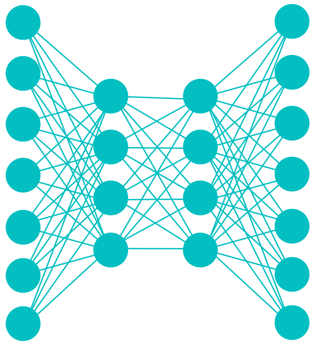
Lecture Notes for

Neural Networks and Machine Learning

Multi-Task and Demo



Logistics and Agenda

- Logistics
  - Schedule Updated!
- Agenda
  - Multi-Task Examples
  - Multi-Task Demos
  - Multi-Task Town Hall
  - *if time: AlphaCode*
- Next Time
  - Circuits

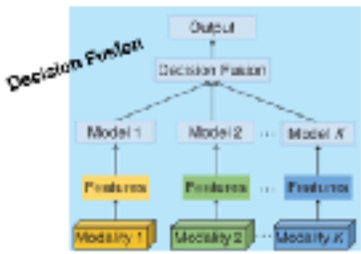
Remaining Presentations

Student		Total Assigned	Responded?	Data Augmentation	Group Normalization	Perceptual Losses	Radford GANs	AlphaFold
Barajas, Cynthia	[illegible] Sect 4	0	1	x		x		
Carter, Austin	[illegible] Sect 4	1	1			1	0	
Collins, Joel	[illegible] Sect 4	1	1					1
Du, Jinyu	[illegible] Sect 0	1	1	1			0	
Ebrahimian, Jonathan	[illegible] Sect P	1	1				1	
Emami, Hessam	[illegible] Sect 4	1	1	1			0	
Gage, Nathan	[illegible] Sect 0	1	1					
Gao, Qing	[illegible] Sect 4	1	1	1			0	
Geng, Zicheng	[illegible] Sect 0	1	1	1			0	
Havard, Andrew	[illegible] Sect P	1	1				1	
Hu, Yvon	[illegible] Sect 0	1	1	0				
Klinkert, Jake	[illegible] Sect 0	1	1		1		0	
Larsen, Nicholas	[illegible] Sect 4	1	1					
Larsen, Steven	[illegible] Sect 4	1	1					
Lazaro, Irvin	[illegible] Sect 4	0	1	x		x		
Lu, Yifan	[illegible] Sect 4	1	1	0	1			
vage, Ayesh Madushanka	[illegible] Sect 0	1	1					1
McNitt, Troy	[illegible] Sect P	1	1					
Moros, Jonas	[illegible] Sect 0	1	1					
Rajapandian, Khoushik	[illegible] Sect 4	1	1				x	x
Rosenblatt, Jack	[illegible] Sect 0	1	1	0				1
Srirama, Nathan	[illegible] Sect 0	1	1	0			0	
Tsai, Amor	[illegible] Sect 0	1	1	0				
Wall, Nick	1321 Sect 0	0	0					
Wang, Kuo	[illegible] Sect 0	1	1					
Yang, Chenyu	[illegible] Sect 4	0	1	x				
Yassien, Sam	1321 Sect 0	1	1			1		
Yu, Hongjin	[illegible] Sect 0	1	1			0	0	
Zepede, Juan	[illegible] Sect 4	0	1	x	x			
			Num Assigned	4	2	2	2	3

Last Time

- **Early Fusion:** Merge sensor layers early in the process
- **Assumption:** there is some data redundancy, but modes are conditionally independent
- **Problem**: architecture parameter explosion
  - Need dimensionality reduction

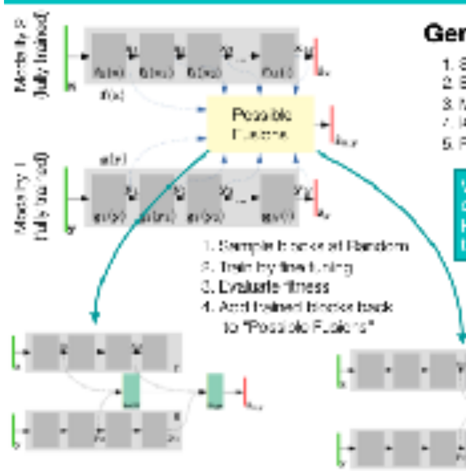
- **Late Fusion**: Merge sensor layers right before flattening
- Use Decision Fusion on outputs
- **Assumption**: little redundancy or conditional independence—just an ensemble architecture
- **Problem**: just separate classifiers, limited interplay



Ramachandram and Taylor, 2017

80

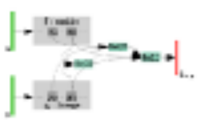
Neural Architecture Search for Mode Fusion



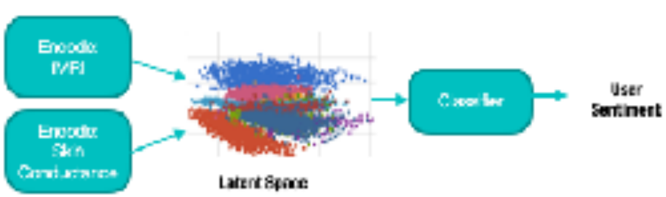
Genetic Algorithm

1. Sample new candidates
2. Evaluate fitnesses
3. Mutate and Crossover
4. Keep the best solutions
5. Repeat

Very computational when starting, because candidates are all untrained. However, as more blocks start from "mostly trained" positions, training time reduces.


Found solution for AV-MNIST

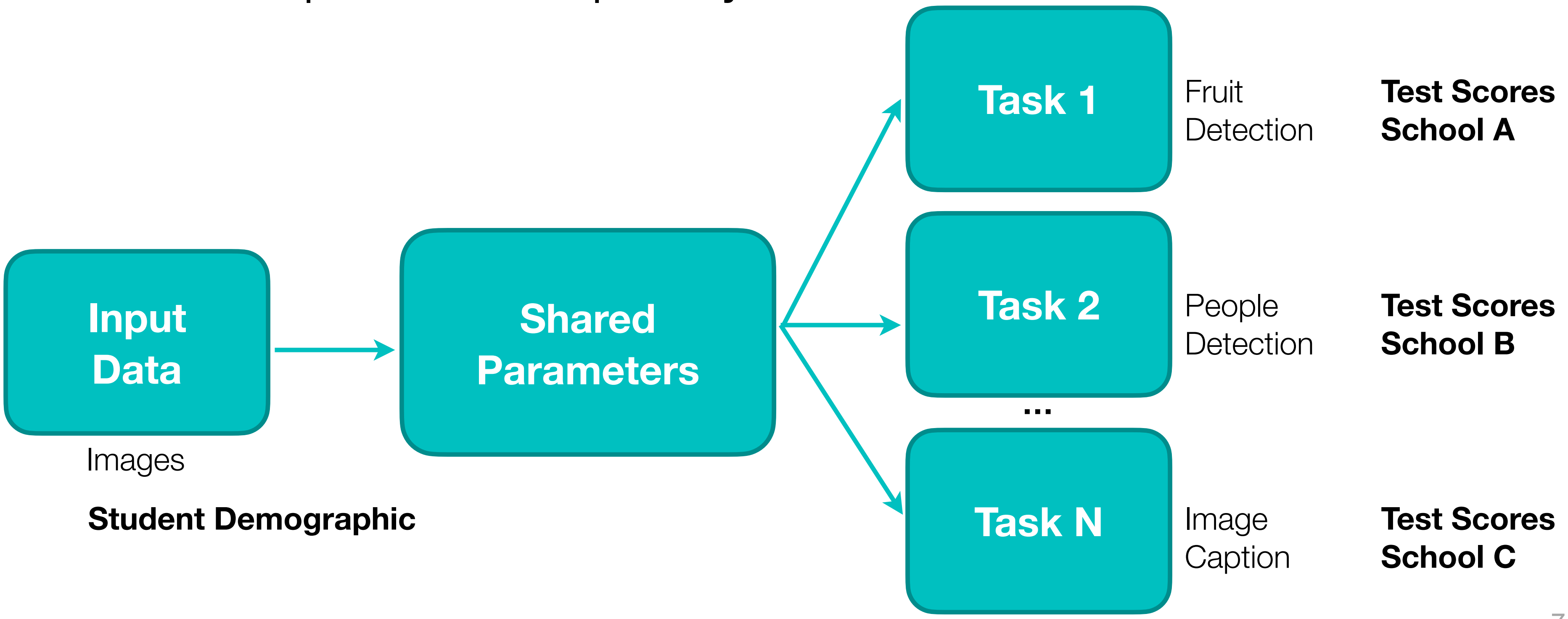
- Latent Space Transfer (universality)
  - From another domain, map to a similar latent space for the same task
  - Useful for unifying data based upon a new input mode when old mode is well understood
    - for example, biometric data
    - **I have never seen a research paper on this…**



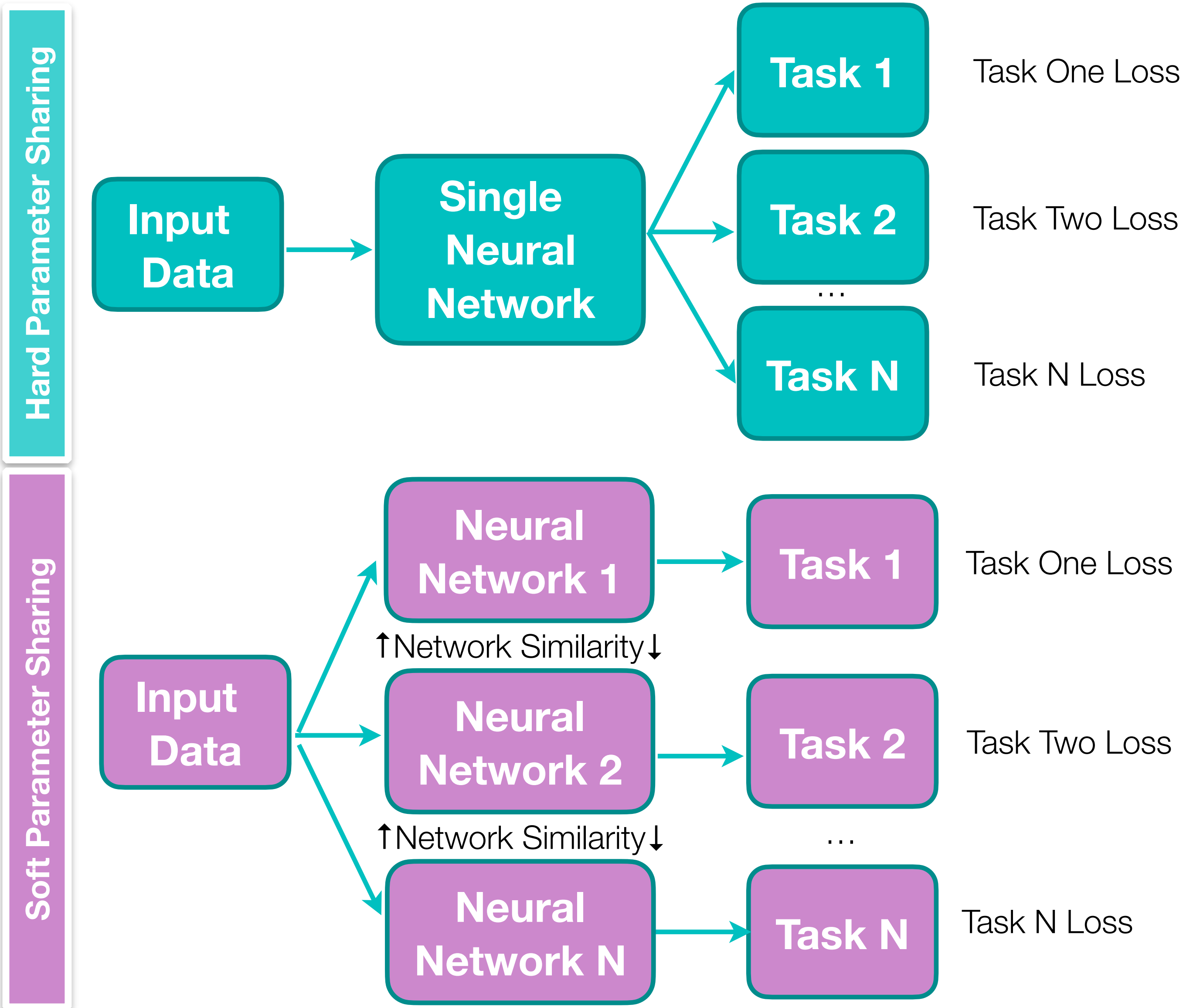
Multi-Task Models

Multi-task learning overview

- For deep networks, simple idea: share parameters in early layers
- Used shared parameters as feature extractors
- Train separate, unique layers for each task



Multi-task Learning Parameter Sharing

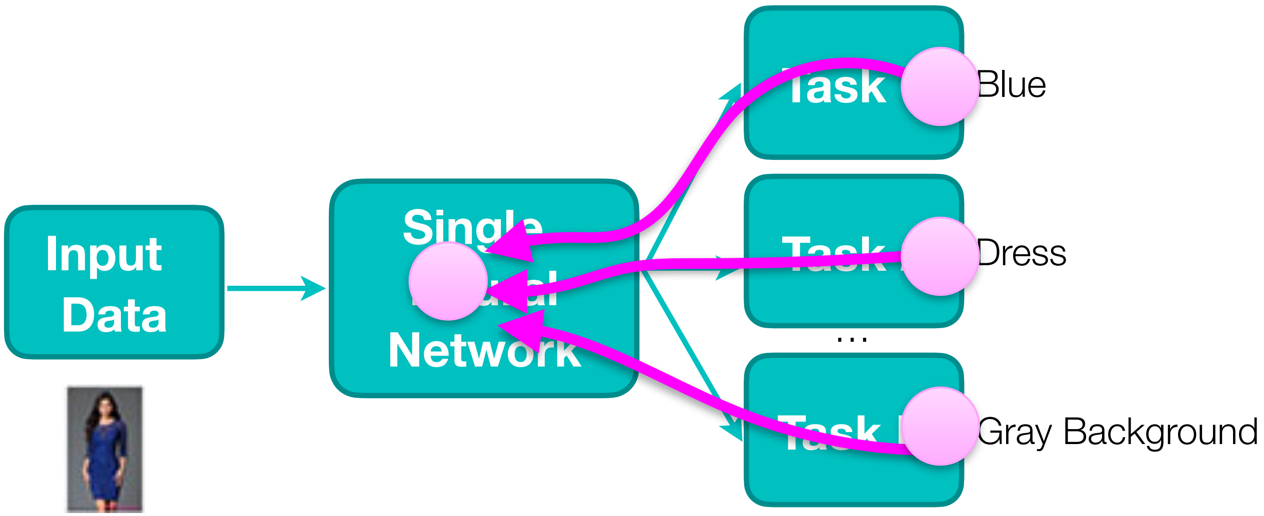


Pool Losses
Over Multiple Batches
From Multiple Tasks,
Update via BackProp

Pool Losses
Over Multiple Batches
From Multiple Tasks,
Add Intra-Network
Similarity Loss
Update via BackProp

Multi-task Optimization

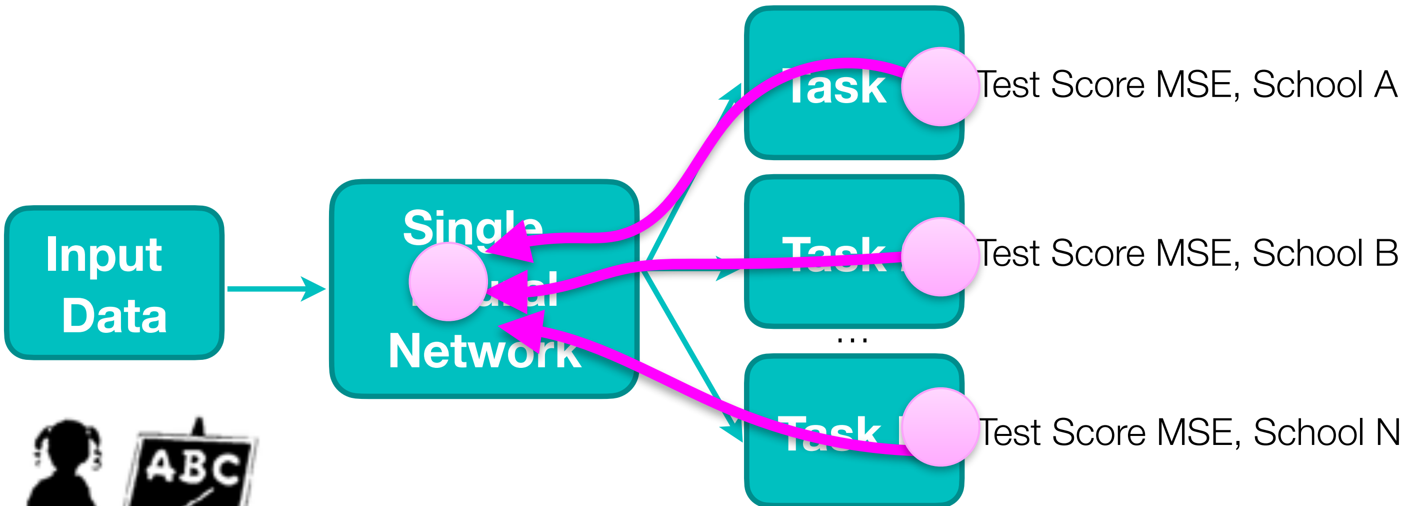
Multi-Label per Input



Measure Loss for each label simultaneously
Back propagate everything at one for a given batch

Multi-task Optimization

Single Task Label per Input



Method One: Batch updates across multiple tasks
need to perform customized gradient calculations

Method Two: Update small batches using a random task
easier, but can cause instability in training

Multi-Task Learning in Keras with Multi-Label Data

Fashion week, colors and dresses

Optional Demo

Follow Along: https://www.pyimagesearch.com/2018/06/04/keras-multiple-outputs-and-multiple-losses/

Multi-Task Learning

School Data, Computer Surveys

Traian Pop

Luke Wood

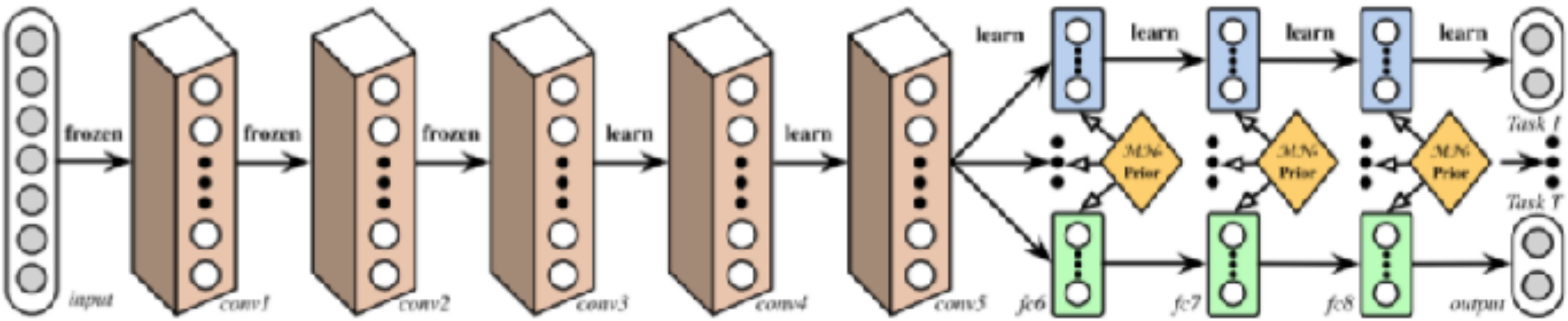
Follow Along: LectureNotesMaster/
03 LectureMultiTask.ipynb

Multi-Task Model Examples

> He uses statistics like a drunken man uses a lamp post, more for support than illumination.

-- Andrew Lang

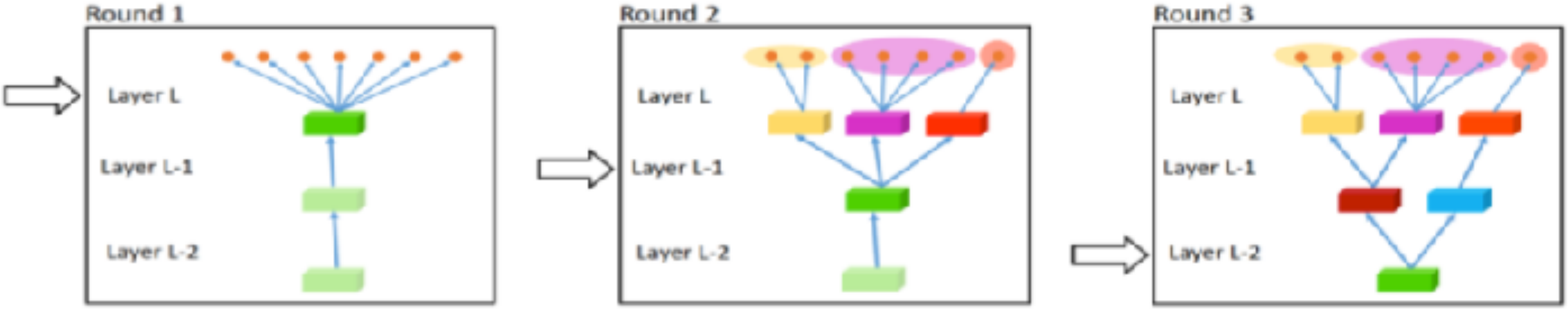
Multi-task: Deep Relationship Networks



- Start training traditionally (CCE)
- Minimize Kroenecker Product of fully connected task specific layers (here matrices are vectorized and therefore it is an outer product)
    - intuitively: make Covariances between tasks close to a given prototype Covariance
    - encourages feature maps in each task to be **less correlated** to feature maps of another task

https://arxiv.org/pdf/1506.02117.pdf

Multi-task: Adaptive Feature Sharing



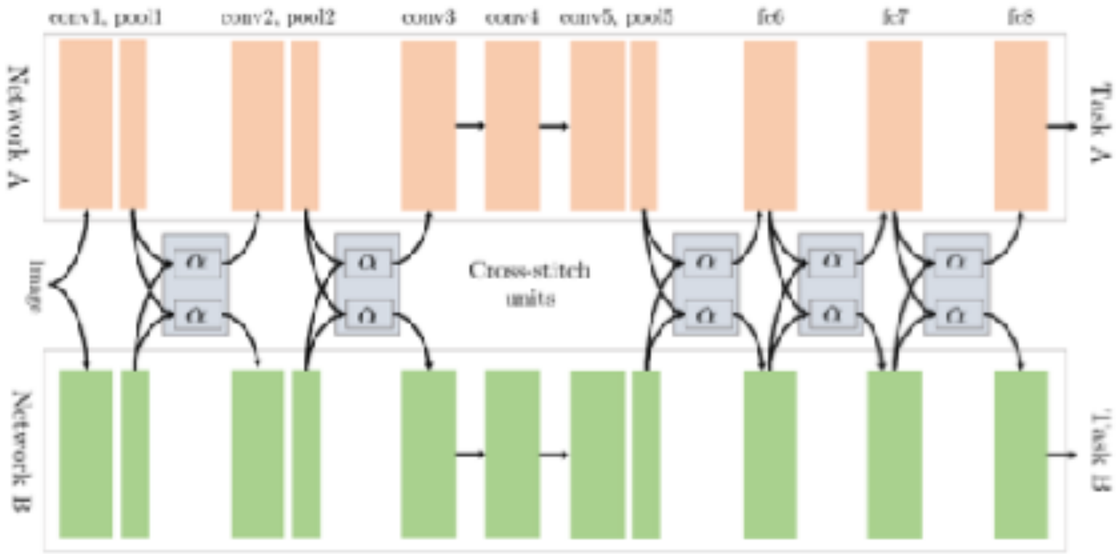
- Trai
- Rep
  - D
    - cluster affinity of branch if not final layer
  - Cut weights and retrain (fine tune) network
  - Decrement current layer index

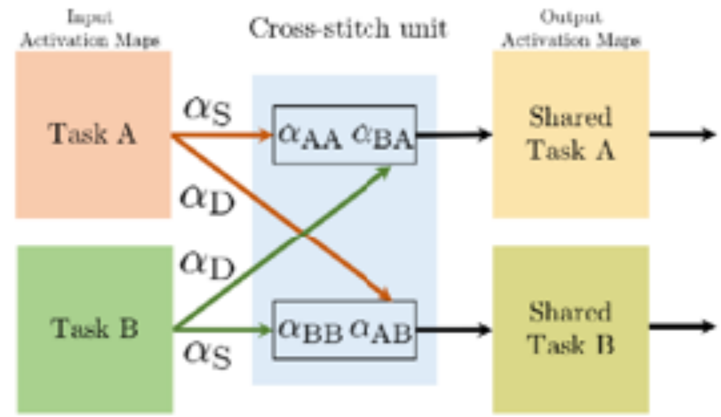
$$A^\star, \omega^\star(l) = \underset{A \in \mathbb{R}^{d \times d'}, |\omega| = d'}{\arg\min} ||W^{p,l} - AW^{p,l}_{\omega:}||_F, \qquad (2)$$

where $W^{p,l}_{\omega:}$ is a truncated weight matrix that only keeps the rows indexed by the set ω. This problem is NP-hard, however, there exist approaches based on convex relaxation

http://openaccess.thecvf.com/content_cvpr_2017/papers/Lu_Fully-Adaptive_Feature_Sharing_CVPR_2017_paper.pdf

Multi-task: Cross Stitch Networks





$$\begin{bmatrix} \tilde{x}_A^{ij} \\ \tilde{x}_B^{ij} \end{bmatrix} = \begin{bmatrix} \alpha_{AA} & \alpha_{AB} \\ \alpha_{BA} & \alpha_{BB} \end{bmatrix} \begin{bmatrix} x_A^{ij} \\ x_B^{ij} \end{bmatrix}$$

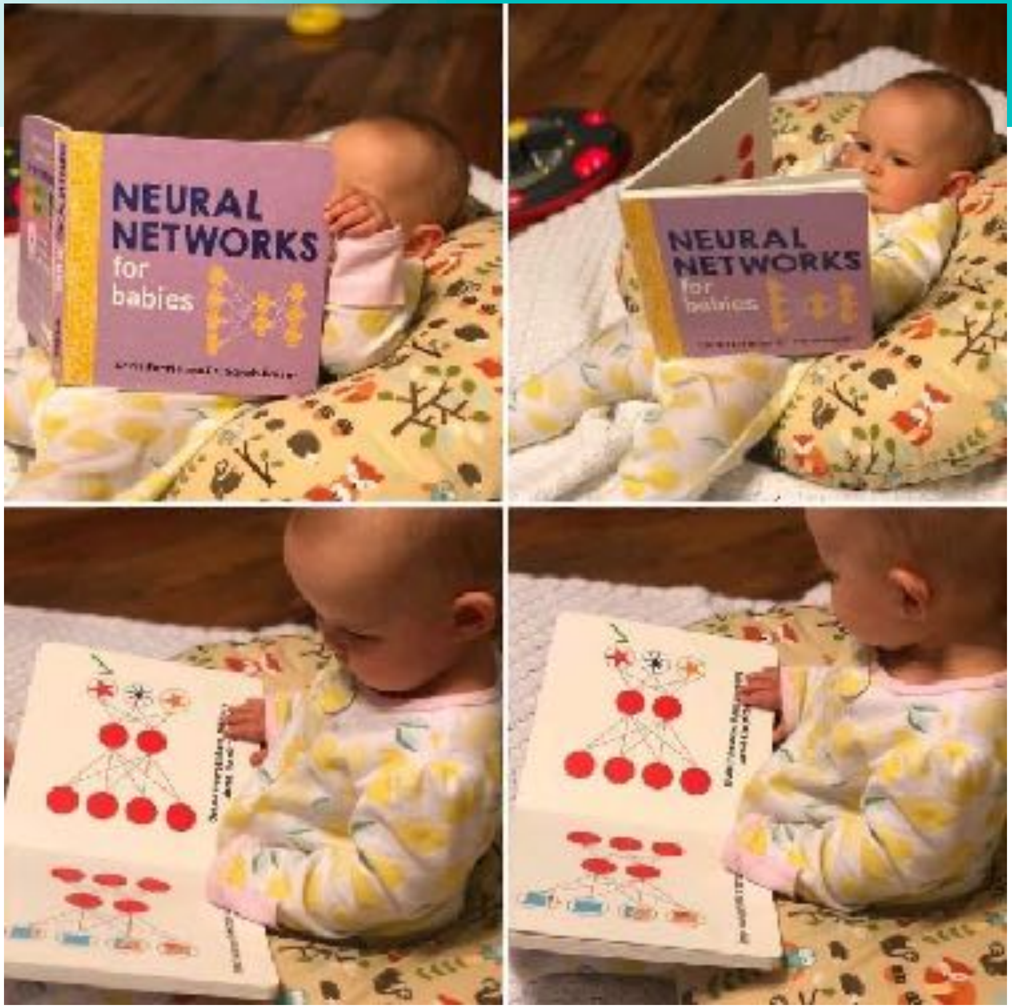
- Only works for simultaneous multi-label problems
    - like semantic segmentation and surface normal segmentation (clustering similarly facing objects)
- Take a learned weighted sum of the activations
- Works a little better than single task, but no worse

https://arxiv.org/pdf/1604.03539.pdf

Current Multi-task Research

- Incredibly diverse sets of solutions
- Mostly not evaluated on similar datasets
- Reasoning given is mostly ad-hoc…
- Theory is wildly under developed
    - because the problem is incredibly difficult
- Neural architecture search is an option…

Lab One Town Hall



Multi-Task Networks
Multi-Modal Networks

Lecture Notes for

Neural Networks and Machine Learning

Demo Multi-Task

Next Time:
Circuits

Reading: Chollet 8.1-8.5