Efficient Computer Vision

Danna Gurari

University of Colorado Boulder Fall 2024



https://dannagurari.colorado.edu/course/recent-advances-in-computer-vision-fall-2024/

Review

- Previous lectures:
 - Student-led lectures
- Assignments:
 - Project presentation poster due in 1 week
 - Project presentation due in 1.5 weeks
 - Peer evaluation due in 1.5 weeks (in-class activity)
 - Project report due in 2.5 weeks
- Questions?

Efficient Computer Vision

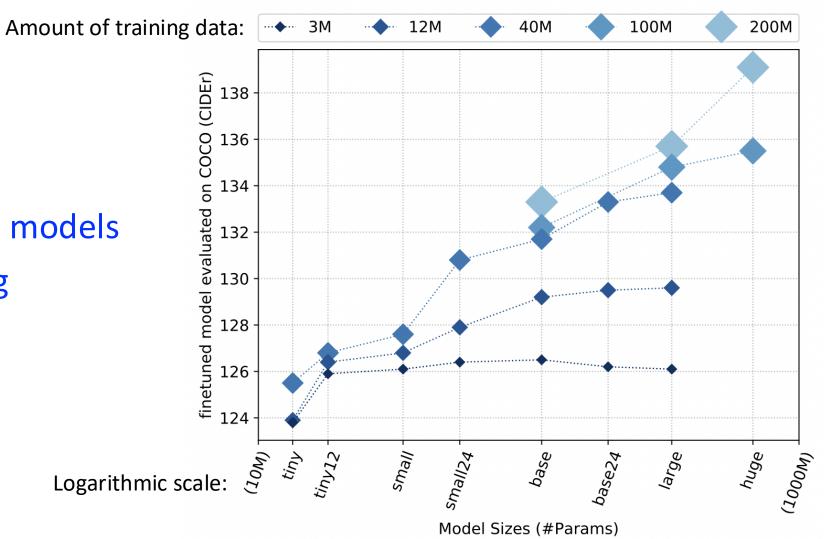
- Motivation
- Model Compression
- Curriculum Learning
- Active Learning
- Faculty Course Questionnaire (FCQ)

Efficient Computer Vision

- Motivation
- Model Compression
- Curriculum Learning
- Active Learning
- Faculty Course Questionnaire (FCQ)

What Are Common Trends for CV Models?

- 1. Parameter-heavy models
- 2. Extensive training



Hu et al. Scaling Up Vision-Language Pre-training for Image Captioning. CVPR 2022

Trend: Parameter-Heavy Models

How many parameters are estimated to be in GPT-4 (was used for ChatGPT)?

- (a) 176 million
- (b) 1.76 billion
- (c) 17.6 billion
- (d) 170.6 billion
- (e) 1.76 trillion

https://en.wikipedia.org/wiki/GPT-4

Trend: Extensive Training

How many training examples led to top performance in Vision Transformers?

- (a) 3 million
- (b) 30 million
- (c) 300 million
- (d) 3 billion
- (e) 30 billion

It took 2,500 TPUv3- core-days to train this model

Zhai et al. Scaling Vision Transformers. CVPR 2022

Modern Neural Networks Are a Mismatch for Many Real-World Applications

- Time-consuming (e.g., incompatible for real-time applications)



Boss: What did you do last month?

You: Trained the model for one epoch.





Boss: Umm, fine, what is your plan for next month?

You: Train... train the model for one more epoch?





https://hanlab.mit.edu/files/course/slides/MIT-TinyML-Lec13-Distributed-Training-I.pdf

Modern Neural Networks Are a Mismatch for Many Real-World Applications

- Time-consuming (e.g., incompatible for real-time applications)
- Large memory footprint (e.g., incompatible with edge devices)



https://www.ephotozine.com/article/19

-things-to-look-out-for-in-a-smartphone-

camera--31055

https://en.wikipedia.org/ wiki/Wearable_technolo gy



https://www.buzzfeednews.com/article/kat ienotopoulos/facebook-is-making-cameraglasses-ha-ha-oh-no



https://aws.amazon.com/blogs/machine-learning/demystifyingmachine-learning-at-the-edge-through-real-use-cases/ Modern Neural Networks Are a Mismatch for Many Real-World Applications

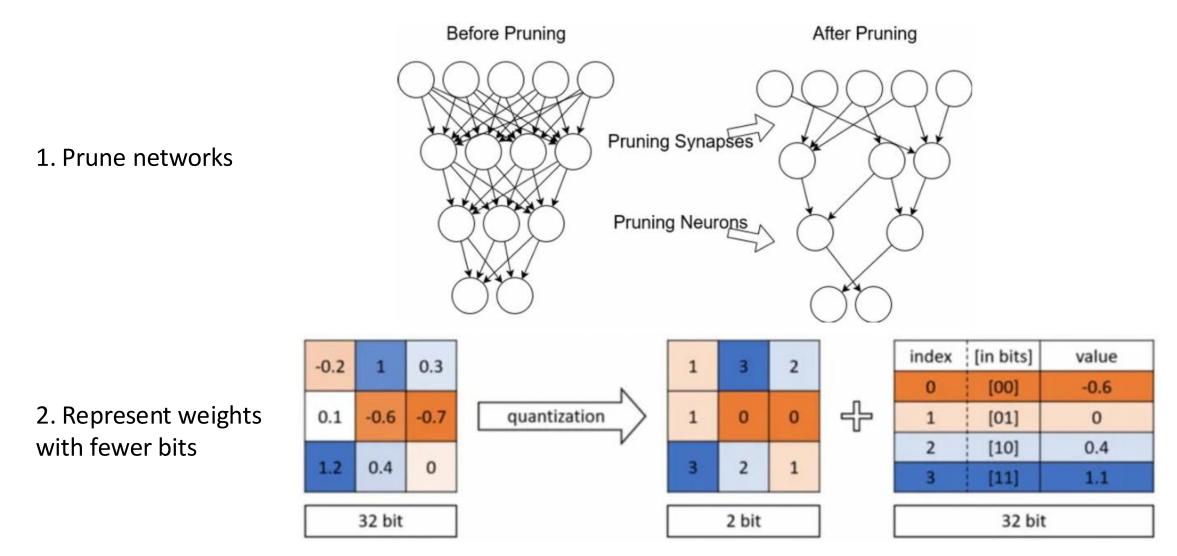
- Time-consuming (e.g., incompatible for real-time applications)
- Large memory footprint (e.g., incompatible with edge devices)
- Large computational cost (e.g., large environmental costs)

Idea: develop models that are more compact and learn more efficiently (i.e., faster and with less data)

Efficient Computer Vision

- Motivation
- Model Compression
- Curriculum Learning
- Active Learning
- Faculty Course Questionnaire (FCQ)

Dated Compression Approaches



https://xailient.com/blog/4-popular-model-compression-techniques-explained/

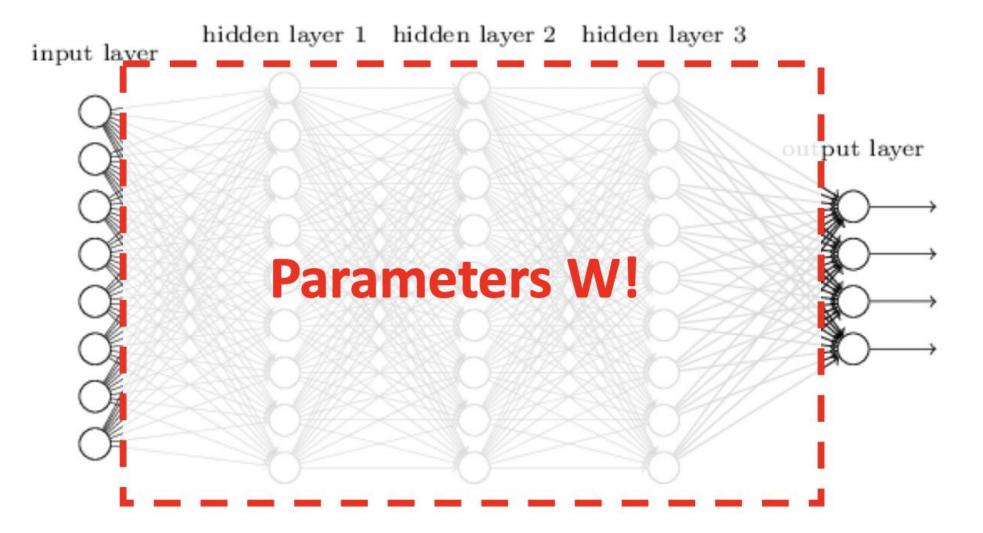
Today's Popular Approach: Knowledge Distillation



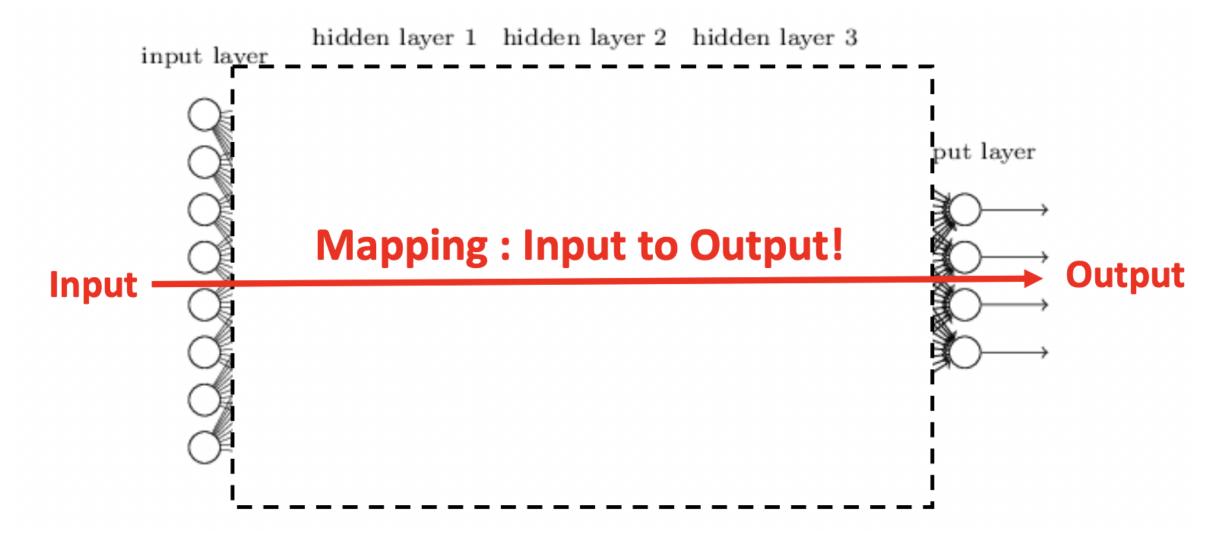
A student learns from a knowledgeable teacher

Image source: https://www.waterford.org/education/teacher-student-relationships/

Key Question: What is Knowledge?

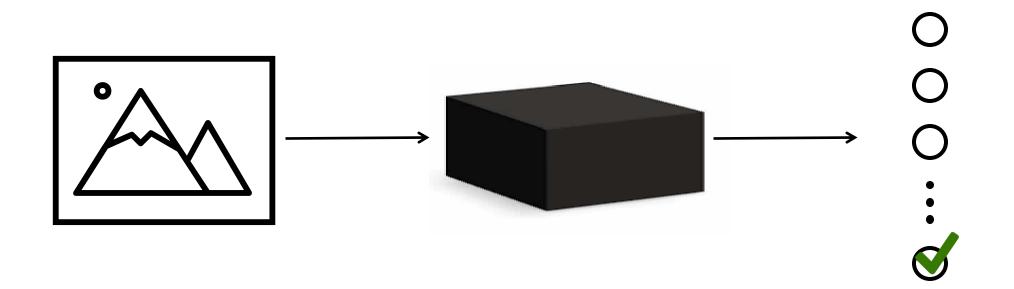


http://ir.hit.edu.cn/~xiachongfeng/slides/Knowledge%20Distillation.pdf

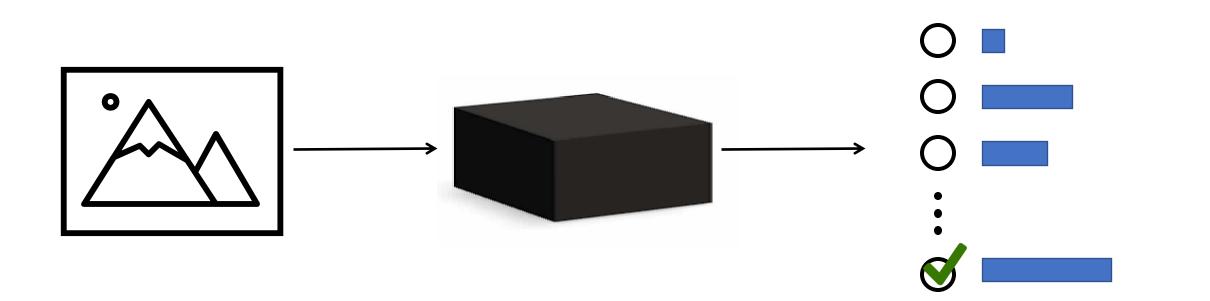


http://ir.hit.edu.cn/~xiachongfeng/slides/Knowledge%20Distillation.pdf

Target mapping: ground truth (1-hot vector)

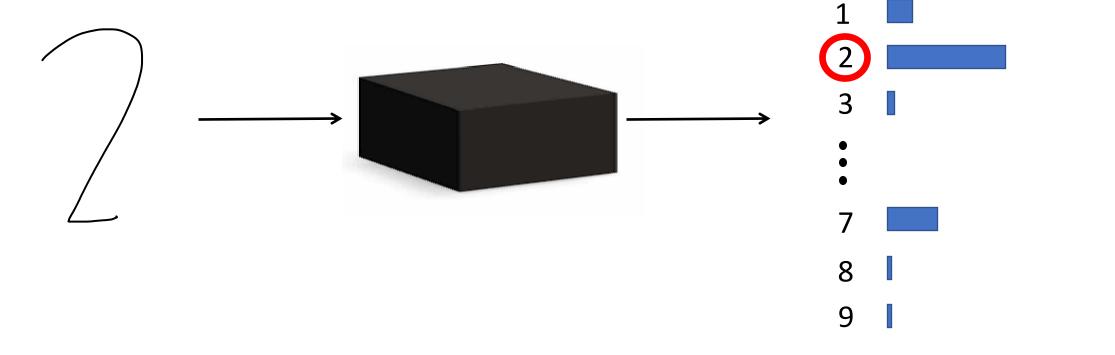


Target mapping: probability distribution from a model offers further insights into similarities and differences of categories



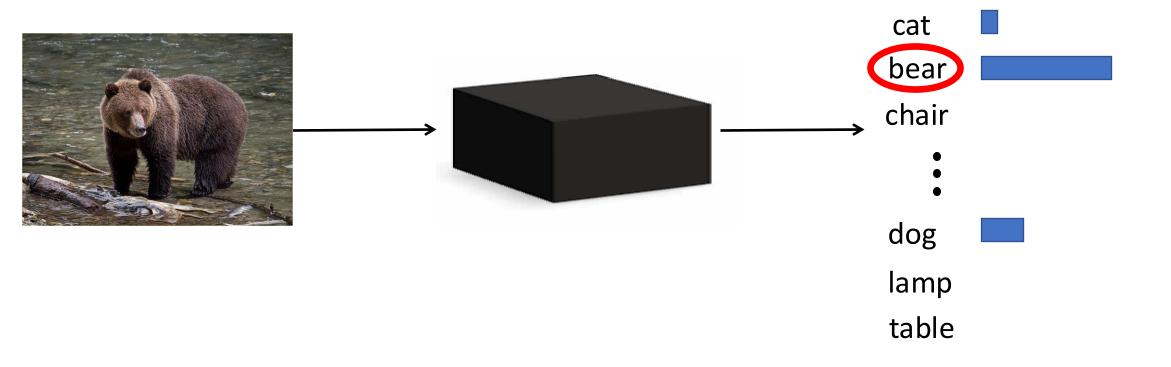
Target mapping: probability distribution from a model offers further insights into similarities and differences of categories

- Attempts to identify ground truth category
- Also, shares that 2 has similar characteristics to 7 and 1



Target mapping: probability distribution from a model offers further insights into similarities and differences of categories

- Attempts to identify ground truth category
- Also, shares that bear has similar characteristics to dog and cat



 \mathbf{O}

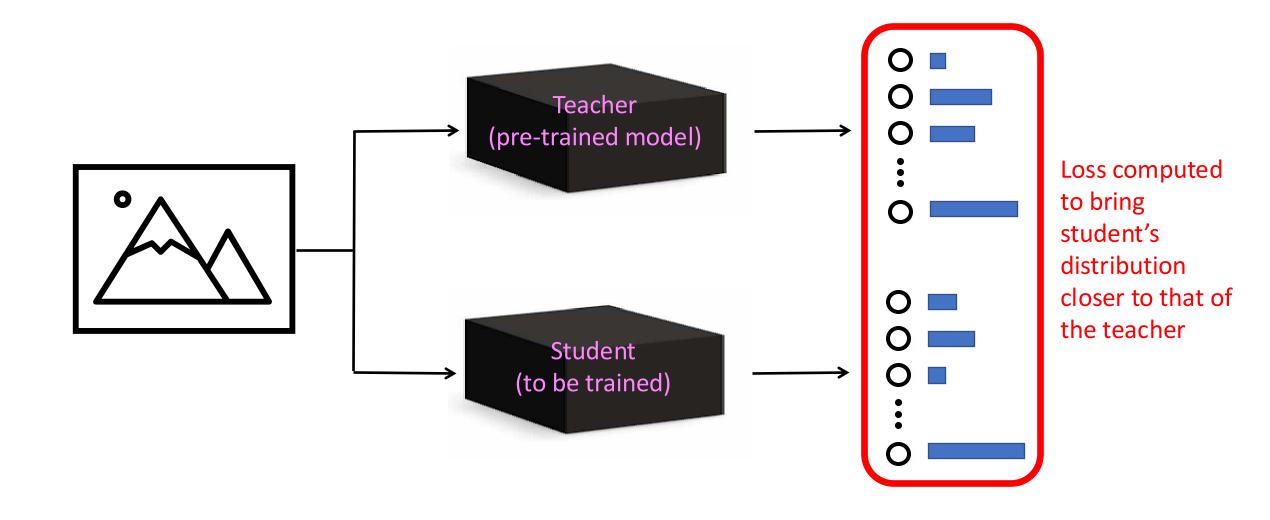
Target mapping: probability distribution from a model offers further insights into similarities and differences of categories

- Attempts to identify ground truth category
- Also, shares that bear has similar characteristics to dog and cat

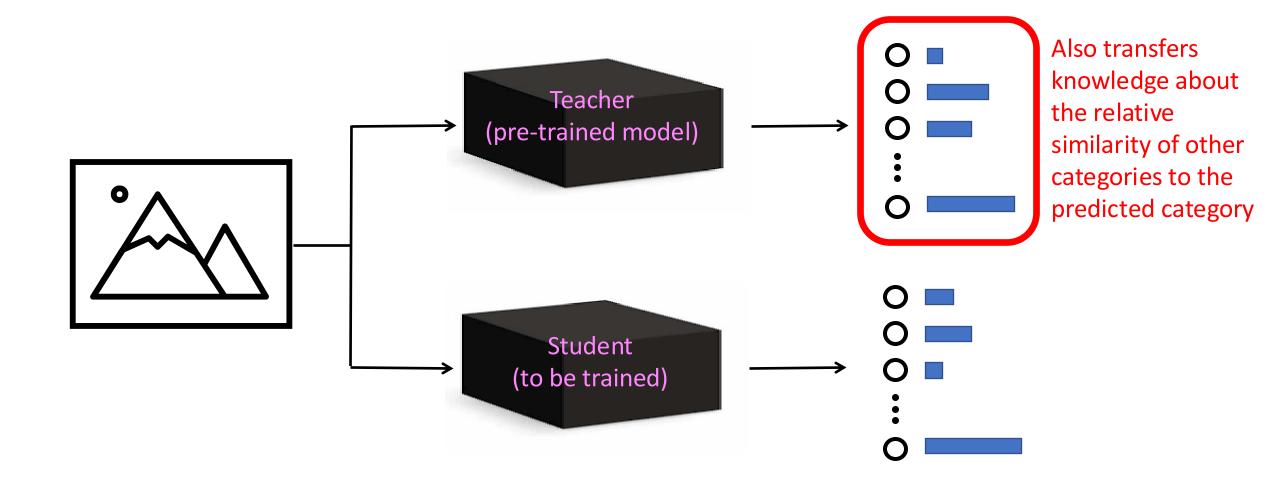
Idea: teach about ground truth and its relationships to other categories

Hinton, Vinyals, and Dean. Distilling the knowledge in a neural network. arXiv 2015.

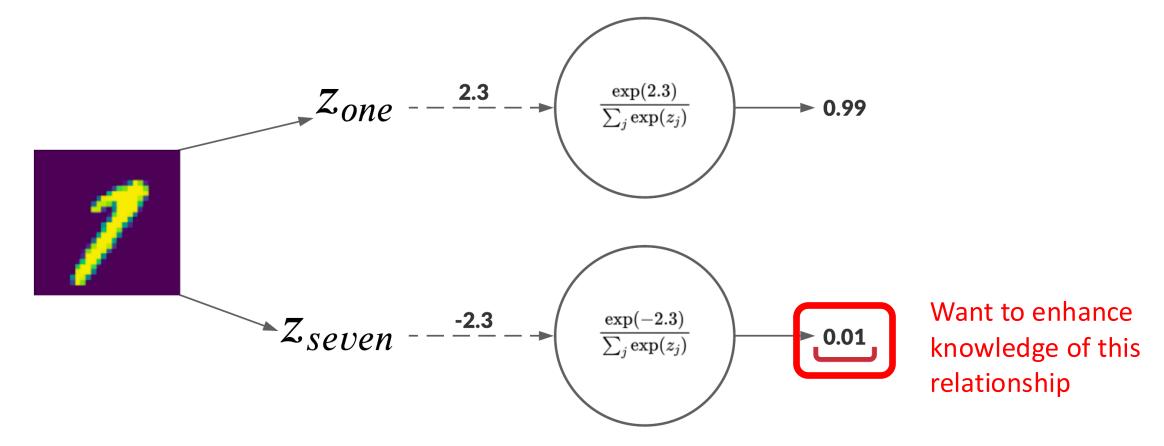
Knowledge Distillation: Teach Student the "Dark Knowledge" of Teacher



Knowledge Distillation: Teach Student the "Dark Knowledge" of Teacher



Recall Softmax: converts scores into a probability distribution that sums to 1



https://wandb.ai/authors/knowledge-distillation/reports/Distilling-Knowledge-in-Neural-Networks--VmlldzoyMjkxODk

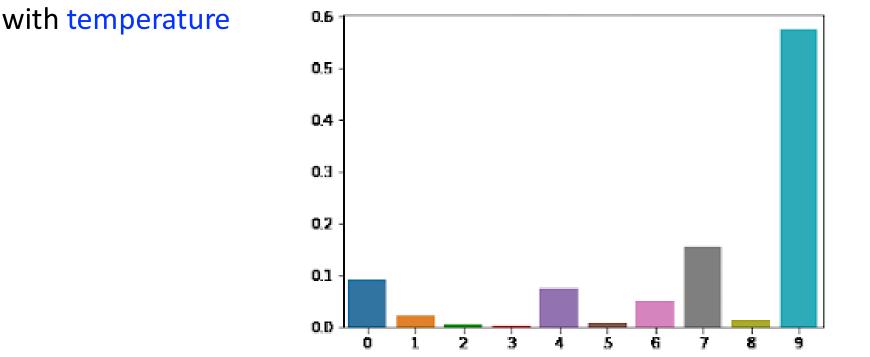
Generalized Softmax: converts scores into a probability distribution summing to 1, with temperature

 $\sigma(\mathbf{z})_i = \frac{\exp(\mathbf{z}_i/\mathbf{T})}{\sum_i \exp(\mathbf{z}_i/\mathbf{T})}$

What is the typical value of T used for softmax?

Idea: set the temperature to a value greater than 1

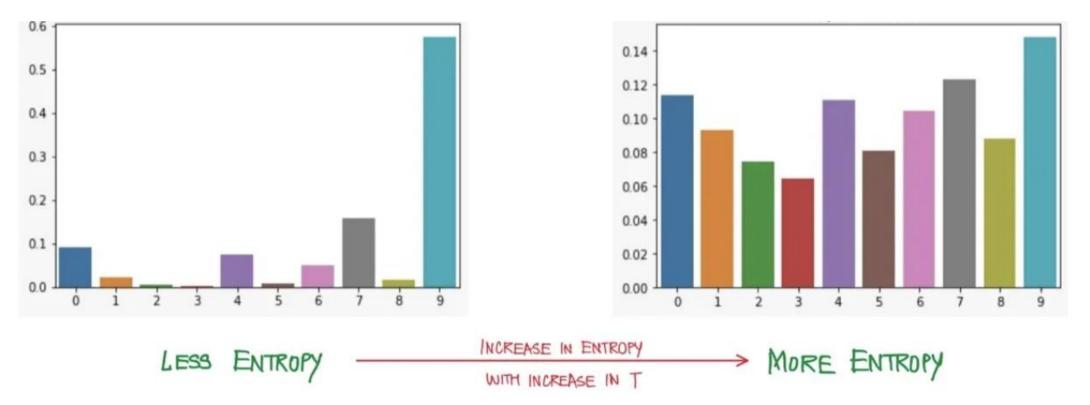
Generalized Softmax: converts scores into a probability distribution summing to 1,



Larger T values means more information is available about which categories the teacher found similar to the predicted category

https://medium.com/@harshit158/softmax-temperature-5492e4007f71

Generalized Softmax: converts scores into a probability distribution summing to 1, with temperature



https://medium.com/@harshit158/softmax-temperature-5492e4007f71

Generalized Softmax: converts scores into a probability distribution summing to 1, with temperature; e.g.,

	1		1		
0.997	Homework	0.935	Homework	0.637	Homework
0.000	Cake	0.0001	Cake	0.021	Cake
0.002	Book	0.046	Book	0.191	Book
0.001	Assignment	0.017	Assignment	0.128	Assignment
0.000	Car	0.0001	Car	0.021	Car
0.000		10000000000000000000000000000000000000			
		<u> </u>	1	T _ C	
T_1		T-2		T=5	

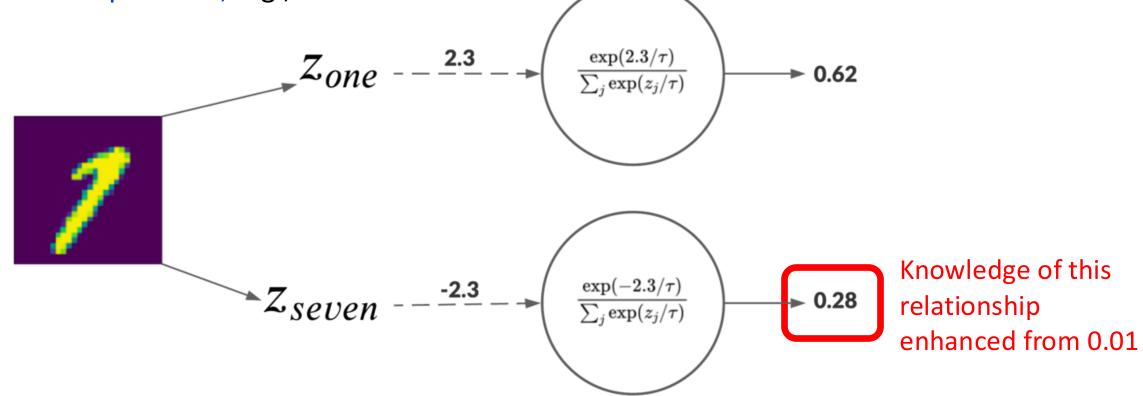
Larger T values means more information is available about which categories the teacher found similar to the predicted category

T=2

T=1

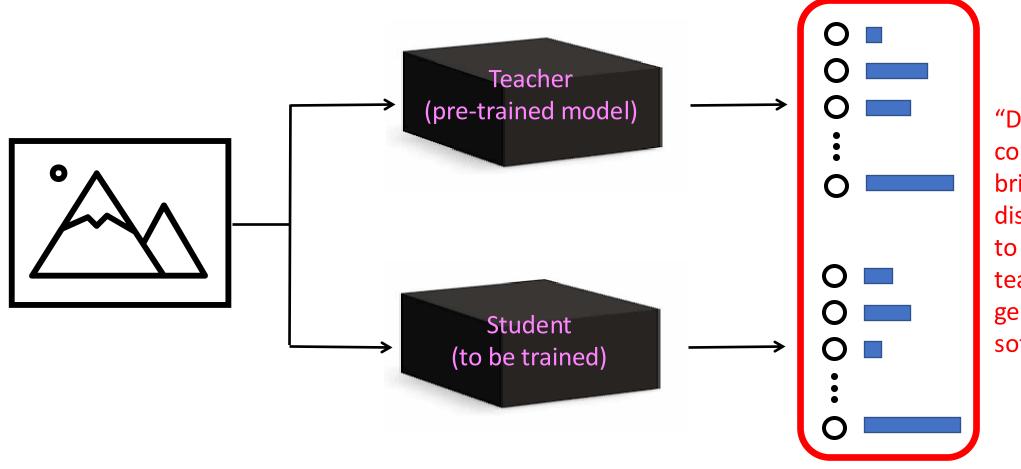
https://static.packt-cdn.com/downloads/9781838821593_ColorImages.pdf

Generalized Softmax: converts scores into a probability distribution summing to 1, with temperature; e.g., T=5



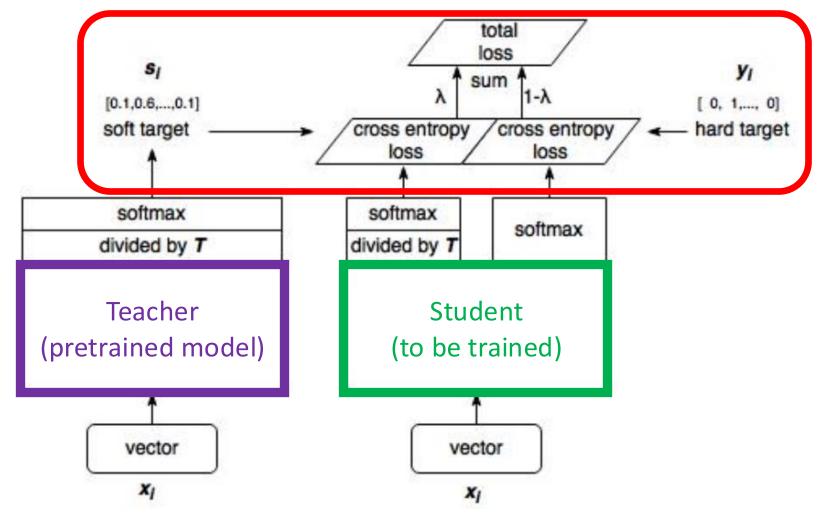
https://wandb.ai/authors/knowledge-distillation/reports/Distilling-Knowledge-in-Neural-Networks--VmlldzoyMjkxODk

Knowledge Distillation: Teach Student the "Dark Knowledge" of Teacher



"Distillation" loss computed to bring student's distribution closer to that of the teacher, using the generalized softmax equation

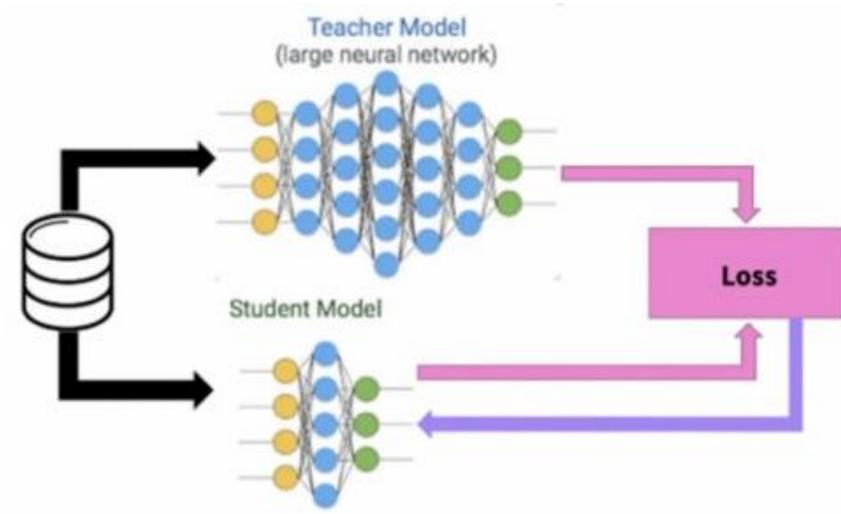
Knowledge Distillation: Teach Student the "Dark Knowledge" of Teacher



Total loss computed during training is a weighted sum of the conventional cross entropy loss and the "distillation loss"

https://blog.csdn.net/qq_22749699/article/details/79460817

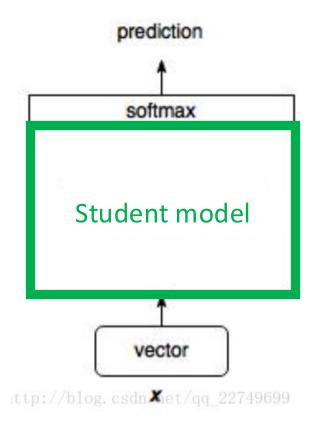
Knowledge Distillation: Teach Student the "Dark Knowledge" of Teacher



Total loss computed during training is a weighted sum of the conventional cross entropy loss and the "distillation loss"

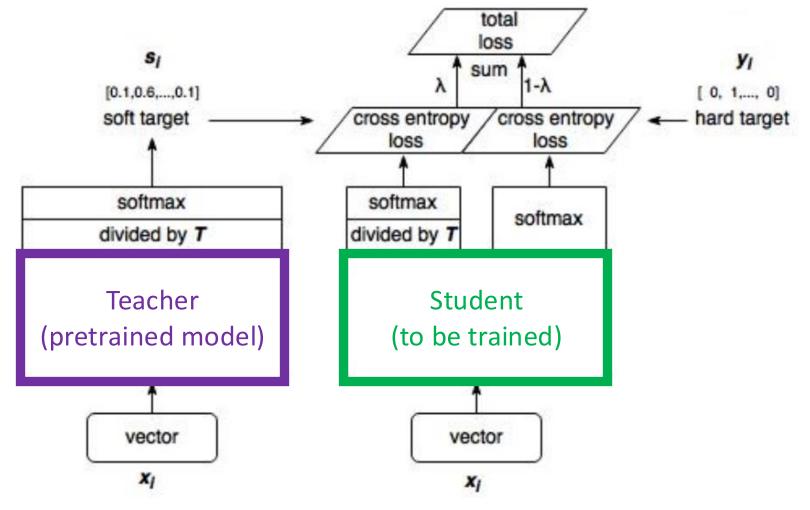
https://xailient.com/blog/4-popular-model-compression-techniques-explained/

Knowledge Distillation: At Test Time



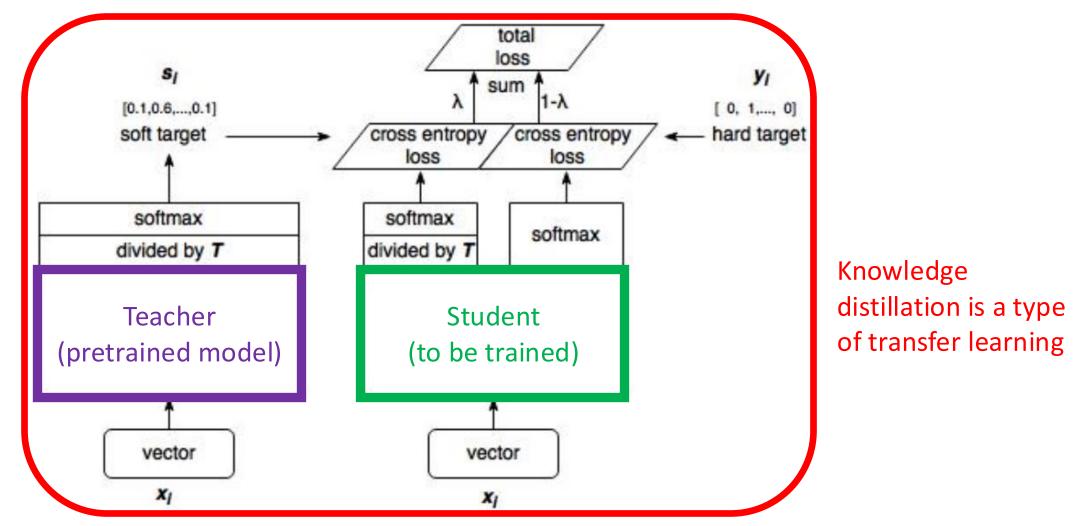
https://blog.csdn.net/qq_22749699/article/details/79460817

Arguably, Any Neural Network Student Could Learn from Any Neural Network Teacher



https://blog.csdn.net/qq_22749699/article/details/79460817

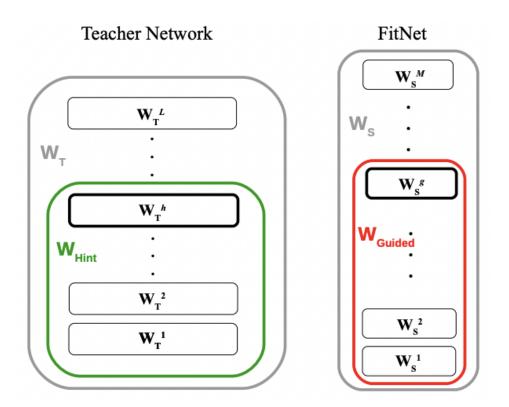
Arguably, Any Neural Network Student Could Learn from Any Neural Network Teacher



https://blog.csdn.net/qq_22749699/article/details/79460817

Knowledge Distillation Enhancement: Hints

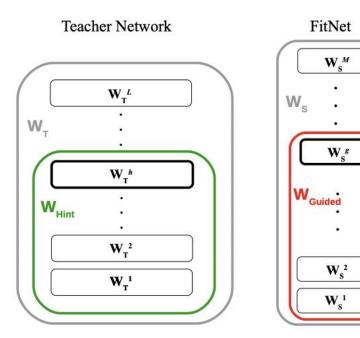
Encourage student (FitNet) to mimic the teacher's feature responses; e.g., output of guided layer should match the output of hint layer



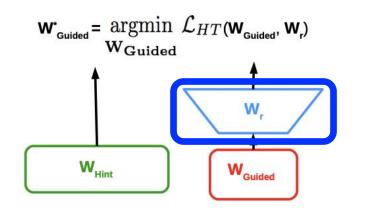
Romero et al. Fitnets: Hints for thin deep nets. ICLR 2015.

Knowledge Distillation Enhancement: Hints

Encourage student (FitNet) to mimic the teacher's feature responses; e.g., output of guided layer should match the output of hint layer



Training conducted to learn the intermediate feature



Layer added to match size of the hint's output layer

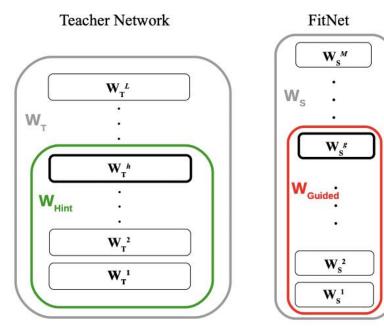
(a) Teacher and Student Networks

(b) Hints Training

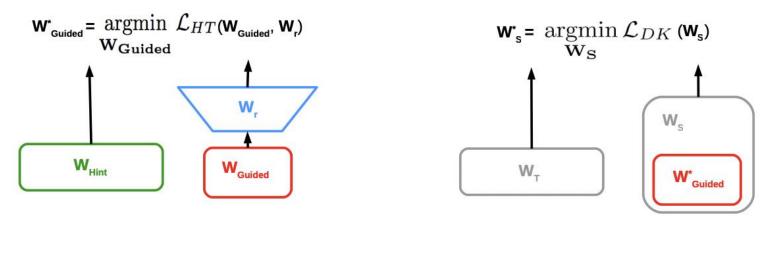
Romero et al. Fitnets: Hints for thin deep nets. ICLR 2015.

Knowledge Distillation Enhancement: Hints

Encourage student (FitNet) to mimic the teacher's feature responses; e.g., output of guided layer should match the output of hint layer



After learning the intermediate features, the whole student network is trained



(a) Teacher and Student Networks

(b) Hints Training

(c) Knowledge Distillation

Romero et al. Fitnets: Hints for thin deep nets. ICLR 2015.

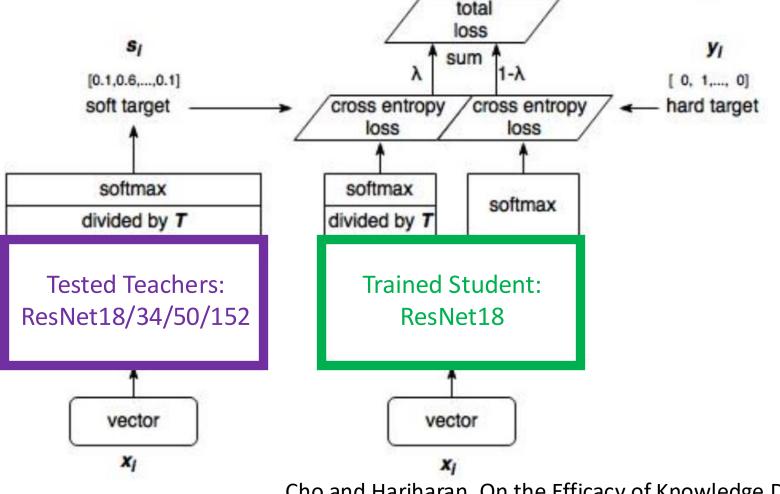
Example: Predict Category from 1000 Options

- Evaluation metric: % correct (top-1 and top-5 predictions)
- Dataset: ~1.5 million images
- Source: images scraped from search engines, such as Flickr, and labeled by crowdworkers



J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. 2009

Example: Do Bigger, More Accurate Models Make Better Teachers?



Cho and Hariharan. On the Efficacy of Knowledge Distillation. ICCV 2019 Figure source: https://blog.csdn.net/qq_22749699/article/details/79460817

Example: Do Bigger, More Accurate Models Make Better Teachers?

(% = Top-1 error rates)

Teacher	Teacher Error (%)	Student Error (%)	
ResNet18	30.24	30.57	
ResNet34	26.70	30.79	
ResNet50	23.85	30.95	

What is the student's performance trend from larger, more accurate teachers?

Example: Do Bigger, More Accurate Models Make Better Teachers?

(% = Top-1 error rates)

Teacher	Teacher Error (%)	Student Error (%)	
-	-	30.24	
ResNet18	30.24	30.57	
ResNet34	26.70	30.79	
ResNet50	23.85	30.95	

Student performance not only drops for larger teachers but the models distilled from teachers perform worse than training the student from scratch!

Example: Why Might Student Performance Drop as Teacher Size Grows?

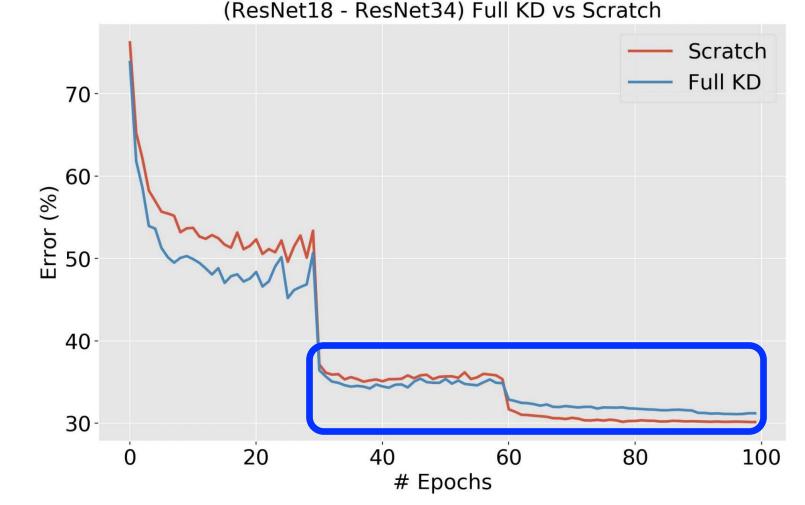
- 1. More accurate models are more confident and so need higher temperatures to learn the "dark knowledge" of category relationships
- 2. Student mimics teacher but the loss function is mismatched from the evaluation metric

3. Student fails to accurately mimic teacher

Experimental analysis suggests this is the reason

Example: Why Might Students Fail to Mimic Teachers?

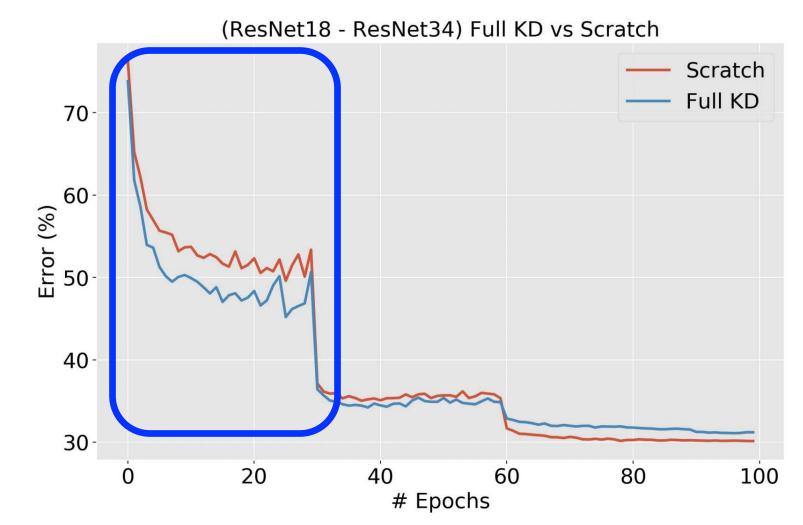
Hypothesis: student is underfitting from smaller capacity and so "minimizing one loss (KD loss) at the expense of the other (cross entropy loss)"



Example: Why Might Students Fail to Mimic Teachers?

How to overcome this issue?

- Early stopping with KD loss (ESKD) to leverage its benefit at the start of training



Example: How Does ESKD Compare To Training A Student from Scratch?

Teacher	Top-1 Error (%, Test)
ResNet18	30.57
ResNet18 (ES KD)	29.01
ResNet34	30.79
ResNet34 (ES KD)	29.16
ResNet50	30.95
ResNet50 (ES KD)	29.35

Training a model with early stopping knowledge distillation loss leads to better results than training from scratch!

Example: Are Results from ESKD Better When Using Bigger, More Accurate Models As Teachers?

Teacher	Top-1 Error (%, Test)
ResNet18	30.57
ResNet18 (ES KD)	29.01
ResNet34	30.79
ResNet34 (ES KD)	29.16
ResNet50	30.95
ResNet50 (ES KD)	29.35

No; the student may still be struggling with underfitting due to an insufficient representational capacity

Example: To Address The Capacity Problem Why Not Instead Distill to Intermediate Sizes?

Performs almost identically to a model that is distilled directly from a large to small size; does not address the core problem:

The student must be in the solution space of the teacher

What's Currently Interesting? e.g.,

What Knowledge Gets Distilled in Knowledge Distillation?

Utkarsh Ojha* Yuheng Li* Anirudh Sundara Rajan*

Yingyu Liang Yong Jae Lee

(Neurips 2024)

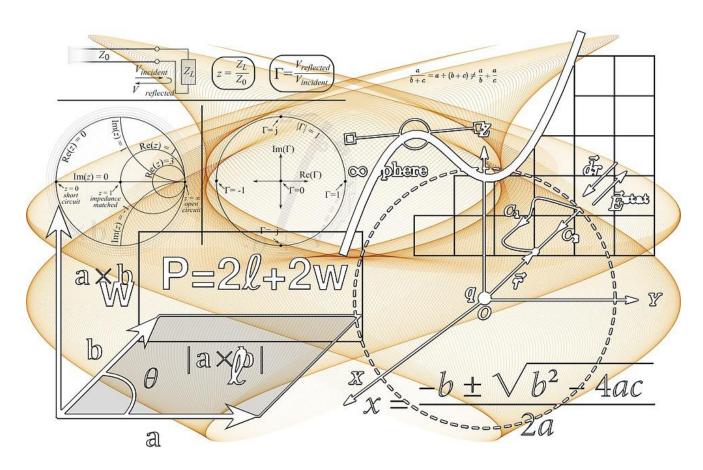
Efficient Computer Vision

- Motivation
- Model Compression
- Curriculum Learning
- Active Learning
- Faculty Course Questionnaire (FCQ)



Intuition: How to Teach a Child Math?

Random Order of Examples



Meaningful Order of Examples

Indexp National States Insuper: Making out Value etg. Americal States Insuper: Making out Value etg. I Aper 1440 Respect Making Value etg. I Aper 1440 Respect Making Value etg. I Aper 1440 Parameterge etg. I Aper 1440 Parameterge etg. I Aper 1440 Parameterge etg. J Aper 1440 Parameterge etg. Parameterge etg. Parameterge etg. Parameterge etg. Parameterge etg. Parameterge etg. Parameterge etg. Parameterge etg. Parameterge etg. Parameterge etg.<	Back Astronom and Public States Wight Discretion Wight Discretion Name of Astronom Maining To Appropriate Weight Advag Statiscretion Advag Statiscretion Main Statiscretion Advag Statiscretion Advag Statiscretion Mark Statiscretion	Lose her Disk Dis	- N -	Some for the Soliday	-
Wig The Voluments in Hallmander' Hapf: Acceleration using PMAMA Particle Research Particle Section Balancing Data National Section Balancing Data National Section Balancing Data Section Balancing Data Section Balancing Data Section Balancing	Wig Un forestere a laboration of the laboratio of the laboration of the laboration of the laboration of			Endormal Strips	
Parante Recise 1 Marche Determini Bissioning De Approprise Westglefe 1 Ward, Handmann Ward, Handmann 1 Bargare Bissioning De Approprise Westglefe 1 Bargare Bissioning De Approprise 1 Bargare Bissionin Managare 1 Bargare </td <td>Point Rein 1 March Determ Binage Contractions March Stations ddiag instructions 1 March Stations ddiag instructions 1 March Stations 17 And Held 1 March Stations 18 And There 1 March Stations 19 And Held 1</td> <td></td> <td></td> <td></td> <td></td>	Point Rein 1 March Determ Binage Contractions March Stations ddiag instructions 1 March Stations ddiag instructions 1 March Stations 17 And Held 1 March Stations 18 And There 1 March Stations 19 And Held 1				
Bissing Ro Appopulation Weights 1 Wach Rocking redding Biothecities Biggert Biggert shift Oblack Dags 1 Biggert 1 Mar Field Biggert Biggert 1 Mar Field Papers Biggert 1 Mar Field Papers Biggert 1 Mar Field Papers Papers 1 Amer Field Papers Pa	Binaming the Appropriate Week/4 1 When Realism dding Realismental dated Velocity Degree Region Instance Region Instance If the Fridal Real Fordal P Real Fordal Region Instance If the Fridal Real Fordal P Real Fordal Real Fordal If the Fridal P Real Fordal Real Fordal Real Fordal P Real Fordal Real Fordal Real Fordal P Real Fordal Real Fordal Real Fordal P Real Fordal Real Fordal Apple Tam Band P Real Fordal Real Fordal Mand Fordal P Real Fordal Real Fordal Tam Ford Fordal P Real Fordal Real Fordal Tam Ford Fordal (Model) P Real Fordal				
 Marge Markers States Marge Markers Markers <l< td=""><td>Address instructions Magnet: Address instructions Magnet: Mainer Mitchell Statum Magnet: Marchild Magnet: Marendi</td><td></td><td></td><td></td><td></td></l<>	Address instructions Magnet: Address instructions Magnet: Mainer Mitchell Statum Magnet: Marchild Magnet: Marendi				
nddeig hol sectors staniy toldadir y Sagan // Barger Billing and Salawing (Salawing Salawing // Sal	Address instructions image instructions Analy Failed Volume image instructions Analy Faile image instructions Instructions image instructions	 Balance in Appropriate Possible 			
Anny Solitable (Segue) 1 Image: Mologing and Decimany I Need Fields 2 Readed Numbers (Section) Early Male 2 Readed Numbers (Section) Reade Novie 3 Participany Mactional Novie 3 Participany Mactional Novie 3 Participany Mactional Novie 3 Participany Mactional Novie 3 Participany Stated Novie 3 Participany Novie 3	Amir Maladir Diagon 4 Image: Maladir Diagon 1 Next Hala 5 Radical Readers 1 Next Hala 6 Radical Readers 1 Next Hala 7 Readers 1 Next Hala 8 Readers <td>사람은 옷을 잘 가지 않는 것이 같아요.</td> <td></td> <td></td> <td></td>	사람은 옷을 잘 가지 않는 것이 같아요.			
France Frank Resident Resolver, Proceedings Frank Frank Barrier Resolver, Procession Frank Frank Barrier Resolver, Procession Frank Frank Procession Jone Frank Procession Theore Frank Procession Theore Frank Procession Procession Frank Procession <td>Free Feld Rained Readers Not Feld Rained Readers Not Feld Rained Readers Feld Rained Readers Note Florid Rained Readers The State Prome Note Florid Rained Readers Joint Florid Rained Readers State Florid Readers Readers Rained Readers</td> <td></td> <td></td> <td></td> <td></td>	Free Feld Rained Readers Not Feld Rained Readers Not Feld Rained Readers Feld Rained Readers Note Florid Rained Readers The State Prome Note Florid Rained Readers Joint Florid Rained Readers State Florid Readers Readers Rained Readers				
Inter Frank Barner Rushest, Frankes Karlf Book 1 Road Rock 1 Road Road 1 <tr< td=""><td>Invest Head Participant Kord Head Participant Fabbel Hook Participant Researd Hook Participant Text Data Block Participant Marchael Participant Participant Participant Participant Participant Participant Participant Participant Participant Participant Participant Participant Participant Participa</td><td>damy follow Mapping</td><td></td><td>underst specifications assessed.</td><td></td></tr<>	Invest Head Participant Kord Head Participant Fabbel Hook Participant Researd Hook Participant Text Data Block Participant Marchael Participant Participant Participant Participant Participant Participant Participant Participant Participant Participant Participant Participant Participant Participa	damy follow Mapping		underst specifications assessed.	
Burn Processor Particle Processor Branch Rock 1 Branch Rock 1 Theo Torr Brang Particle March Rock 1 March Rock 1 March Rock 1 March Rock 1 March Rock 10 March Rock 10 March Rock 10 March Rock 10 Submit Park 10 Park Park 10 Rest Park 1	Nort Next. ** Extensit Number: Nextensit Naski Nork 1 Pressite Naski Nork 1 Ratio Ten: Ten Next 1 Ratio Marchank 1 Ratio Marchank 17 Pressite Scher Addit 17 Pressite Scher Addit 18 Pressite Three Too Rend. 10 Pressite	A reason builds			
Factor Bank 1 Name Branch Bank 1 Particion Jorder John 1 Particion Jorder John 10 Particion Jorder John 10 Particion Jorder John 10 Particion Jorder John 11 Particion Jorder John 12 Particion Jorder John 13 Particion Jorder John 14 Particion Jorder John 15 Particion Particion John 16 Particion Particion John 17 Particion Particion John 18 Particion Restrick Last, Bords 19 Particion Restrick Last, Bords 10 Particion Restrick	Field Quel 1 Name Read Rock 1 Person Read Rock 1 Person Jacobia 1 Person Marchall 1 Person Ream Public 1 Sec and Realism Schall Road 1 Person Theor Tax Schall 1 Person Read Schall 1 Person	traff Book	1.46		
Result Tork 1 Press Then Dim Hang 9 Partinia Aver Validi Prescription Prescription Aver Validi Prescription Prescription Aver Validi Prescription Prescription S. Pari Validi Prescription Prescription S. Pari Validi Prescription Prescription S. Pari Validi Prescription Prescription Three Tool Nock Nationame Prescription Prescription Three Tool Nock Nationame Prescription Prescription Prescription Prescription Prescription Advertifier Bresc Prescription Prescription Prescription Bresc Prescription Prescription Prescription Bresc Prescription	Receit Rock 3 Press Terr Terr Hand 4 Partic Terr Terr Hand 4 Partic Marchink 10 Partic Marchink 10 Partic Marchink 10 Rest And Partic Marchink 10 Rest And Partic State Volds 10 Rest And Partic State Volds 10 Rest And Partic State Volds 10 Rest And Partic Terr Too Event 10 Rest Notion Terr Too Event 10 Pereor Terr Too 10 Pereor <td< td=""><td></td><td></td><td></td><td></td></td<>				
Theor Start Manage 8 Programme JParey Veliditi Jac March Starting JParey Veliditi Jac March Starting March Starting Jac March March Starting Jacker Veliditi Jac March March Starting Start March Starting Start March Starting March March March Starting Start March Marc	Theor Short March State Programmer J-Auer Antili Journal Fordition March State JP Parter March JP State March JP State March JP State March JP Theory Decord JP Theory Decord JP March March JP Theory Decord JP March March JP Marer March JP M				
j. Ange Vedati Marchinals	J. Ange Ander Marchitekin				
Marchool J2 Text Number Foreset Jones (Inco. 11 Algebraic Process, soil Numbers S. And Addr. 12 Algebraic Process, soil Numbers S. And Addr. 12 Algebraic Process, soil Numbers S. And Addr. 12 Algebraic Process, soil Numbers S. And Addr. 13 Progenetic Tarter (Do Dentic, Soil) 16 Registration Theory Too Dentic, Soil 16 Registration Theory Tool Soil 16 Registration Teacher Tool Soil 17 Marchool Soil Teacher Tool Soil 17 Marchool Soil Teacher Tool Soil 18 Sopenser Marchool Soil 18 Sopenser Anteretheral Ages 18 Sopenser Angebrachteral Ages 18 Sopenser Angebrachteral Ages 18 Sopenser	Maximuk PP Text Number Network Porter Hors PP Address Personal and Numbers School Hors PP Address Personal and Numbers School Hors PP Non-Nether Network School Hors PP Personal Theor Por Person Personal Personal Theor Port Person Personal Personal Theor Port Port Person Personal Personal Theor Personal Lize				
Parter Serret 11 Applicato Automatic and Numitians New York 12 Served Automatic and Numitians Safety Tark Event 12 Reparation Tarter Tark Event 12 Reparation Aver Tark 14 Reparation Aver Tark 15 Reparation Reserved 15 Reparation Reserved Rest Event 15 Reparation Reserved Statt Event 16 Reserved Reserved Statt Event 16 Reserved Reserved Statt Event 16 Reserved New York Event 16 Reserved Reserved Tark Event Fail 16 Reserved Reserved Tark Event Fail 16 Reserved Reserved Tark Event Faile 16 Reserved Reserved	Particip Rent 11 Applicacy Parameters Interim Publ. 12 Status Planations Schell Rent 12 Status Planations Table Table 12 Reprinter Table Table 12 Reprinter Table Table Rent 12 Reprinter Table Table Rent 12 Reprinter Table Table Rent 13 Reprinter Table Table Rent 14 Reprinter Table Table Rent 15 Reprinter Table Table Rent 16 Reprinter Exercite Rent 16 Reprinter Exercite Lint, Porth 17 Matinter Render Station 18 Response Render Station 18 Response Near Stating Table 17 Matinter Near Stating Table 17 Matinter Near Stating Table 18 Response Near Stating Table 18 Response Near Stating Table 18 Response Near Stating Table		0.00	the second se	
Immu Pold 12 Append or community community 1. Peri Male Expension Expension Schedi Real 10 Expension Three Tax Book Valuese 15 Regeneration Three Tax Book Valuese 16 Regeneration Present Tax Book 16 Regeneration Present Tax Book 16 Regeneration I Aver Polds 16 Reference I Aver Polds 16 Reference I Aver Polds 16 Reference I aver pold 16 Reference I aver pold 17 Reference I aver pold 18 R	Insue Feld II Append of Section and Texames 3-Rei Voldi Expendition Expendition Scheil Reich 10 Expendition Theory Fox Book Volutioner 55 Regenition Rever Fox Book Volutioner 56 Regenition Rever Fox Book Volutioner 56 Redenition Rever Fox Book Volutioner 56 Redenition Rever Fox Book Volutioner 56 Redenition Rever Fox Book Volutioner 56 Redenition and Potensiner Rever FoxBook 37 Potence Rever and Expension Rever FoxBook 37 <t< td=""><td></td><td></td><td></td><td>100</td></t<>				100
Search Section Search Section School Section 10 Tarter Too Event 10 Theorie Too Event 10 Alexer Folds 10 Too Event 10 Alexer Folds 10 Too Event 10 Too Event Tool 10	S. April Mail: Section Research Constraints Section Research Con			Alphone Parents and American	
Scholl Find. 11 Pageword Theor Tax Event. 24 Pageword 15 Theor Tax Event. 25 Registrice. 15 Present of the book Valuence 35 Registrice. 16 Present of the book Valuence 35 Registrice. 16 Present of the book Valuence 35 Registrice. 16 Present of the book Valuence 36 Pageword 16 Reserved Katt Door 36 Pageword 16 Reserved Valuence 37 Pageword 16 Pageword Acceleration 32 Pageword 16 Pageword 16 Acceleration 32 Pageword Registrice 16 16 16 16 16 16 16 16 16 16 16 16 16 16	Math Real. 11 Paper Ref. Theor Text Real. 14 Represent Theor Text Real. 15 Represent Procent Post In Model 16 Realization of Patentine Procent Post In Model 16 Math Realization Read Post Inf. Post 17 Math Realization Read Post Inf. 18 Sequence Nearbox Vold 18 Sequence Nearbox Vold 18 Sequence Nearbox Vold 24 Math Real Nearbox Vold 25 Marcant Name Vold 26 Nearbox Name Vold 27 Nearbox Name Vold 28 Name Nearbox Name Vold 29 Name Nearbox Name Vold 24 Name Nearbox Name Vold 24 Name Nearbox	House Paid concerns a second	- 340	Key and Relation	
Bare Car Lor Card 34 Agenetics Agenetics Theory for book Valuence 55 Regeneration Regeneration Present for to Motion 56 Reference Regeneration Present for to Motion 56 Reference Regeneration Present for to Motion 56 Reference Reference Present for Motion 57 Management Reference Reservice Local, Fords 57 Management Reference Reservice Local, Fords 57 Management Reference Reservice Local, Fords 57 Management Reference Reservice Local, Contract 57 Management Reference Reservice Contract 57 Management Reference Associate Roma 58 Reference Reference Associate Roma 58 Reference Reference Reference Contract Statts 58 Reference Reference Reference Contract Statts 58 Reference Reference Reference Contral Statts	Base-Do Best H Repairing Theor Do Book Valuemen 55 Respective Present Por Nok Valuemen 55 Residue and Paramen Present Por Nok Valuemen 55 Residue and Paramen Rev Fulde 56 Residue and Paramen Rev Fulde 56 Matagaine Leverly Lest Port 57 Matagaine Leverly Lest Port 56 Matagaine Navelage Static 58 Matagaine Navelage Static 58 Matagaine Navelage Static 56 Matagaine Navelage Static 50 Matagaine Static Static 50 Matagaine Static Static 56 Matagaine Static Static 56 Adaptin Static Static Static 56 Adaptin Static S	3. Part Adds		Experience	
The out of the Stationes 55 Incernance Present First in Stationes 56 Relationes and Processors Leveler Level 56 Relationes and Processors Leveler Level 57 Mathematical Processors Leveler Level 57 Mathematical Processors Reserved 57 Mathematical Processors Reserved 58 Preserve Reserved Level 58 Preserve Reserved 58 Preserve Reserved Rese 58 Preserve Reserved Rese 59 Preserve Autorities Rese 59 Preserve Reserved Reserved Reserved Reserved Reserve 70 Reserv	Theor Tab Book Valuemen 55 Incernation Processing Factor in Model 16 Balances and Potentiamen Linewise Loss (Model) 16 Matalance Linewise Loss (Potentiamen) 17 Matalance Linewise Loss (Potentiamen) 17 Matalance Reser Mall 17 Matalance Reser Mall 18 September Reser Mall 18 September Reser Mall 20 Matalance Reser Mall 21 Matalance Reser Mall 22 Matalance Autoretine Rese 21 Matalance Autoretine Rese 22 Matalance Autoretine Rese 23 Matalance Autoretine Rese 24 Priver Tage Valente of Partition 26 Augein Toolog and Line Graph and 26 Augein Toolog and Cloub Services 26 Privageine Toolog and Cloub Services 27 Matalance Toolog and Cloub Services 26 Privageine<	Territoria Breathanne and an announcement and		Parperties	
Process of Food in Station (b) Relations and Processors 2. Rev Folds Parter Parter Leveler Lends Death (f) Maximum Leveler Lends Death (f) Maximum Leveler Lends Death (f) Maximum Trace Tell Noor (f) Maximum Star State State (f) Supersor Trace Tell Noor (f) Maximum Trace Tell Noor (f) Supersor Trace Tell Noor (f) Supersor Trace Tell Noor (f) Noorkeet Tage Tell Nool (f) Noorkeet Tage Tell Nool (f) Noorkeet Tage Noorkeet of Parts (f) Noorkeet Tage Noorkeet of Parts (f) Naple Tage Noorkeet Tell Noor (f) Naple <td>Pyromet Field is 100000 (b) Balaisses and Potensini 2 New Fields Process Process Lensing Loss (Loss (Post) Process Maintain New Field Incom Process Maintain Year Edit Process Maintain Year Edit Process Maintain Year Edit Process Maintain Year Edit Process Process Stage Scander of Parts Process App Internet Process and Factor of Parts Process App Inte</td> <td>- These Tele Rent.</td> <td>2000 (MC</td> <td>Ingenitries, contract of the second second</td> <td></td>	Pyromet Field is 100000 (b) Balaisses and Potensini 2 New Fields Process Process Lensing Loss (Loss (Post) Process Maintain New Field Incom Process Maintain Year Edit Process Maintain Year Edit Process Maintain Year Edit Process Maintain Year Edit Process Process Stage Scander of Parts Process App Internet Process and Factor of Parts Process App Inte	- These Tele Rent.	2000 (MC	Ingenitries, contract of the second second	
Aper Folds Lever Loss, Book	Parer Fable Parer Fable Learning Last, Deah Parer Parer Pable Learning Last, Deah Parer	"These Tab Book Maintines",		Republic constants and an experimental	
Avery Fulds Largency Louis, Fords Ford	Arren Funds Learning Least Fronton Transford Funds Transford Tr	Prepared Fact in McKilli	11111	Relations and Potentians	
Larence Lost, Bosh FP Multisolution Point DA Noor B Person and Deponation Person and Deponation Reschapt Vall B Separation Person and Deponation Person and Deponation Reschapt Vall B Separation Person Person Person Tag Vall B Separation Person Person Person Tag Vall Data Data Person Person Person Tag Vall Data Data Person Person Person Person Tag Vall Data Data Person	Levery Lett, Bort Programmi, Prog			Futors	
Pain Fill Now Hit Memory and Exponent Reaching Vall 28 Sequence Sequence Reaching Vall 28 Sequence Sequence Top Tak Reak 20 Sequence Sequence Associate Reat 29 Frist Toos and Loss Sequence Top Tak Reak 20 Sequence Sequence Sequence 29 Frist Toos and Loss Sequence Top Tak Reak 20 Asglits Asglits Problem time Tolks 26 Asglits Sequence Problem time Tolks 28 Polyages Sequence Problem time Tolks 28 Polyages Sequence Charditor 28 Topolyage Sequence Charditor 23 Region Sequence Sequence Charoney Noles 28 Regio Tol	Prior Ed Noor Hit Prevents of Population Bandop Fold Bit Prevents of Population Bandop Fold Bit Supprevent Working Catle Dit Supprevent Top Tax Reads Dit Supprevent Supprevent Reads Dit Prevents Supprevent Reads Dit Reads Preduction Reads Dit Reads Prevents Reads Dit <td></td> <td></td> <td>Malighter</td> <td></td>			Malighter	
Bandapi Faill IP Device state paperse Nacolagi Faill 22 Separation Nacolagi Faill 22 Maximum Tage Tain Road 22 Maximum Tage Tain Road 23 Maximum Association Road 24 Point Association Road 25 Point Tage Samotor of Paris 26 Point Tage Samotor of Paris 26 Paris Trading into Paris 26 Applies Today and Paris 26 Applies Today and Paris 26 Applies Continues Today and Paris 26 Point Today and Paris 28 Point Today and Paris 28 Point Today and Paris 28 Point Chow Hays 29 Point Chow Hays 20 Today and Wandwiney Road 20 Today and Value of Samoton 20 Today and Value of Samoton 20 Today and	Basebay Solid Dire Direct Space Nauskay Calu Ji Supervise Nauskay Calu Ji Maxwait Tay Tao Boak Ji Maxwait Tay Tao Boak Ji Maxwait Say Scontor of Parts Ji Print Say Scottor of Parts Ji Augint Training Name Tails Ji Augint Tail Tails Ji Augint Training Calus Inter Tails Ji Augint Training tails Ji Maxwait Training Calus Inter Tails Ji May Scottor of Parts Training Calus Inter Tails Ji Magint Training Calus Inter Tails Ji May Scottor of Parts Training Calus Inter Tails Ji May Scottor of Parts Training Calus Inter Tails Ji May Scottor of Parts Training Calus Inter Tails Ji May Scottor of Tails Char Graph Calus Inter Tails Ji May Scottor of Tails Char Graph Calus Inter Tails Ji May Scottor of Tails			Weinstein auf Polymouth 11111	
Number Office Distance Distance Transform Dami Distance Distance Distance Tage Table Bands Distance Distance Distance Tage Table Bands Distance Distance Distance Tage Table Bands Distance Distance Distance Dar (Vp Bank Distance Distance Distance Dar (Vp Bank Distance Distance Distance Dar (Vp Bank Distance Distance Distance Distance Class Note Section Distance Distance Distance Distance Class Note Section Distance Distance Distance Distance Class Note Section Distance Distance Distance Distance of	Neurity (Tate Dis Support Trans. Event from			Prevent and Expension	
Page State State 20 Decempt Tage Xin Houls 20 Question Associate State 23 Question Associate State 23 Printer Jage Konstor of Parts 900 900 Dari Vo Brish 26 Applie Totaling in Tribins 26 Applie Printer Tribins 26 Applie Totaling in Tribins 26 Applie Decisities Tribins 26 Printer Chrony Tribins 26 Printer Chrony Tribins 26 Printer Chrony Tribins 28 Printer Chrony Tribins 28 Printer Chrony Tribins 28 Tribung tribins Chrony Tribins 28 Tribung tribins Printer 23 Tribung tribins Printer 23 Tribung tribins Printer 28 Tribung tribins Printer 20 Tribung tribins Printer 20 Tribung tribins <td>Process 22 Second Tay Dat Boat 22 Second Accepted Base 23 Prime Accepted Base 24 Prime Say Scientific of Parts 26 Accepted Base Day Scientific of Parts 26 Accepted Base Day Scientific of Parts 26 Accepted Base Problem two Parts 27 Prime Problem two Parts 28 Prime Problem two Parts 29 Primare Problem two</td> <td></td> <td></td> <td>- September - Constant - Constant</td> <td></td>	Process 22 Second Tay Dat Boat 22 Second Accepted Base 23 Prime Accepted Base 24 Prime Say Scientific of Parts 26 Accepted Base Day Scientific of Parts 26 Accepted Base Day Scientific of Parts 26 Accepted Base Problem two Parts 27 Prime Problem two Parts 28 Prime Problem two Parts 29 Primare Problem two			- September - Constant - Constant	
Tay Tak Rush 32 Octomenty According Rush 34 Friend According Rush 34 Friend Day Variable of Parts 1000 and Use Segments Root Day Variable of Parts 34 Root Root Table of Parts 35 Aughts Root Root Table of Parts 35 Aughts Sectionities Root	Tage Data Hauk 22 Occurring Accuration Breet 24 Prints Accuration Breet 1000 and Line Segments Breet Tage Variation of Parist 26 Acquire Tage Variation of Parist 26 Acquire Toology Internet Chair or Graph 27 Magins Toology Costs later Strain 28 Pringers Charange Costs later Strain 28 Pringers Charange Strain 29 Pringers Toology Strain 20 Pringers <td< td=""><td></td><td></td><td>Maximum</td><td></td></td<>			Maximum	
Accession them	Accession Borns			and the second second second second second	
Any Kender of April Free and Los Segment Tog Vp Brok 25 Aught Toking min Plan. 25 Aught Segment Toking min Plan. 26 Aught Segment Toking min Plan. 26 Aught Segment Toking min Plan. 28 March Toking min Plan. 28 Toking min On only Christian 28 Toking min Oncorp Wag from 28 Toking in Wagdowing Rold 28 Toking from Proport Sing Ann Xope Values Toportement Proport Sing Ann Xope Values Toportement	Jany Scienter of Paris Lines and Line Segments Tay for third 25 Tables use Plats 26 Tables use Tables 27 Tables use Tables 28 Chronope Dig Tables 26 Chronope Dig Tables 26 Tables Use Tables 28 Chronope Dig Tables 28 Tables Use Tables 28 Chronope Dig Tables 28 Tables Use Tables 29 Tables Use Tables 20				
Jacy Kendor of Parts Back Back<	Any Kender of Paris Par Vie Paris Par Vie Paris Par Vie Paris Pari				
Darj Ve Book 25 Aagle Tridding min Fillion 26 Aagle Reconstitution Passard Tasks Units or English 28 Reconst Passard Tasks Units and English 28 Polygeme Convertings 28 Tridingsion Convertings 28 Tridingsion Workshort Red 28 Tridingsion Producting Red 20 Bigler Diagelon Producting Red 20 Bigler Diagelon Producting Red 20 Bigler Diagelon	Day (r) Einit 20 Aug/tit Tricking inter Parker 26 Aug/tit Problem inter Parker 26 Aug/tit Problem inter Parker 26 Aug/tit Problem inter Parker 27 Maxial Problem inter Parker 28 Tradington Changer King Brain 28 Bight Diarright Vacidation (Rest) 28 Bight Diarright Vacidation (Rest) 28 Bight Diarright Property King Math 29 Diarright	Jusy Monitor of Parts			
Telding real Fullio De product fullion De pro	Texture (Volumentary) Description Descrip Description <thdescrip< th=""></thdescrip<>				
Instance - Continue Training 21 Paring Tooling - Contribute Training 28 Tooling-or Christian Training 28 Tooling-or Christian Training 28 Tooling-or Christian Training 28 Tooling-or Christian Training 28 Tooling-or Voodulery Book 28 Tooling-or Proporti Staing Awari State Training Fragmentery	Instance Code of York 01 Point Totaling A Code and Taxalin 20 Point Christian Code and Taxalin 20 Point Vacidation Taxalin 20 Point Trapont code And 20 Point Propont code And 20 Point				
Description 24 Transport Christing 25 Transport Christing 28 Transport Christing 28 Transport Versioning 28 Transport Producting Name 28 Transport Producting Name 29 Transport	Today 4 Cells and Malla 24 Chrysling 4 Ch				
Overviewit 3% Vragos Crosopy Nag Bog 80 Night Darger Vacadady Rod 33 Night Darger Proposition Finds 30 Night Darger Proposition Finds 33 Night Darger Proposition Finds Operations from the second secon	Christeller 24 Volages Councy Hay Bols 24 Volages Southern Mark 24 Volages Nordelary Role 23 Reple Volages Southern Mark 24 Volages Southern Mark 24 Volages				
Charger Hay Base III Bight Toangint Yandhalary Bost III Dight Toangint Proport Joing Alath Operating Charger Manager III Proport Joing Alath Operating State	Creater Nat None N Bight Danger Technicy Red 27 Right Danger Proportions Allen Proportions Allen				
Vieldwiry Bolt. 31 Nije Vierph Tigermere . Preject inlig faith	Productory Book 21 Repair Visionale Commenters				
Project integration Open date in	Projects indep Adds Qualifying A	 Methodary Book 		Window Colorado - Coloradorea	
		Persident Indian Lange		One design of the second se	
			111		
		Farmers Farmers	100	And the second s	

Big Book of Math; Dinah Zike

Intuition: How to Teach a Child To Read



Random Order of Examples



Meaningful Order of Examples



Idea: Teach Machines As We Teach Humans

Curriculum

Train with simpler examples first and progressively harder examples over time

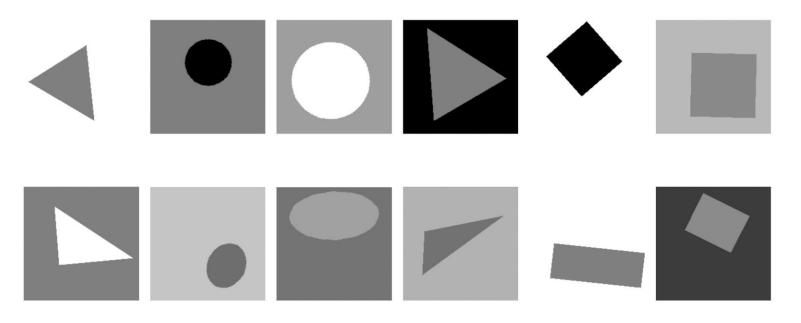
Jeffrey L. Elman. Learning and development in neural networks: The importance of starting small. Cognition, 1993.

Key Evaluation Metrics

- Training convergence speed
- Generalization performance on test data

Pioneering Task: Shape Prediction

Classify each shape as rectangle, ellipse, or triangle



Solution: 3-layer neural network

1. Easy (Basic): less shape variability (squares, circles, and equilateral triangles); 10,000 examples

2. Hard (Geom): more shape variability (rectangles, ellipses, and triangles); 10,000 examples

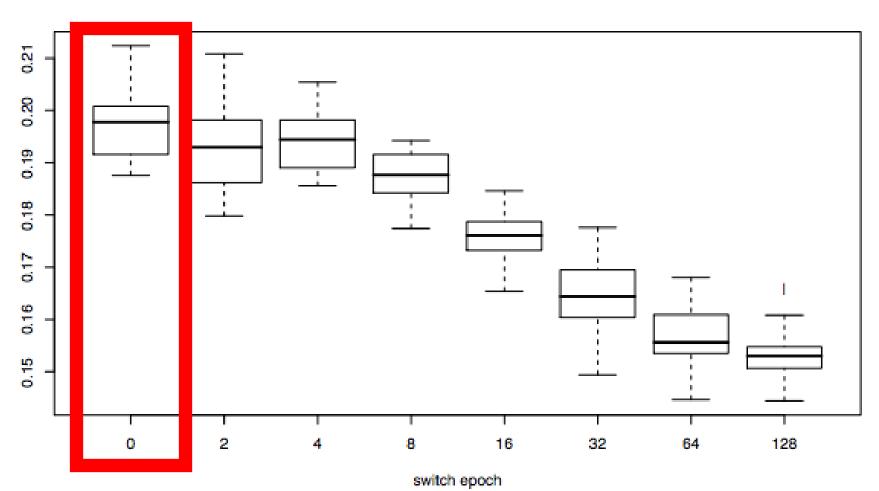
Bengio et al., Curriculum Learning, 2009

Shape Prediction: Curriculum Learning

Results of training on "easy" examples for *n* epochs and then training on "hard" examples until 256 epochs (20 random initializations).

What are benefits of curriculum learning?

How many epochs should the algorithm train with easy examples before switching to difficult examples?



No curriculum

Error

Bengio et al., Curriculum Learning, 2009

EfficientTrain: An ICCV 2023 Paper

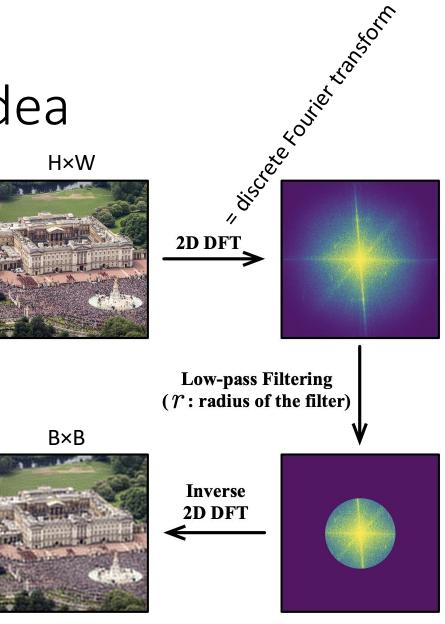
EfficientTrain: Exploring Generalized Curriculum Learning for Training Visual Backbones

Yulin Wang¹* Yang Yue¹* Rui Lu¹ Tianjiao Liu² Zhao Zhong² Shiji Song¹ Gao Huang^{1,3⊠} ¹Department of Automation, BNRist, Tsinghua University ²Huawei Technologies Ltd. ³BAAI {wang-yl19, yueyang22}@mails.tsinghua.edu.cn, gaohuang@tsinghua.edu.cn

Key idea: eliminate difficult patterns from all training examples at earlier learning stages by removing higher-frequency content

EfficientTrain: Key Idea

~20% training cost eliminated by initially training on lower resolution, low-frequency images to learn lowfrequency information typically learned first during training



(a) Low-pass Filtering (DFT: discrete Fourier transform)

B×B patch cropped in frequency domain

Key Questions In Creating "Curriculum"

- How to define what is "easy" versus "hard"?
- How many levels to include in the curriculum from easy to hard?

Efficient Computer Vision

- Motivation
- Model Compression
- Curriculum Learning
- Active Learning
- Faculty Course Questionnaire (FCQ)

How to teach machines with minimal human supervision?





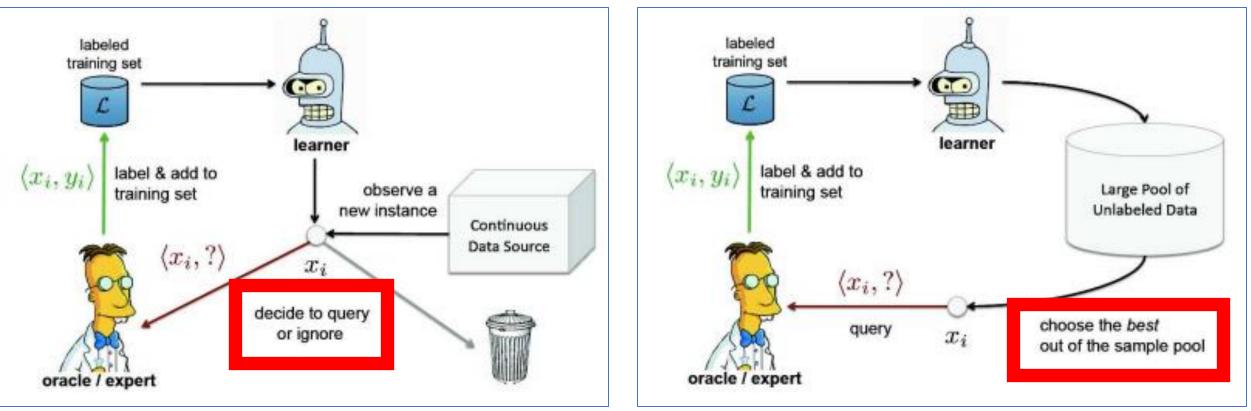
e.g., limited access to (expert) annotators

e.g., limited funding

Idea: Choose Most Informative Data to Label

Stream-Based

Pool-Based



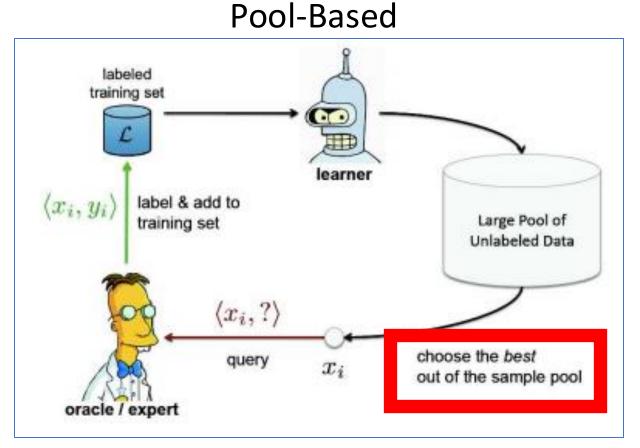
Consider one example at a time

Consider many examples at a time

Image Credit: https://www.cs.utah.edu/~piyush/teaching/10-11-slides.pdf

Active Learning for Neural Networks: Status Quo

Iteratively add more labelled training examples after *n* epochs; different from curriculum learning because labels need to be collected for the added data



Consider many examples at a time

Image Credit: https://www.cs.utah.edu/~piyush/teaching/10-11-slides.pdf

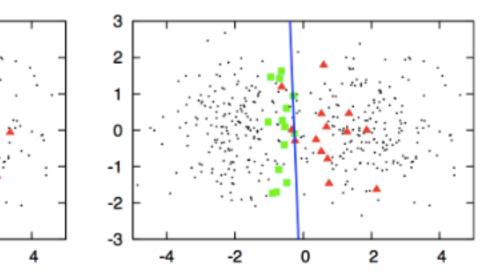
What approach might be effective in identifying the most informative data to label?

Common Approach: Uncertainty Sampling

Query instance(s) the classifier is most uncertain about.

True Representation (Assume Labels Are Not Known) з 2 -1 -1 -2 -2 -3 -3

Passive Learner (Random Selection) Active Learner (Uncertainty Sampling)



http://burrsettles.com/pub/settles.activelearning.pdf

e.g., Uncertainty Estimation for Neural Networks Using Robustness Testing

Use model's predictions on random augmentations of the input to measure consistency/uncertainty; e.g.,



Mirror Image



Figure Source: https://learnopencv.com/understanding-alexnet/

Elezi et al. Not all labels are equal: rationalizing the labeling costs for training object detection. CVPR 2022

e.g., Uncertainty Estimation for Neural Networks Using Ensembles (Two Approaches)

1. Dropout with different masks at inference time

2. Multiple neural networks

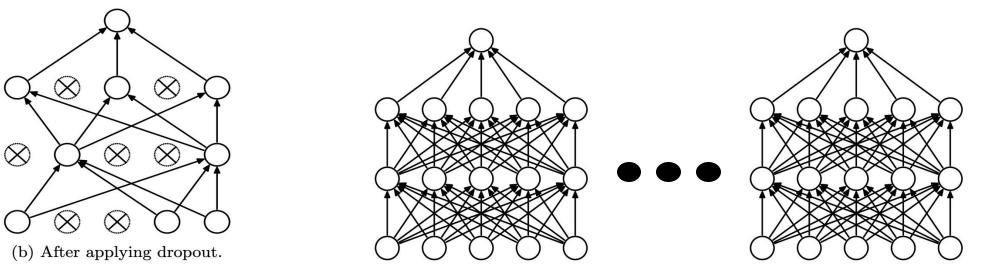


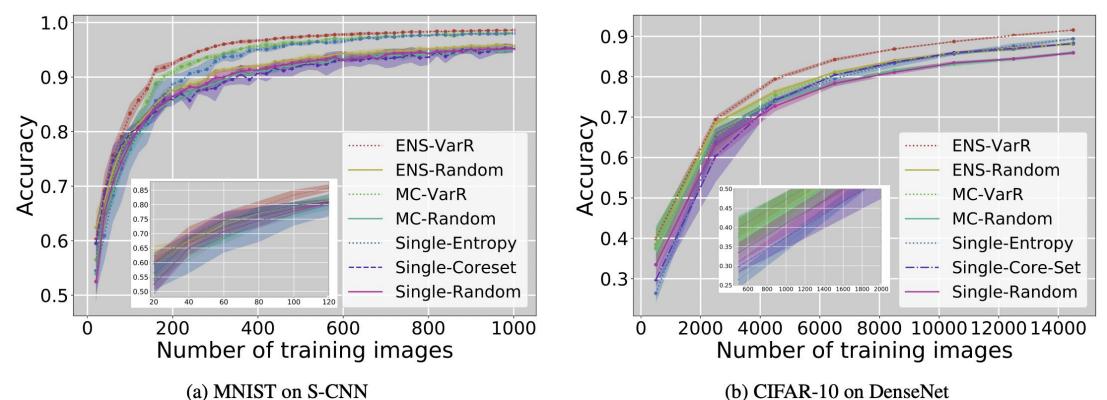
Figure Source: Srivastava et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research. 2014

Predicted softmax probabilities used to estimate uncertainty (e.g., entropy across softmax values), with average taken across all ensemble's softmax distributions

Beluch et al. The power of ensembles for active learning in image classification. CVPR 2018

e.g., Uncertainty Estimation for Neural Networks Using Ensembles (Two Approaches)

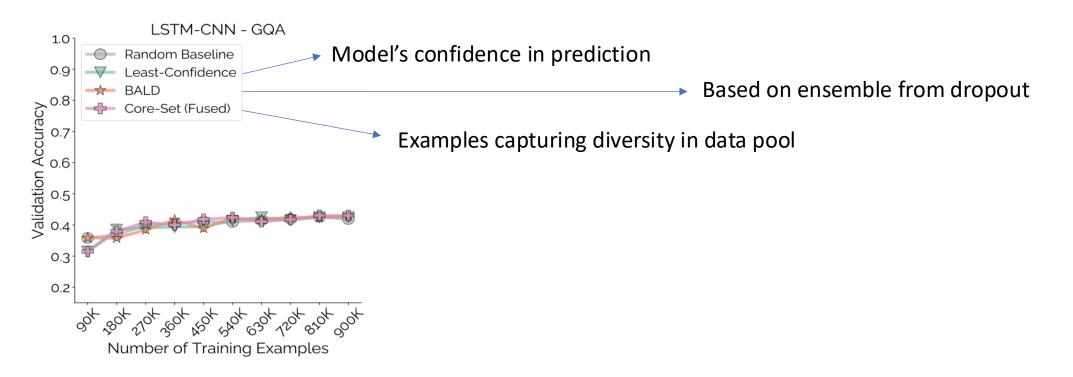
Active learning methods lead to faster learning and reduced human annotation effort than passive (random) learning for two image classification datasets



Beluch et al. The power of ensembles for active learning in image classification. CVPR 2018

Common AL Techniques Have Mixed Results

- Successes: image classification, object detection
- Failure: VQA (e.g., AL methods label 10% of overall pool per iteration; initial model trained on 10% of pool)



Karamcheti et al. Mind your outliers! Investigating the negative impact of outliers on active learning for visual question answering. Association for Computational Linguistics (ACL) 2021

Common AL Techniques Have Mixed Results

Why might AL methods perform comparable or worse to random selection? - Challenging examples to learn are sampled; e.g.,



External knowledge: What does the symbol on the blanket mean?



Underspecification: What is on the shelf?







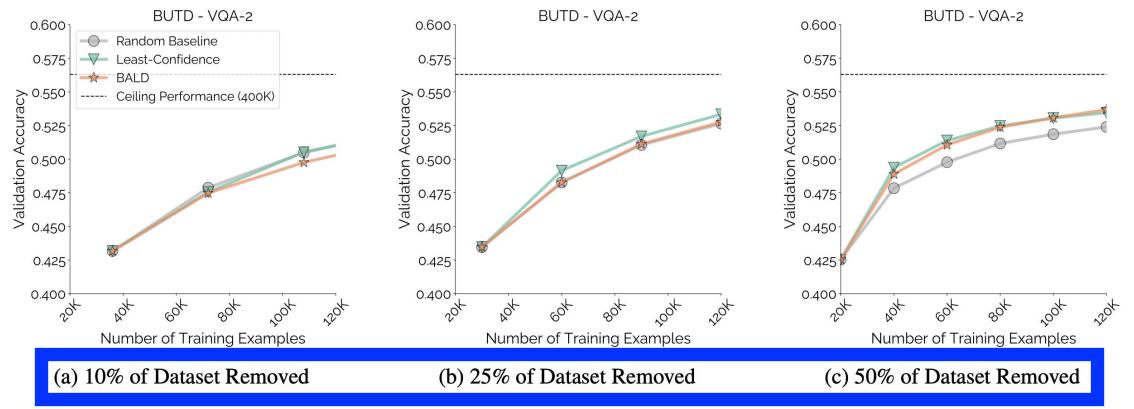
Multi-hop reasoning: What is the vehicle that is driving down the road the box is on the side of?

Figure 7: Example groups of collective outliers in the VQA-2 and GQA datasets.

Karamcheti et al. Mind your outliers! Investigating the negative impact of outliers on active learning for visual question answering. Association for Computational Linguistics (ACL) 2021

Idea: Remove "Unlearnable" Data from Pool

Performance compared to random selection improves for AL approaches when removing "challenging" examples from data pool



Karamcheti et al. Mind your outliers! Investigating the negative impact of outliers on active learning for visual question answering. Association for Computational Linguistics (ACL) 2021

Recent Works; e.g., (ICCV 2023 Papers)

Heterogeneous Diversity Driven Active Learning for Multi-Object Tracking

HAL3D: Hierarchical Active Learning for Fine-Grained 3D Part Labeling

ALWOD: Active Learning for Weakly-Supervised Object Detection

 $ng^{1,\dagger}$

Yuting Wang¹, Velibor Ilic², Jiatong Li¹, Branislav Kisačanin^{3,2}, and Vladimir Pavlovic¹

¹Rutgers University, NJ, USA
²The Institute for Artificial Intelligence Research and Development of Serbia, Novi Sad, Serbia
³Nvidia Corporation, TX, USA
yw632@rutgers.edu, velibor.ilic@ivi.ac.rs, jiatong.li@rutgers.edu,
b.kisacanin@ieee.org, vladimir@cs.rutgers.edu

Efficient Computer Vision

- Motivation
- Model Compression
- Curriculum Learning
- Active Learning
- Faculty Course Questionnaire (FCQ)

