Exploiting Convolutional Filter Patterns for Transfer Learning

Chris Blythe
What Is Transfer Learning? - Intuitively

- Transfer learned knowledge to new conditions
Why Is Transfer Learning Important?

- Deep training requires lots of time and data
- Labeled data is expensive
- Recycling previously learned knowledge
Exploiting CNN Filter Patterns - Intro

- **Standard Loss**

\[
\min_w \sum_n L(z(x_n, w), y_n)
\]
Exploiting CNN Filter Patterns - Intro

- Are Filters separable into clusters?
- How similar are filters inside a cluster?
- How filters are distributed over clusters?
Exploiting CNN Filter Patterns - Intro

- Multivariate Gaussian Distribution

\[ \mathcal{N}(w; \mu, \Sigma) = \frac{1}{(2\pi|\Sigma|)^{-\frac{1}{2}}} \exp\left(-\frac{1}{2}(w-\mu)^T \Sigma^{-1} (w-\mu)\right) \]

- Gaussian Mixture Model

\[ P(w|\pi, \mu, \Sigma) = \sum_{k=1}^{K} \pi_k \mathcal{N}(w|\mu_k, \Sigma_k) \]
Exploiting CNN Filter Patterns – Statistical Analysis

- Analyzed known CNN
- 3x3 Filters
- K-Means Clustering
  - $K=10$
Exploiting CNN Filter Patterns – Statistical Analysis
Exploiting CNN Filter Patterns – Statistical Analysis

Alexnet - Imagenet
Alexnet - Places
VGG16 - Imagenet
VGG16 - Places
Exploiting CNN Filter Patterns - Goals

- Goal: Transfer filter relational statistics
  - Better representations quickly
- “What makes a Good Filter?”
  - Proven filter pattern distributions
Exploiting CNN Filter Patterns - Goals

Well trained Network

New Network

The distribution of network is learned with GMM

The new network is forced to show similar distribution with regularization param.
Exploiting CNN Filter Patterns - Method

- Regularization
  \[ R(w_i) = -\log(P(w_i)) \]

- Total Regularization
  \[ R(w|\mu, \Sigma) = \sum_{i=1}^{N} -\log \left( \sum_{k=1}^{K} \pi_k N(w_i|\mu_k, \Sigma_k) \right) \]
Exploiting CNN Filter Patterns - Method

- Derivative of likelihood expensive to calculate
- Approximate derivative by max contribution
- Simplified Derivative: \( \frac{\partial R}{\partial w_i} = (w_i - \mu_s)\Sigma_s^{-1} \)
Exploiting CNN Filter Patterns - Method

- Completed loss term:

\[
\min_w \sum_n L(z(x_n, w), y_n) + \sum \alpha \frac{1}{2} w^2 + \sum \lambda \cdot R(w)
\]

- Standard Loss
- Weight Decay
- Regularization
Exploiting CNN Filter Patterns - Generizability

- Reg vs No Reg
- CIFAR-10
- Frozen Layers
  - 10k, 15k, 20k, 25k
Exploiting CNN Filter Patterns – Task Transfer

- VGGF-F
  - Trained ImageNet
  - Finetune Places2
- Frozen layers
  - 10k, 25k, 50k
- Transfer = Less Training
Exploiting CNN Filter Patterns – Cross-Modal

- One Scene
- Cross Modal Representations
- Modal Alignment
Exploiting CNN Filter Patterns – Cross-Modal

- Mode Networks
- Specific Low level
- Shared Layers
- Cross-Modal

High level features
Exploiting CNN Filter Patterns – Cross-Modal

- VGGF-F
  - Trained ImageNet
  - Finetune Clip Art
  - Finetune Sketches
Exploiting CNN Filter Patterns – Conclusion

- Showed statistical patterns in known networks
- Capturing patterns with GMM + Regularization term
- Better representations in less time
  - Generalized
  - Task Transferrable
  - Cross-Mode Transferrable