Cross-stitch Networks for Multi-task Learning

Ishan Misra*, Abhinav Shrivastava*, Abhinav Gupta, Martial Hebert
Multiple supervision

Input

Labels

Surface Normals

Semantic Segmentation

Surface orientation

Semantic labels
“Free” supervision

Input

Labels

Surface Normals

Semantic Segmentation

Surface orientation

Semantic labels

Using Kinect: Free

Manually marked: Expensive
No. of pixels per segmentation class
No. of pixels per segmentation class

Classes with less data

- bag
- lamp
- paper
- floor-mat
- mirror
- counter
- other furniture
- chair
- wall

No. of pixel labels (millions)
No. of pixels per segmentation class

Classes with less data

- bag
- lamp
- paper
- floor-mat
- mirror
- counter
- other furniture
- chair
- wall

No. of pixel labels (millions)
Multi-task: Shared network

Input

Task A

Task B

Complete sharing of parameters between tasks

- Gray: Shared Layers
- Orange: Task A layers
- Green: Task B layers
Split at a lower layer

Input

Task A

Task B

Split here?

- Gray: Shared Layers
- Orange: Task A layers
- Green: Task B layers
Split at a lower layer

Input

Task A

Task B

- Shared Layers
- Task A layers
- Task B layers
Train and Test **all** of them
Standard Multi-task Learning

Relative performance wrt. one-task network

Better
Relative Performance
Worse

Surface Normal
Segmentation
Pick the best

Relative performance wrt. one-task network

Better

Relative Performance

Worse

Surface Normal

Segmentation

Generic Network
All Parameters Shared

Specific Network
No Parameters Shared
New Tasks?

Input

Labels

Attributes

Has saddle

Four legs

Detection

Object location
Training and measuring performance

Generic Network: All Parameters Shared

Specific Network: No Parameters Shared

Relative performance wrt. one-task network

Better

Relative Performance

Worse

Attributes
Detection
No single best network

Best for Detection

Best for Attributes

Relative performance wrt. one-task network

Attributes

Detection

Better

Worse
Problems with standard approach
Problems

Given a new set of tasks:

• Enumerate many ConvNet architectures

No principled way of exploring
Problems

Given a new set of tasks:
• Enumerate many ConvNet architectures
• Train all of them

Practically Expensive
Problems

Given a new set of tasks:
• Enumerate many ConvNet architectures
• Train all of them
• Pick the best one for your tasks

Best architecture depends on tasks
Problems

Given a new set of tasks:

- Enumerate many ConvNet architectures
- Train all of them
- Pick the best one for your tasks

One architecture may not perform well on all tasks
Our contribution

• Principled exploration of architectures
  • *Without* training all of them
Our contribution

• Principled exploration of architectures
  • Without training all of them
• Finds network architecture automatically
  • Generalizes across tasks
Our contribution

- Principled exploration of architectures
  - *Without* training all of them
- Finds network architecture automatically
  - Generalizes across tasks
- Implicitly explores **more architectures** than brute-force
- **Performs better** than architectures found via brute-force
Two one-task networks

Task A:
conv1, pool1 → conv2, pool2 → conv3 → conv4 → conv5, pool5 → fc6 → fc7 → fc8

Task B:
Image → Image → Image → Image → Image → Image → Image
Idiom: A stitch in time saves nine
Our proposed unit: Cross-stitch
Learns a linear combination
Results
Surface Normals (Median Error) Lower is better

- One-Task network: 19.0
- Best split architecture (Brute-force search): 18.6
- Cross-stitch: 18.2

Segmentation (mean IU) Higher is better

- One-Task network: 18.4
- Best split architecture (Brute-force search): 19.2
- Cross-stitch: 19.3

Attributes (mean AP)

- One-Task network: 60.9
- Best split architecture (Brute-force search): 59.7
- Cross-stitch: 63.0

Detection (mean AP)

- One-Task network: 44.9
- Best split architecture (Brute-force search): 45.0
- Cross-stitch: 45.2
Data starved segmentation
Data starved segmentation

- bag
- lamp
- paper
- floor-mat
- mirror
- counter
- other
- furniture
- chair
- wall

No. of pixel labels (millions)

Change in performance
Data starved Attributes

+4.6% mAP for bottom 10 attributes
+4.3% mAP for bottom 20 attributes
Cross-stitch Networks

Poster # 22
Session 3-2 (4:45pm-6:45pm)