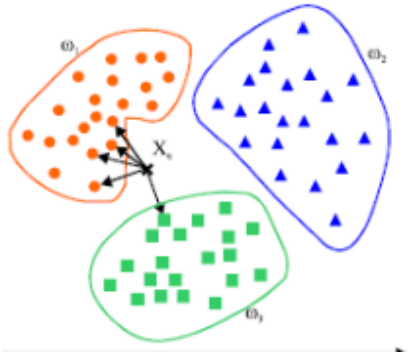
Machine Learning

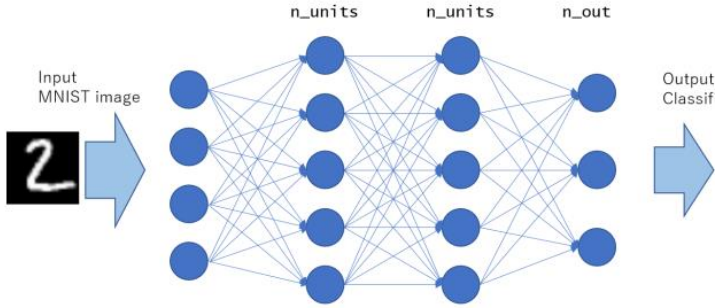
Handwritten Recognition Process



Local Descriptors







SVM

Support Vector Machine

KNN

K - Nearest Neighbor

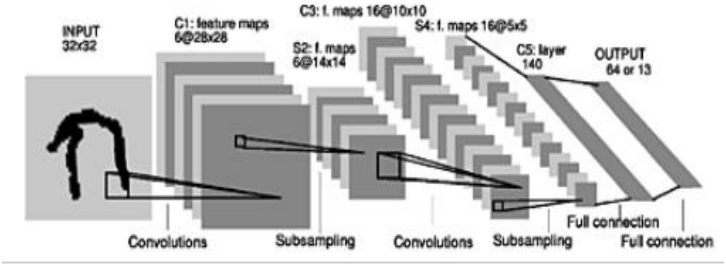
MLP

Multi-Layer Perceptron

Research Contributions:

Images

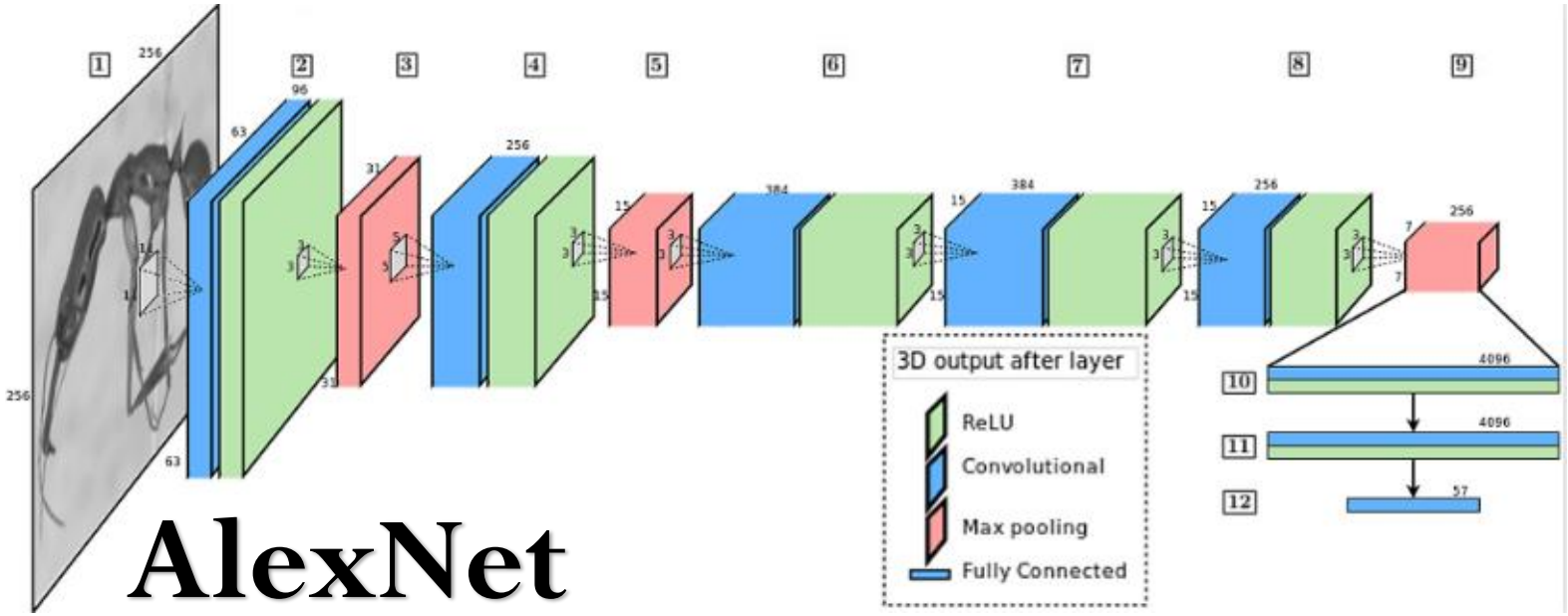
Recognition (Machine Learning)



Contributions:

- Improve the efficiency recognition of thai handwritten recognition by CNN
- Comparison of the CNN Architectures
  - VGGNet and Inception-ResNetV2 with feature-based SVM
  - Learning style, Scratch and Transfer Learing

CNN Background

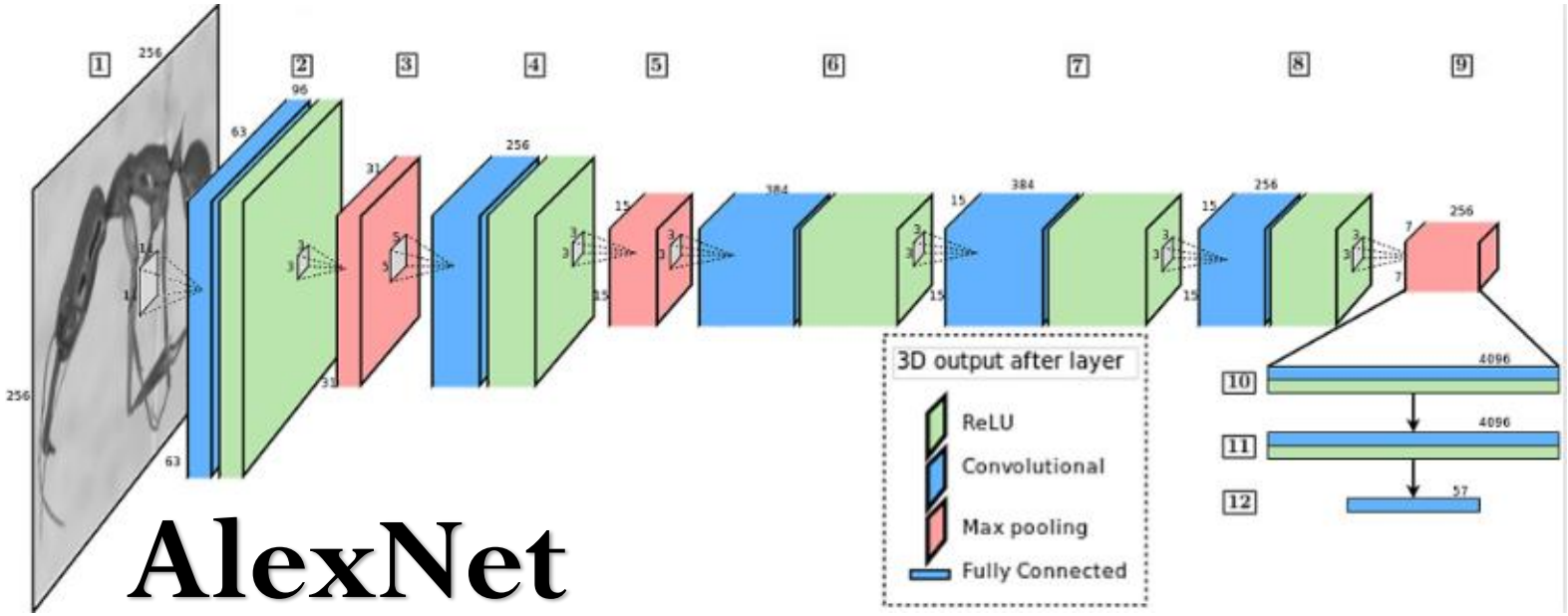


https://figshare.com/articles/CNN_architecture_used_to_perform_image_classification_/5843691

Layers :

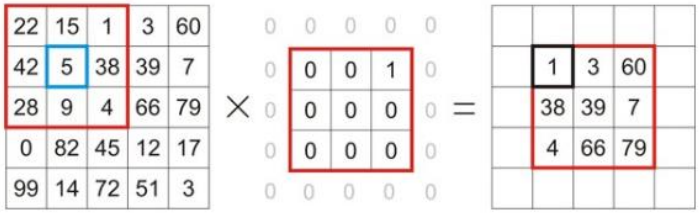
- Convolution Layers
- Pooling Layers
- Fully Connected Layers

CNN Background



AlexNet

Convolution Layer



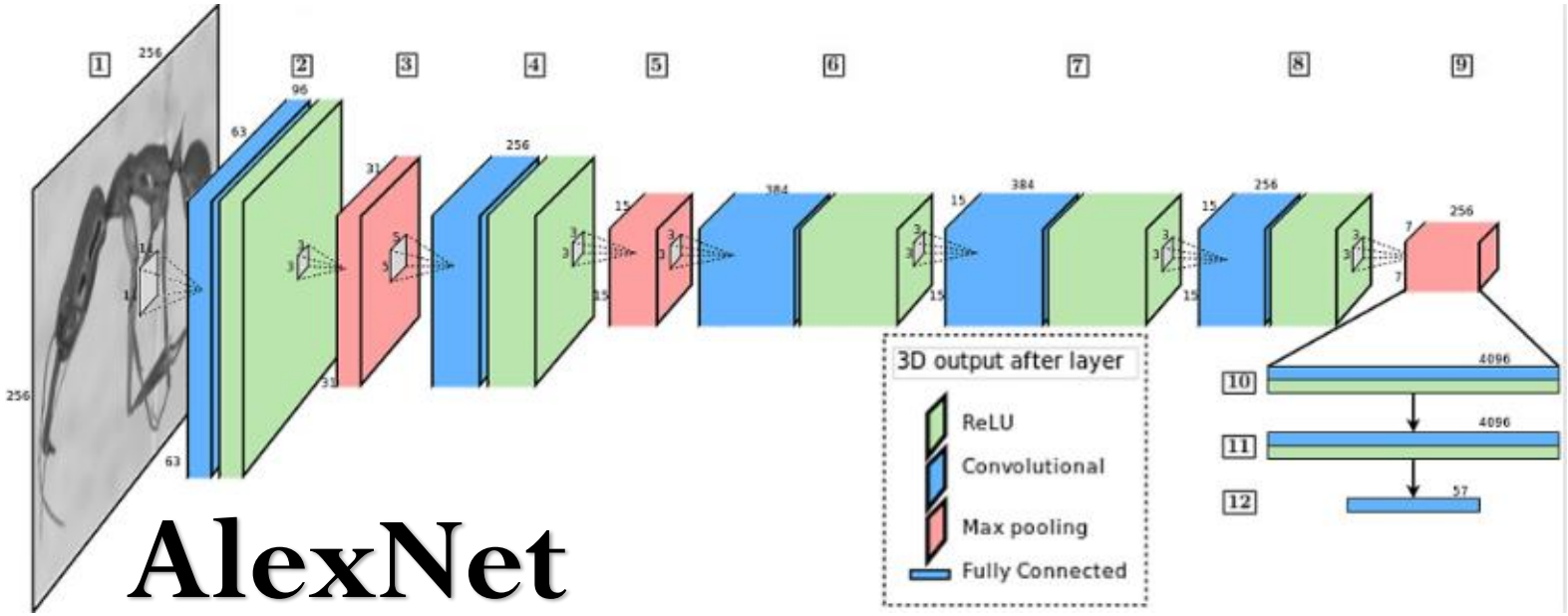
$$X_{k,l,n}^{p+1} = \sum_{i,j,m} K_{i,j,m,n}\, x_{k+i-1,l+j-1,m}^{p}$$

Activation Function

Rectified Linear Unit (ReLU)

$$ReLU(x) = \max(0, x)$$

CNN Background



AlexNet

Pooling Layer

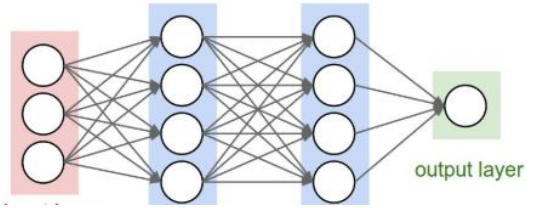
1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

max pool with 2x2 filters and stride 2 →

6	8
3	4

Fully Connected Layer

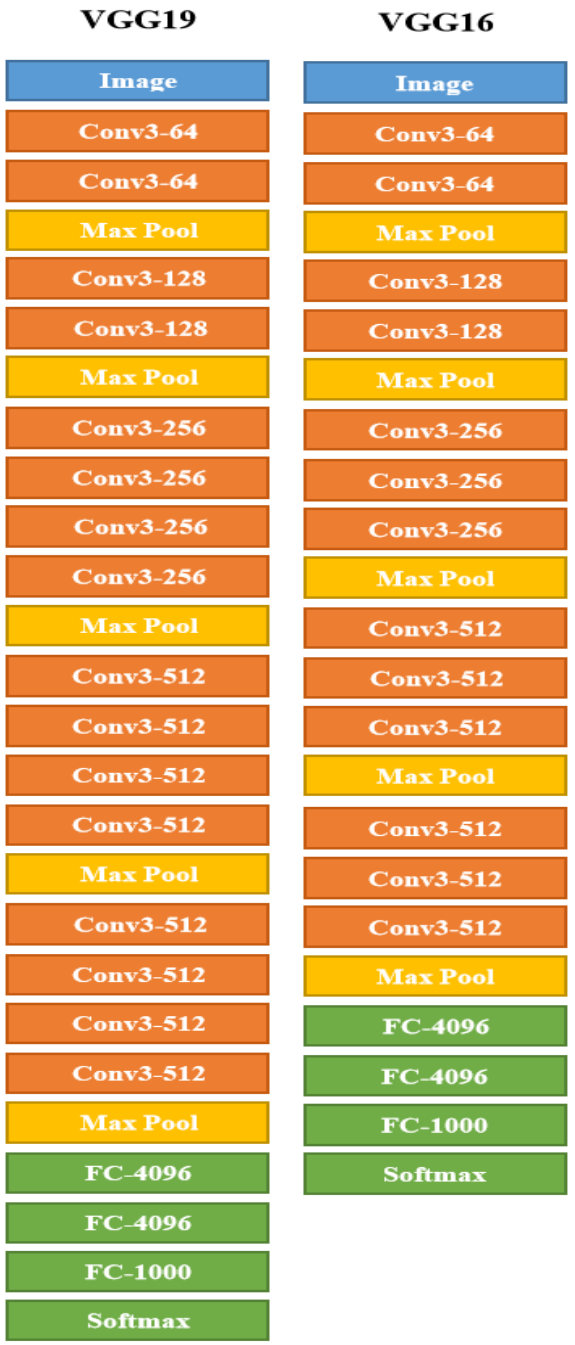
Softmax Function



$$softmax(x) = \frac{\exp(x)}{\sum_i^N \exp(x_i)}$$

VGGNet

- **Very Deep Convolutional Networks for Large-Scale Image Recognition[2013]**, VGGNet is the use of a convolution filter that is very small, only 3x3 filter when using convolution processing
- 64, 128, 256, 512 channels

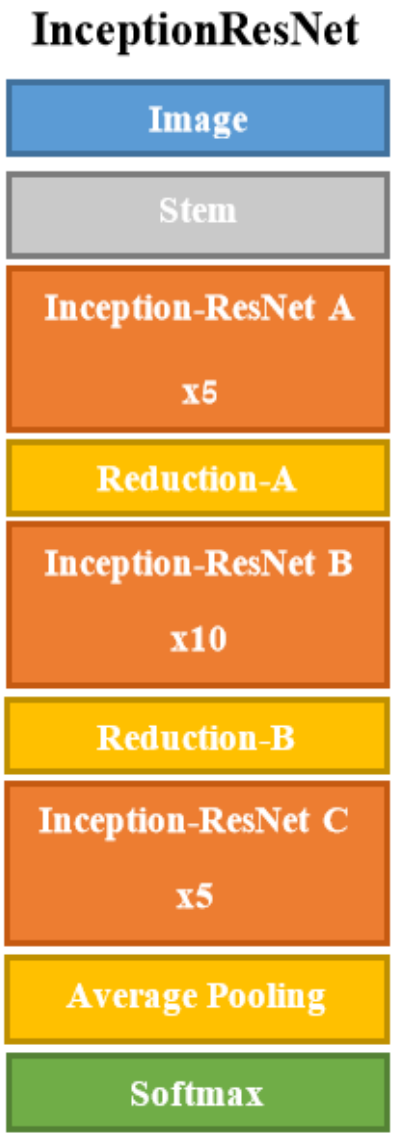


Inception-ResnetV2

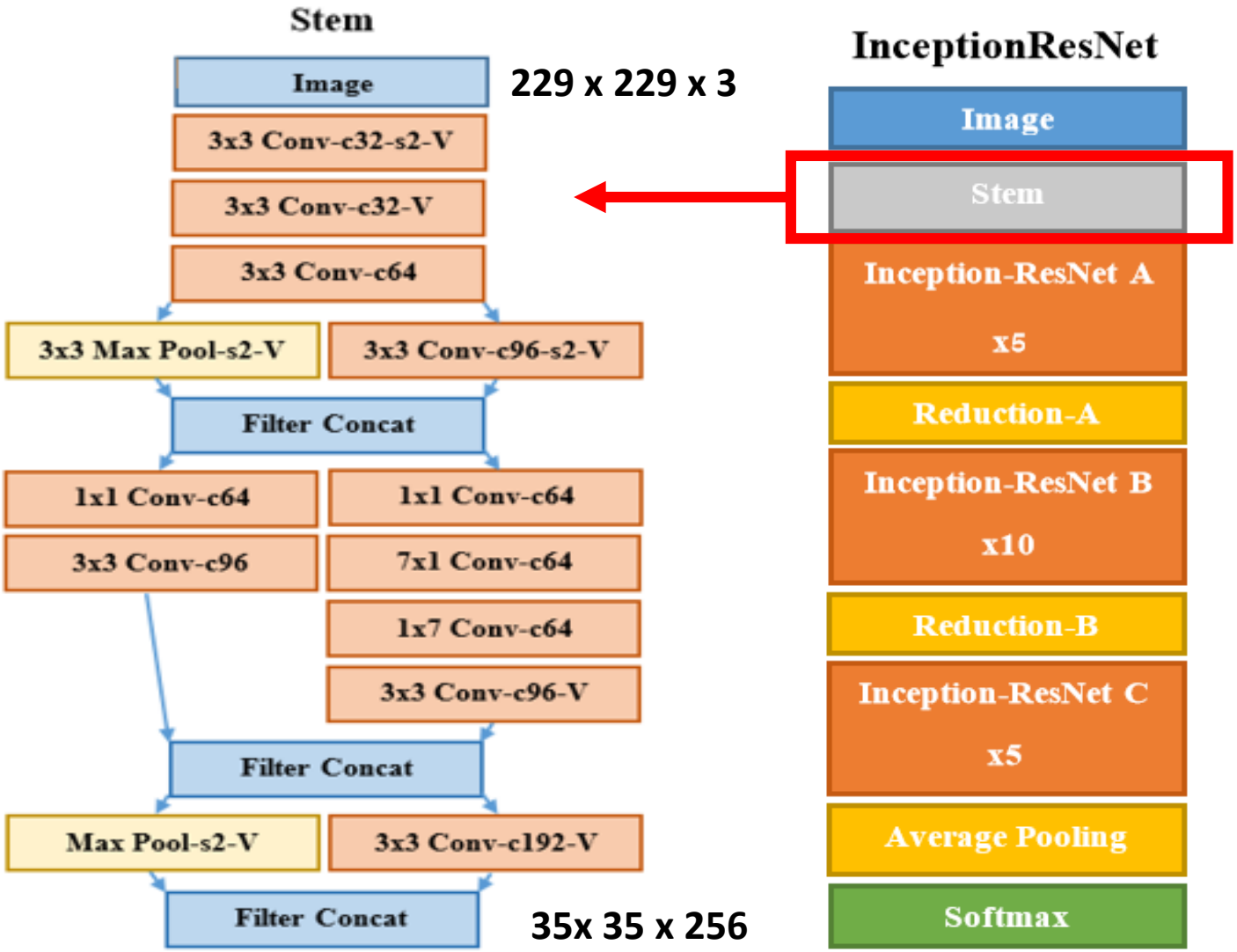
- Inception model series
  - Inception-v1 , GoogLeNet (2013)
  - Inception-v2, Batch normalization (BN)
  - Inception-v3 Factorization convolutions, such 7x1,7x1 etc
  - Inception-v4, ResNet

    Inception-v4, Inception-ResNetv2 and the Impact of Residual Connections on Learning

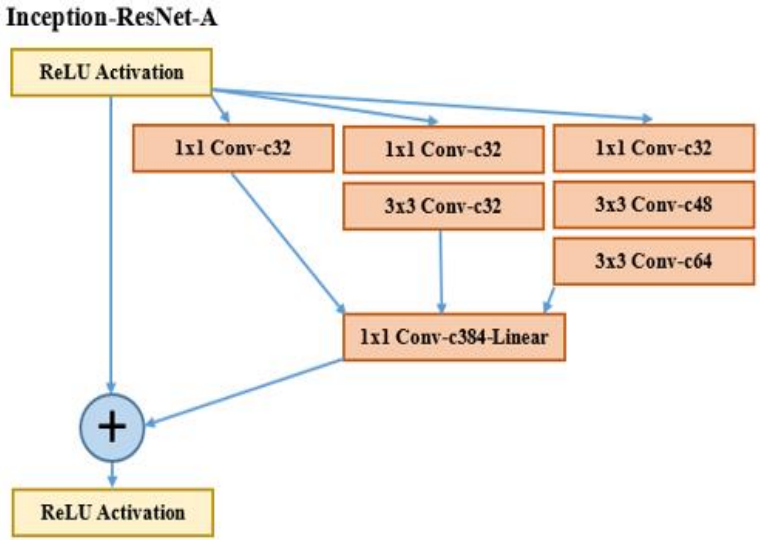
- InceptionResNetV2
  - Stem block
  - Inception-ResNet A B and C blocks
  - Reduction A and B Blocks

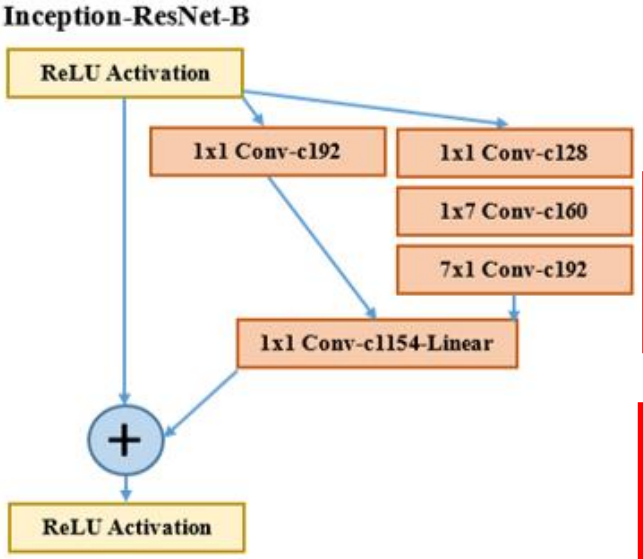


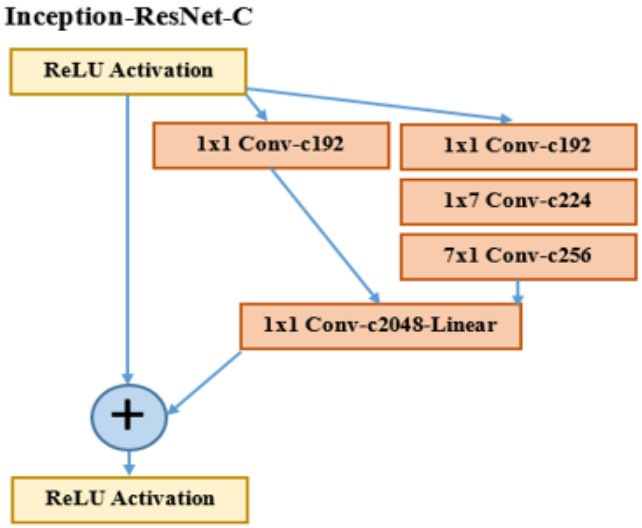
Stem block


Stem
Image
229 x 229 x 3
3x3 Conv-c32-s2-V
3x3 Conv-c32-V
3x3 Conv-c64
3x3 Max Pool-s2-V
3x3 Conv-c96-s2-V
Filter Concat
1x1 Conv-c64
1x1 Conv-c64
3x3 Conv-c96
7x1 Conv-c64
1x7 Conv-c64
3x3 Conv-c96-V
Filter Concat
Max Pool-s2-V
3x3 Conv-c192-V
Filter Concat
35x 35 x 256
InceptionResNet
Image
Stem
Inception-ResNet A
x5
Reduction-A
Inception-ResNet B
x10
Reduction-B
Inception-ResNet C
x5
Average Pooling
Softmax

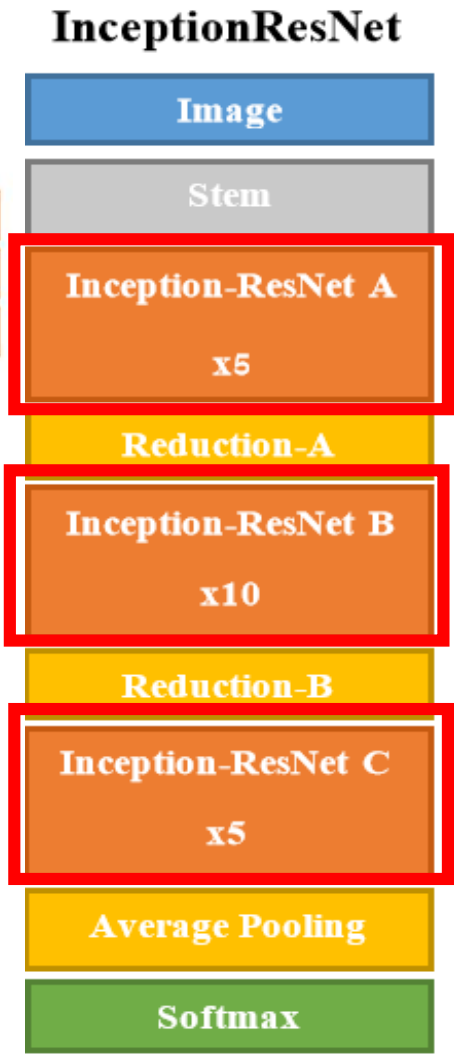
Inception-Resnet Block



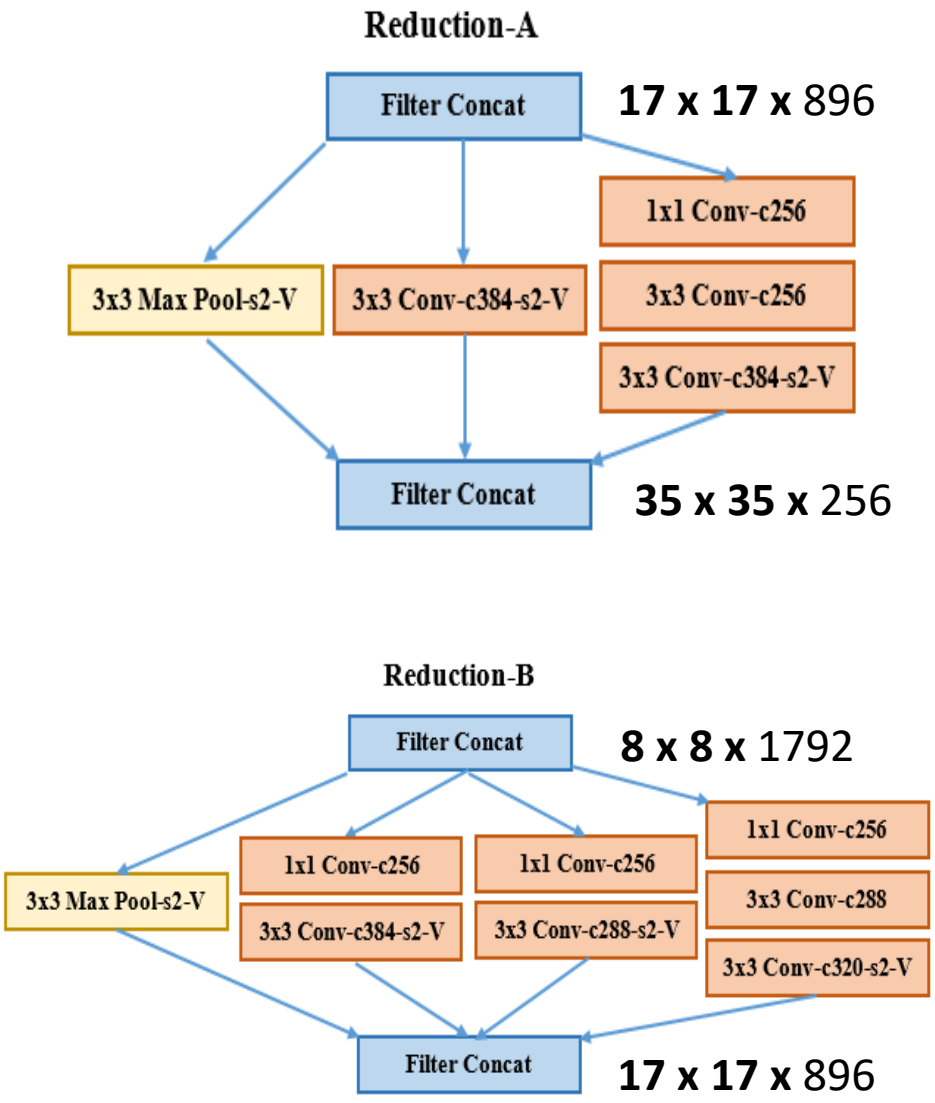


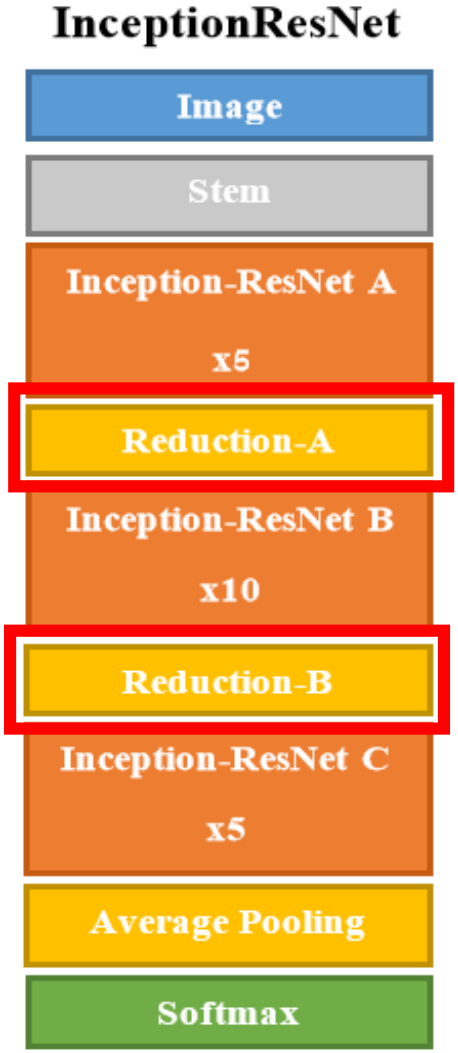


Residual Connection + Inception Block

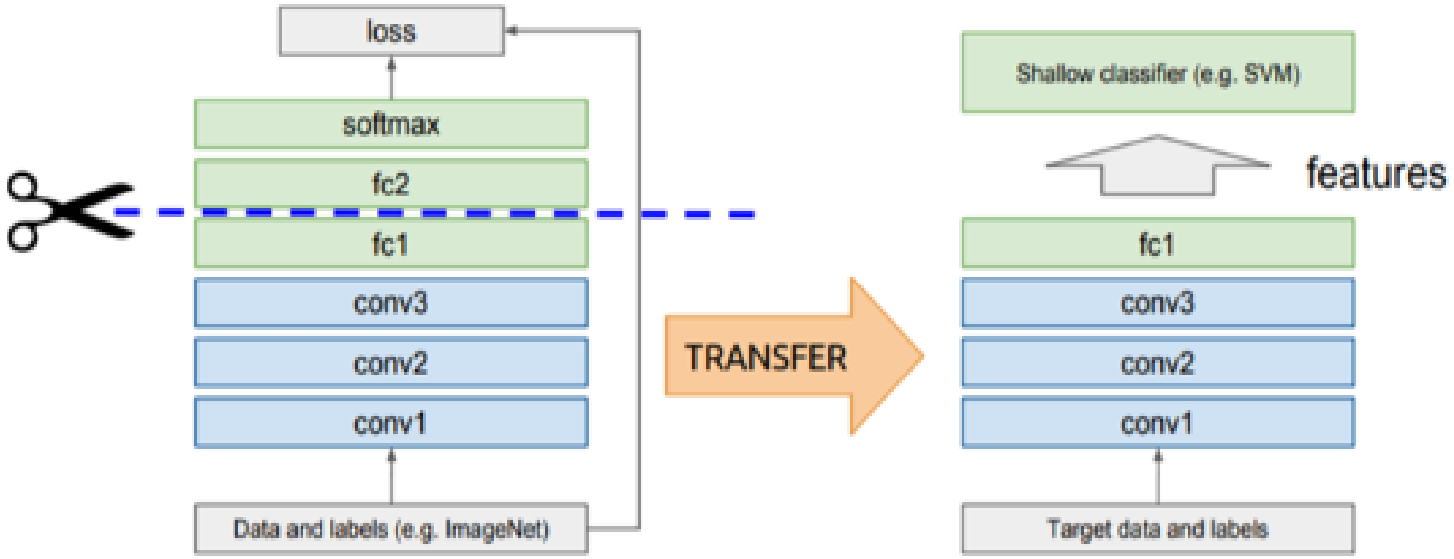


Reduction Block





Scratch and Transfer Learning



- ***Scratch Learning*** is a complex process and takes a long time to learn due to the learning beginning with creation of a random weight,
- ***Transfer learning*** is applying knowledge from previous domains that have been learned. It is called the ***Pre-trained model*** which directly results in faster training and higher effectiveness.

Thai Handwritten Characters Dataset

- **ALICE-THI dataset ,** 78 Thai characters; from150 undergraduate students, aged 20-23,

- THI-C68 14,490 characters, 68 classes including:
  - 44 consonants,
  - 17 vowels,
  - 4 tones
  - 3 symbols

Experimental Setup

- CPU 3.0 GHz, Quad-Core, GPU GeForce GTX 1080Ti, Memory 16GB
- The 5-fold and 10-fold following ratios; Train:Valid:Test , 7:1:2 and 8:1:1, respectively.
- Stochastic Gradient Decent (SGD) with momentum , learning rate 0.001, momentum 0.9 , decay learning rate 0.0001
- 100 epoch , and 32 batch size
- Pre-trained CNN Model ImageNet Dataset

Properties	CNN Models	
	VGGNet	Inception-ResNet-v2
Input Image Size (Pixel)	128x128	128x128
Memory (MB)	160.6	437.5
Parameters (M)	20	54.44
Train Times (sec/epoch)	41.00	74.18
Test Times (sec/image)	0.0014	0.0043

Experimental Result

Methods	Accuracy Rate (%)	
	10-cv	5-cv
SiftD-SVM [18]	94.34	-
HOGFoDRs-SVM [5]	-	98.76
VGGNet-Scratch	97.93 ± 0.55	96.93 ± 0.48
Inception-ResNet-Scratch	98.15 ± 0.24	97.79 ± 0.29
VGGNet-Transfer	**99.20 ± 0.27**	**98.81 ± 0.25**
Inception-ResNet-Transfer	98.88 ± 0.24	98.61 ± 0.14

- [1] Surinta, O., Karaaba, M.F., Schomaker, L.R.B. and Wiering, M.A. 2015. Recognition of handwritten characters using local gradient feature descriptors. Engineering Applications of Artificial Intelligence. 45, (Oct. 2015), 405–414.
- [2] Inkeaw, P., Bootkrajang, J., Marukatat, S., Gonçalves, T. and Chaijaruwanich, J. 2019. Recognition of Similar Characters using Gradient Features of Discriminative Regions. Expert Systems with Applications. 134, (Nov. 2019), 120–137.

Conclusions

- Effective in recognizing Thai handwritten characters (THI-C68) with a high rate of recognition.
- Experimental 2 CNN models are VGGNet-19 and Inception-ResNet-v2 architectures.
  - CNNs is higher than previous works, which are feature-based SVM
  - Transfer learning is a way to reduce learning time and increasing the efficiency of recognition.
- **VGGNet-19** architecture with transfer learning has an accuracy rate at 99.20% of Thai handwritten characters

- VGGNet-19 is an appropriate model to solve the problems of ***"Thai Handwritten Character Recognition"***

Thank you

61011262002@msu.ac.th