Categorisation images of clothes with convolutional neural networks (CNN)

Author: Taras Sereda
LeNET - 5 (Zip code recognition)

Yan LeCun
AI lab director at Facebook (NYU)

Convolution operation


Dot product between image patch and kernel shifted by bias.
Max Pooling

pooling neighboring features, reducing parameters amount.

AlexNet (ImageNet 2012 winning solution)

[CONV -> POOL] * 2 -> CONV -> CONV -> CONV -> POOL -> FC -> FC -> SOFTMAX

ReLU activation function

\[ f = \max(0,x) \]

- speeds up training
- more expressive features

GoogleNet (ImageNet 2014 winning solution)

CONV1 weights

Similar to V1 layer of visual cortex
Gathering datasets

- Nearly 800k of user photos
- ~120 classes in total
- Noize ~30-40%
- Manually pick 100 images for each class (human)
- Train CNN
- Use network for gathering data.
Finetuning for Clothes classification with Caffe

- Small learning rate: 0.001
- Batch_size: 256
- Classes: 56
- Dataset: 20k
- time: 20 hours on K520 GPU
- Amazon: g2.2xlarge
Finetuning Results

1,8k Test set accuracy

Wrong 14%
Correct 86%
Online data augmentation #0

- Random crops
- Random rotations [-50;50]
- For generalization and rotation invariance.
Online data augmentation #1

-90 degree
"accessory.glasses": "0.78519"

original image
"accessory.glasses": "1.0000"

-60 degree
"accessory.glasses": "0.90990"
Online data augmentation #2

-30 degree
"feet.sport.low trainers": "0.95052"

original image
"feet.sport.low trainers": "1.00000"

90 degree
"feet.sport.low trainers": "0.38777",
"feet.sport.high trainers": "0.27408"
Client side

Garment adding process.
Client side
Response as text + icon.
Classes Hierarchy

- entity
  - legs
  - body
  - feat
  - head
DARTS (hedging your bets)

Solving accuracy specificity tradeoff

\[
\maximize_f R(f) \\
\text{subject to } \Phi(x) \geq 1 - \epsilon
\]

[7] DARTS, hedging your bets

\( \Phi(x) \) accuracy

\( R(f) \) reward
DARTS Benefits

- get still correct predicts but less specific
- body.knitwear.cardigan is also body.knitwear
- needs only posterior probabilities on leaf nodes.
Future plans

- add more classes, move to 120 in total.
- build ansambel of specialized models
- add image segmentation
- apply attention concepts for image captioning
- move towards online learning
Thanks!

Any questions?

taras.y.sereda@gmail.com